## Table of Contents

### Track – Cloud Computing


### Track – Big Data

6. [PID-2] Panagiotis Nikitopoulos, “TEXT ANALYTICS EVALUATION ON BIG DATA PROJECTS USING ELASTIC SEARCH”

### Track – Digital Services, Social Media and Digital Collaboration

11. [PID-57] Brad McKenna, Michael Myers and Lesley Gardner, “ANALYSING QUALITATIVE DATA FROM VIRTUAL WORLDS: USING IMAGES AND TEXT MINING”
12. [PID-58] Shang Gao, Zhe Zang and Hong Guo, “A STUDY ON USERS’ ATTITUDE TO Pervasive Games in China”
17. [PID-105] Nikolaos Basias, Marinos Themistocleous and Vincenzo Morabito, “AN INNOVATIVE DECISION MAKING FRAMEWORK FOR E-BANKING INTEGRATION”

Track – E-Government

19. [PID-42] Ryma Abassi and Sihem Guemara El Fatmi, “PRIVACY PRESERVATION IN A PATTERN-AWARE AUTHENTICATION SCHEME FOR VEHICULAR AD HOC NETWORKS”

Track – Enterprise Systems

23. [PID-61] Luay Ahmad Anaya, Leif Skiftenes Flak and Dag Håkon Olsen, “START-UP COMPANY? GET YOUR ERP SYSTEM ASAP!”
24. [PID-104] Josephine Chong and Swati Nagar, “MARKET REACTIONS TO ENTERPRISE SYSTEMS INVESTMENTS IN AN EUROPEAN ECONOMY: AN EVENT STUDY”

Track – Green Information Systems

26. [PID-31] Nursen Aydin, Afshin Mansouri and Habin Lee, “INCREASING ENVIRONMENTAL SUSTAINABILITY BY OPTIMIZING VESSEL SPEED IN LINER SHIPPING”
27. [PID-63] Hojeong Yoon RELATIONAL BENEFITS AND CUSTOMER ENGAGEMENT IN SOCIAL MEDIA AND BRAND COMMUNITIES”
28. [PID-64] Saowanit Lekhavat, Nursen Aydin and Habin Lee, “DECISION SUPPORT SYSTEM FOR MULTI-OBJECTIVE SUSTAINABLE MARINE SHIPPING”
29. [PID-69] Oluwakayode Aluko, “FACTORS INFLUENCING ADOPTION OF GREEN SHIPPING PRACTICES AND INFLUENCE ON ORGANISATIONAL PERFORMANCE”

Track – Healthcare Information Systems


33. [PID-115] Leandros Stefanopoulos, “A SMARTPHONE APP FOR EFFICIENT NOTE KEEPING IN CONCURRENT MULTI-PARTICIPANT BEHAVIOURAL STUDIES”

Track – Innovative Research Projects

34. [PID-4] Nikolaos Chelmis, “A NOSQL DATABASES COMPARISON FOR E-HEALTH PROJECTS”


37. [PID-110] Vasileios Anagnostopoulos, Nikos Vlastaras and Dimosthenis Kyriazis, “IMPLEMENTING FAILURE DETECTION AND REPORTING IN A LIBVIRT CLOUD THROUGH AD-HOC MECHANISMS”


40. [PID-3] Leonidas Katelaris “E-Referral - PINCLOUD”

Track – Management and Organisational Issues in Information Systems


44. [PID-71] Ozgun Imre, “LEARNING BY NEGOTIATION – IMPLEMENTING A JOURNAL MANAGEMENT SYSTEM”

45. [PID-103] Hamed Sarbazhosseini, “ORGANISATIONAL ISSUES INVOLVED IN PROJECT PORTFOLIO MANAGEMENT IN TERMS OF ORGANISATIONAL STATE-TRANSITION APPROACH”
46. [PID-32] Wafi Al-Karaghoul, Mohamed Haffar, Ahmad Ghoneim and Sulaf Assi A QUALITATIVE STUDY OF THE SIGNIFICANT MOTIVATORS FOR SHOPING ONLINE AMONG UNDEGRADUATE STUDENTS IN THE UK

Track – Security and Privacy Protection for Information Systems

47. [PID-7] Konstantina Vemou, Gavriela Mousa and Maria Karyda, “ON THE LOW DIFFUSION OF PETS IN SNS: RESULTS OF AN EMPIRICAL INVESTIGATION”


49. [PID-66] Martin Brodin, Jeremy Rose and Rose-Mharie Åhlfeldt, “MANAGEMENT ISSUES FOR BYOD”

50. [PID-68] Ehinome Ikhalia and Alan Serrano, “A FRAMEWORK FOR DESIGNING AN EFFECTIVE SECURITY AWARENESS SYSTEM FOR ONLINE SOCIAL NETWORK USERS”

51. [PID-74] Norilyz Figueroa and Ramakrishna Ayyagari, “TRAINING USERS ON INFORMATION SECURITY: EVIDENCE FROM JAVA APPLETS”
Big Data And Analytics

Chairs: Dr. Paulo Rubino Cunha
Table of Contents

Track – Big Data

1. [PID-2] Panagiotis Nikitopoulos, “TEXT ANALYTICS EVALUATION ON BIG DATA PROJECTS USING ELASTIC SEARCH”
TEXT ANALYTICS EVALUATION ON HEALTHCARE USING ELASTICSEARCH

Panagiotis Nikitopoulos: Department of Digital Systems, University of Piraeus, Greece, nikp@unipi.gr
Marinos Themistocleous: Department of Digital Systems, University of Piraeus, Greece, mthemist@unipi.gr

Abstract

Due to the explosion in the amount of medical information available, search techniques are gaining much attention in the medical domain. These techniques have so far been applied exclusively to applications focusing on users with technical knowledge in the medical domain (i.e. doctors). Personal healthcare systems are emerging as the new standard way of treating patients these days. The fact that these systems can be used by non-medical related users, makes the Information Retrieval process even more challenging. In this paper we study the feasibility of implementing search engines on medical related information, for systems that can be used by members of the general public (i.e. patients). To this end, we propose a solution which handles complex text queries execution based on Elasticsearch. The proposed solution is tested for efficiency and effectiveness in the context of the Pincloud project.

Keywords: Medical Information Retrieval, Text Analytics, Elasticsearch.

1 INTRODUCTION

Several research studies in the field of medical information management deal with the problem of medical information retrieval. These studies can generally be classified into one of two broad categories. The first category concerns studies that deal with the information needs of clinicians, and focuses on searching specialized databases in response to specific and well-informed topics. The second category deals with so-called “consumer” information needs: searches conducted by non-medical users (often patients or family members) over the Internet. Earlier medical information retrieval evaluation campaigns have tended to focus on the first of these two groupings (Bedrick and Sheikhshabbafganghi, 2013).

In a world that constantly adopts electronic healthcare solutions, the healthcare industry continues its quest for the ideal computing platform to serve caregivers and patients. Monitoring the physical status and health of humans or patients at home is an interesting solution compared to hospitalization in healthcare facility institutions, since it offers a medical surveillance in a familiar atmosphere for the patient and can reduce the costs of medical treatment. This service is well known with the term “homecare” and the corresponding systems are called homecare systems or platforms. The key benefit in homecare systems is the fact that these systems can be used by non-medical users, at their place, using their own preferred device. A challenging aspect of such systems, is the medical information retrieval subsystem, which should meet the needs of regular every day users, in the context of a system which processes information that may not be very familiar to them. Text analysis could be used in the information retrieval subsystem, in order to help users unfamiliar to the medical context of the homecare system.

The Pincloud project (Pincloud) aims at providing integrated e-Health services for personalized medicine utilizing cloud infrastructure. State of the art technologies like cloud computing and Service Oriented Architectures (SOA) are used to provide efficient, scalable, portable, interoperable and integrated IT infrastructures that are cost effective and maintainable. Despite the significant importance of these technologies, the healthcare sector has yet not paid much attention on these technologies.
Hence Pincloud is as project, where medical information retrieval research studies (Goeuriot et al., 2014, Zuccon et al., 2012, Medsker and Small, 2012) can be applied in order to provide advanced medical services to the end users. It is evident that the project’s tasks require a reliable and efficient text analytics mechanism that takes into consideration the special needs of a medical e-Health project that will be used by non-medical users. In this paper, we will examine the feasibility of applying text analytics mechanisms in the Pincloud personal e-Health systems. To this end, a text analytics mechanism will be conducted and applied to the Pincloud project, by using the Elasticsearch product (Kuc and Rogozinski, 2013, Elasticsearch).

The rest of the paper is structured as follows: Section 2 provides background theory on Information Retrieval field and then discusses related work on medical Information Retrieval, while Section 3 introduces the Elasticsearch product and the implementation we conducted for the Pincloud project. Section 4 concludes the paper.

2 LITERATURE REVIEW

2.1 Background Theory

The goal of information retrieval (IR) is to satisfy an information need from within a large collection of material (Sanderson and Croft, 2012). A large number of IR models has been developed and some of the most common ones include the Boolean model, vector space model and language model. An information retrieval system is an implementation of an IR model.

To get started with IR it may be good to use database systems as a reference to some of the main concepts of IR systems. Some of the most common terms in database systems and their corresponding IR terms are listed in Table 1. Even though most IR terms can be likened to database terms there are some major differences between the two concepts. A database consists of structured data. Queries to a database are also structured and the results from a query are data which are exact matches to the query.

In IR systems both the data and the queries are unstructured, usually consisting of natural language text. The retrieval method in IR systems is probabilistic meaning that data returned from a query are not exact matches. IR systems are said to be searching for their data and thus they are often referred to as search engines.

<table>
<thead>
<tr>
<th>Traditional DB System</th>
<th>IR System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row, Tuple</td>
<td>Document</td>
</tr>
<tr>
<td>Column</td>
<td>Field</td>
</tr>
<tr>
<td>Table</td>
<td>Index</td>
</tr>
<tr>
<td>Database</td>
<td>Collection</td>
</tr>
</tbody>
</table>

Table 1. Common database terms and their corresponding IR terms.

Determining if a document is relevant to a query is a fundamentally hard problem. The only way to really determine if a document is relevant to a query or not is for a user to judge it as relevant or not (Sanderson and Croft, 2012). Most IR models assign a score to documents given a query, where a higher score is more likely to be judged as relevant by a user compared to a document with a lower score. Some of the most common scoring models will be described below:

- **The Tf-Idf scoring model** assumes that a document/document-field with a high frequency of the query-terms is more relevant to the query (Huang et al., 2012). Using only this criterion would however disproportionately discriminate against less common terms in a query. For example, in the query "Computer Science Chalmers" the terms "computer" and "science" probably have a high term-frequency in many documents, but the term "Chalmers" is probably the most important term since it is the most specific. As a result, the tf-weights are combined with the inverse document frequency (idf) weights. Df is calculated by counting the number of postings for a term in the inverted index, and the idf is calculated by dividing the total number of documents by this number as in the following equation:
The combination of tf-weight and idf-weights determines the score for a document given a query of terms as is displayed in the following equation:

$$\text{score}(d, q) = \frac{\vec{v}_d \cdot \vec{v}_q}{|\vec{v}_d| \times |\vec{v}_q|}$$

- In the vector-space model documents and queries are represented as vectors of weighted terms. The weights can be calculated in different manners, but a common approach is to use the Tf-idf weights described in the previous paragraph. Relevance between a document $d$ and a query $q$ is determined by the cosine of the angle between the vectors, often called cosine similarity and is calculated as in the following equation:

$$\text{score}(d, q) = \frac{\vec{v}_d \cdot \vec{v}_q}{|\vec{v}_d| \times |\vec{v}_q|}$$

2.2 Related Work

Search techniques have gained much attention recently by the research community dealing with the context of medical information retrieval, due to an explosion in the amount of medical information available. This includes both patient-specific information (e.g. electronic health records) and knowledge-based information (e.g. scientific papers). All parties involved in medical treatment are regularly faced with an information need that cannot be met from their own store of knowledge. Medical information search is a classic example of a domain-specific search, but the medical domain is so complex that simply limiting the target documents to those in the medical domain and indexing with a standard search engine is not sufficient.

All parties involved in medical treatment are regularly faced with an information need that cannot be met from their own store of knowledge. Medical information search is a classic example of a domain-specific search, but the medical domain is so complex that simply limiting the target documents to those in the medical domain and indexing with a standard search engine is not sufficient. The range of sources of medical information (primary research sources, secondary research sources, web pages and popular publications, etc.) and the range of end users (members of the general public, general practitioners, specialists, researchers, etc.) lead to complex requirements. Recent work in the Khresmoi project (Khersmoi) has collected end user requirements for search in the medical domain through online surveys, interviews and observation.

Although many groups of people search for information in the medical domain, the analysis in the Pincloud project has been restricted to two of the most important groups: search by physicians and search by members of the general public. Different end users have different requirements based on the technical level (accessibly or technically written), level of specificity (overview or in-depth), language of the results, etc. Naturally, the trustworthiness of medical information is of particular importance.

For physicians, an unmet information need has been reported as occurring for 2 of every 3 patients seen (Hersh and Hickam, 1998), or for 41% of the questions they pursued (Ely et al., 2007). This requires that they attempt to meet this information need by using available resources, which has traditionally involved searching in printed sources and asking colleagues, although searching on the Internet is of increasing importance. The Khresmoi survey (Kelly et al., 2013) has shown that the three most common sources of online information used by physicians (in decreasing order of usage) are: general search engines (e.g. Google, Bing, Yahoo!), medical research databases (e.g. Pubmed) and Wikipedia.

Patients also have regular information needs, illustrated by the fact that 61% of American Adults seek out health advice online (Fox, 2011). The Khresmoi survey (Kelly et al., 2013) of the general public revealed that the most common sources of online information used by this group are: general search engines (e.g. Google, Bing, Yahoo!), web sites providing health information (e.g. university, hospital,
pharmaceutical company) and Wikipedia. However, as this group has had less training in interpreting the information found, issues such as trustworthiness of the information and cyberchondria (White and Horvitz, 2009) gain more importance.

A number of publications towards meeting the requirements for searching in the medical domain have recently appeared. This includes work on estimating the granularity of information in a document (Yan et al., 2011), estimating the difficulty of a document (Collins-Thompson et al., 2011, Leroy and Endicott, 2012), the quality of documents (Tang et al., 2006), document summarization (Afantenos et al., 2005) and the use of terminology resources for query refinement (Luo and Tang, 2008). Cross-lingual search is of importance for end users at all levels (Bedrick, 2010).

3 INFORMATION RETRIEVAL USING ELASTICSEARCH

Complex IR systems that deal with text queries are often referred as search engines in literature. Search engines are NoSQL database management systems dedicated to the search for data content. In addition to general optimization for this type of application, the specialization consists in typically offering the following features (Croft et al., 2010):

- Support for complex search expressions.
- Full text search.
- Stemming (reducing inflected words to their stem).
- Ranking and grouping of search results.
- Geospatial search.
- Distributed search for high scalability.

Lucene (Lucene) has emerged in recent years as the most popular free information retrieval library. Lucene is a mature, full-text search, open source library, written in Java, which can be used for indexing and scoring documents. It supports several IR models but is shipped with an implementation of the vector-space model with tf-idf weights as presented in the previous section. Since Lucene has been used in and maintained by many different applications, there is an implicit trust in the basic search capabilities of the products using it. Famous search engine products, such as Solr (Solr) and Elasticsearch (Elasticsearch) use Lucene under the hoods. The Solr product is for standalone applications while Elasticsearch is designed for modern, cloud environments. Thus in the Pincloud project we decided to use Elasticsearch for our text queries.

3.1 The Elasticsearch product

Elasticsearch (Elasticsearch) is an open-source distributed search engine under the Apache 2.0 license. It was built with big data in mind which has given it an emphasis on scalability and reliability. A running instance of Elasticsearch is called a node and together they form a cluster (Kuc and Rogozinski, 2013). As the name implies it is very “elastic” in that it automatically handles rebalancing of indices and shards when new nodes are added or removed from the cluster. As a result developers may add or remove nodes in order to increase or reduce resources allocated for the cluster on demand. Since Elasticsearch uses Lucene, there exists an implicit trust in the basic search capabilities of this search engine. Thus the developers of Elasticsearch have been able to focus on usability, scalability and performance of their product.

Elasticsearch includes out of the box support for the following features:

- **Multiple indices:** Elasticsearch supports sharding of indices, but the number of shards of an index has to be set when the index is created and can thus not increase or decrease on demand (Gormley and Tong, 2014). To compensate for this Elasticsearch has support for multiple indices, a feature which distinguishes it from most other search engines. Instead of increasing the number of shards when the number of documents grows it’s recommended to construct a new index with the same type. A query may be forwarded to one or many indices which can be
specified in the query. Thus, an index may also be referred to as a shard in Elasticsearch, if there are many indices with the same type. From now on a shard in this report will refer to an index in Elasticsearch.

- Query routing: To enhance its scalability further Elasticsearch comes with a sophisticated system for routing queries to nodes. The simplest example is when Elasticsearch is used as a database, where the ID of documents to fetch is specified in a query. In this case only the nodes where the documents reside will process the query. Since Elasticsearch is mainly used for searching this feature only has a limited value. There are other features which may be used for query-routing (Gormley and Tong, 2014). An index in Elasticsearch support different types. Each type may be assigned a specific routing-value. If a routing-value is assigned to a type the default policy is to cluster the documents of that type together in the same shard. If a type is specified in a query it will only be routed to the nodes which contain documents of that type.

In the following subsections, we introduce innovative methods for indexing and querying Greek documents using Elasticsearch

### 3.2 Indexing Documents using Elasticsearch

The main database in the Pincloud project, is a MongoDB instance, which stores all the medical related information for the patients using the web application. While data are being loaded into the MongoDB data store, a mechanism called River is responsible for copying select data from the MongoDB instance to the Elasticsearch cluster. A River is a pluggable service running within Elasticsearch cluster pulling data (or being pushed with data) that is then indexed into the cluster. The River can be configured to store data to different Elasticsearch indices according to the MongoDB collection that the source data belong to.

![Indexing transformation process of medical related data in the Pincloud project using Elasticsearch.](image)

Elasticsearch indices can be configured to transform input data in order to enable complex text queries execution on top of them. Regarding Pincloud project’s technical specifications, input data are transformed in a fashion that enables Greek users (in particular patients) use the web application in a
more familiar way, by not requiring special knowledge in the medical field. The River mechanism provides input data to the Elasticsearch indices. The data transformation process occurs right after the River mechanism has transferred input data in the Elasticsearch cluster, and before these data are stored into memory (main memory, or secondary memory). The exact steps of the indexing process can be configured in the index initialization phase. Every Elasticsearch index is configured to execute a simple or more complex data transformation process before storing its data. For the Pincloud project we conducted a series of steps, called filters in terms of Elasticsearch, which take place in the order shown in Figure 1.

Each step receives input from the previous step, transforms the data and sends output to the next step. The transformation process is explained step by step below, using the phrase “The quick brown fox jumps over a lazy dog,” as an input example:

- **The Standard Tokenizer** filter is responsible for splitting the text into a list of words. The Tokenizer implements the Unicode Text Segmentation algorithm, as specified in (segmentation). The input example is transformed in a list of words like this: "The" "quick" "brown" "fox" "jumps" "over" "a" "lazy" "dog" ".".

- **The Standard Analyzer** filter normalizes each word in lowercase while eliminating any accent and punctuation marks. The list of words becomes like this: "the" "quick" "brown" "fox" "jumps" "over" "a" "lazy" "dog" .

- **The Greeklish Analyzer** is a custom filter that is being used to meet the complex requirements of Greek users. Greek users often use Greeklish (a portmanteau of the words Greek and English) in their computer interactions. Greeklish is the Greek language written using the Latin alphabet. The Greeklish Analyzer is responsible for transforming Greek text into Greeklish text, using a number of rules written in custom java code. The Greeklish Analyzer is in fact a plugin that integrates into Elasticsearch system, enabling it to transform Greek text to Greeklish text. The example phrase is written in English, so there is no impact to the phrase by this analyzer.

- **The Stop Filter** is a filter that removes stop words from token streams. Stop words are words which are filtered out before or after processing of natural language data (text). Examples of stop words include words like “the”, “a”, “this” etc. The list of tokens becomes like this: ”quick” ”brown” ”fox” ”jumps” ”lazy” ”dog” .

- **The Stemmer** is a filter that removes the commoner morphological and inflexional endings from words in a text. For example, the word “jumps” becomes “jump”, the word “foxes” becomes “fox”, the word “lazy” becomes “laz” etc. The list of tokens in transformed in the following way: ”quick” ”brown” ”fox” ”jump” ”laz” ”dog”.

- **The n-gram Tokenizer** is a filter that constructs several tokens of a single input word by adding one letter of the word in each iteration. For example the word ”jump” produces the tokens ”j”, ”ju”, ”jum”, ”jump”. In the Pincloud project we eliminate the first two tokens, that contain only the first two letters of the word, resulting to a more representative list of tokens ”jum”, ”jump”. The resulting list of tokens for the input phrase is relatively large and due to lack of space, it cannot be presented here.

At the end of the aforementioned index data transformation process, the transformed data can be stored either in main or secondary memory. The choice resides on the system administrator, who has to make that decision. Of course main memory storage is a better choice for faster query results, while secondary memory (i.e. hard drives) is a better choice for huge indices. For the Pincloud project we decided to use main memory storage due to the small amount of data being stored into Elasticsearch indices.

### 3.3 Querying Documents using Elasticsearch

The indexed data can later be queried using the JSON based REST API of Elasticsearch. Queries are processed and transformed in a similar fashion to the indexing process. The exact steps of the querying process can be configured in the index initialization phase. Whenever an index is queried, the query transformation process is fired to transform the query text into a more appropriate text, which can be
matched to the indexed data. Different query transformation steps can be applied to different Elasticsearch indices. The Elasticsearch’s REST API allows the caller of the query web service, to specify the index to be queried.

The query process we conducted for the Pincloud project is presented in Figure 2. It is crucial that each step occurring in the query transformation process can be executed relatively fast to enable fast results production by the Elasticsearch product. In the Pincloud project, we carefully designed the index transformation process and the query transformation process, while trying to minimize the impact of the query steps to the performance of the system. Tests have shown that the query transformation process occurs fast, and has minimal impact to the system’s performance.

The steps of the query transformation process have already been explained in the index transformation process. Here the filters “Greeklish Analyzer” and “n-gram Tokenizer” have been excluded to enable better results and faster answers.

The importance of the processes can be shown by considering the following example: the phrase “The quick brown fox jumps over a lazy dog.” is indexed by Elasticsearch according to the way illustrated in the previous subsection. Then a user searches for the aforementioned phrase by writing the following query: “Brown foxes jump over Dogs”. This phrase is transformed to a list of tokens such as "brown" "fox" "jump" "dog". All of these tokens can be found in the list of tokens that was generated while indexing the phrase “The quick brown fox jumps over a lazy dog.” So it is obvious that this phrase is a candidate result for the user’s query, although the query and the phrase do not match exactly. Of course in order for this phrase to be reported in the query results, it must be assigned a score, regarding its relativeness to the original query. If other phrases existing in the index are ranked higher with a better score, then this phrase could be excluded by the query results. We conducted a series of tests for query results evaluation, and we noticed that our process works as expected, reporting the intended results.

4 CONCLUSIONS

In this paper we presented an overview of the Information Retrieval research field, while introducing some common key characteristics of the state of the art search engines. We then focused on related research studies on the field of medical Information Retrieval, showing that recent work focuses mainly on technical related users such as medical stuff. On the other hand we focused primarily on the group
of users that belong to the general public like patients. In Section 3 we introduced some state of the art text analytics search engines that are based on the well-known Lucene library. We provided a list of features that these search engines commonly support and stated the reasons for why we chose Elasticsearch as our search engine in the Pincloud project. We then described the techniques and features we took advantage of the Elasticsearch product, while conducting the data transformation process for the indexing and querying phases. We showed that the general process is efficient and effective, resulting to a more familiar and user friendly application environment, even for the users that belong to the general public.

Acknowledgement

The work reported here has been carried out in the framework of national project Providing Integrated eHealth Services for Personalized Medicine utilizing Cloud Infrastructure (Pincloud), led by the University of Piraeus, conducted in the context of the National Strategic Reference Framework NSRF 2007-2013, Cooperation 2011 and co-funded by the European Commission.

References

IMPROVED SINGLE-LABEL TEXT CATEGORIZATION BY
INSTANCE FILTRATION

Kashif Khan, email: kashif.khan421@gmail.com, National University of Science and Technology, Pakistan

Abstract

Machine learning classifiers are widely used for text categorization however a classifier misclassifies some of the instances into a category that is relevant to their actual category. The categorization ability of a classifier can be improved by filtering dataset with better classifier and removing such category for misclassified instances. In this paper we proposed a two level approach where level-1 filters instances according to their likelihood in each category and reduce training dataset to top ranked 't' categories and their instances whereas level-2 classifier is used to classify instances with filtered training set. We employed Naïve Bayes, SVM and KNN as machine learning classifiers. Experimental evaluations on standard reuters-21578 and 20 Newsgroups datasets showed improved categorization effectiveness as measured by accuracy, precision, recall and f-measure protocols.

Keywords: text classification, SVM, KNN, Naïve Bayes.

1 INTRODUCTION

Automated text categorization plays a vital role in information retrieval and machine learning. Text document categorization is a task of classifying text documents in predefined categories based on their content (Vidhya.K.A and G.Aghila, 2010). In this era of BigData, it is nearly impossible to manually categorize all this data. Therefore, more efficient categorization algorithms are deemed necessary.

Computers cannot recognize human language. In order, to process text document we need to convert them in computer recognizable format. The model usually used for this purpose is known as vector space model. In this model each document is converted into vector of terms (also known as features), and each term is assigned a weight. The basic weighting scheme uses binary weights (w = 1 if term is present in document, and w = 0 otherwise). A better and much referenced weighting scheme is the tf-idf (term frequency-inverse document frequency) scheme.

2 PREVIOUS WORK

Different Automated text categorization approaches are introduced till date. The most famous of these approaches are usually based on Decision Tree (Stephen Marsland, 2009), Naïve Bayes (Vidhya.K.A and G.Aghila, 2010; Tom M. Mitchell, 1997), Support Vector Machine (Pan Hao, Duan Ying and Tan Longyuan, 2009), Neural Network (Tom M. Mitchell, 1997), K-Nearest Neighbor (Peter Harrington, 2012).

In order, to improve individual classifier results for text categorization many methods are studied. These methods includes dimensionality reduction techniques i.e. Information Gain (Charu C. Aggarwal and ChengXiang Zhai, 2012), Mutual Information (Charu C. Aggarwal and ChengXiang Zhai, 2012), PCA (Peter Harrington, 2012), LSI (Peter Harrington, 2012) etc. These methods reduce dimensions and improve classification accuracy of individual classifier however dimensionality reduction may not be the best option in all high dimensional problems (Hua-Liang Wei and Stephen A. Billings, 2007).

The performance of dimensionality reduction techniques can be partially attributed to the classification algorithm and domain characteristics (Dunja Mladeníc, 2006). Tao Liu et.al (Tao Liu et.al, 2003) found that there are few noisy terms in 20NewsGroup and Reuters dataset so feature selection may reduce the dimensions but many terms with discriminative power for classification may
also remove. Evgeniy Gabrilovich et al (Evgeniy Gabrilovich and Shaul Markovitch, 2004) also found that features selection for 20NewsGroup dataset degrades classification accuracy for the SVM. Dunja Mladenić (Dunja Mladenić, 2006) found that Information Gain (Charu C. Aggarwal and ChengXiang Zhai, 2012) completely failed on reuters dataset.


In this paper we focus on individual machine learning classifiers and propose an approach for improving their effectiveness by instance filtration and reducing training sample set to top ranked ‘t’ categories and their instances. The aim of this study is to improve the categorization effectiveness of individual machine learning classifier particularly for the dataset with low noise content.

3 COMPONENTS OF TEXT CATEGORIZATION

There are two kinds of approaches for text categorization. i) Rule based (Pratiksha Y. Pawar and S. H. Gawande, 2012) ii) Machine learning (Pratiksha Y. Pawar and S. H. Gawande, 2012). In this paper we targeted machine learning approaches. Text document categorization is divided into following components

- Text Document Pre-processing
- Text Document Classification

![Block Diagram of Text Categorization](image)

3.1 Text Document Pre-processing

The existence of noise terms i.e. special characters and whitespaces etc. makes it difficult to categorize documents in raw form. Therefore, to achieve optimum performance these raw documents are pre-processed and converted into structured form. Pre-processing also helps in reducing the size of documents (Mita K. Dalal and Mukesh A. Zaveri, 2011). We performed the following pre-processing steps for our dataset

- Dataset Cleaning
- Tokenization
- Stop Words Removal
- Terms Pruning
- Stemming
- Generating N-Grams
- Vector Space Model representation for documents with TF-IDF weighting

3.2 Text Document Classification

Machine learning is classified as supervised (Parag Kulkarni, 2012), semi-supervised (Parag Kulkarni, 2012) and unsupervised (Parag Kulkarni, 2012) machine learning. In this study supervised machine learning techniques are employed.
Text categorization is divided into two main types i) Single-label (Hiteshri Modi and Mahesh Panchal, 2012) ii) Multi-label (Hiteshri Modi and Mahesh Panchal, 2012). In this research we are dealing with single-label text categorization.

Classifiers in the machine learning uses text document vectors and features extracted in the pre-processing phase.

3.2.1 Naïve Bayes

Naïve Bayes is a famous probabilistic machine learning classifier for text categorization. It is very effective despite of the naïve assumption that states that all the attributes in the datasets are independent of each other (Vidhya.K.A and G.Aghila, 2010). It computes posterior probabilities of a document for all categories and assigns the document to a category with maximum posterior probability

\[ P(\mathbf{d} | \mathbf{c}_j) = \prod_{i=1}^{m} P(d_i | \mathbf{c}_j) \]

Where \( P(\mathbf{c}_j) \) is prior probability for class \( \mathbf{c}_j \), and \( P(d_i | \mathbf{c}_j) \) is likelihood of document \( d_i \) in class \( \mathbf{c}_j \). In above equation can be calculated as

\[ P(\mathbf{c}_j) = \frac{N_j}{N} \quad \text{and} \quad P(d_i | \mathbf{c}_j) = \frac{(1 + N_{i,j})}{M + \sum_{k=1}^{K} N_{k,j}} \]

Where \( N_j \) is Number of documents in class \( \mathbf{c}_j \), and \( N \) is Number of documents in the dataset. \( N_{i,j} \) is the number of times the word ‘i’ is found in the documents that belong to class \( \mathbf{c}_j \) in training set and \( M \) is the size of vocabulary.

The two variations of Naïve Bayes are Multivariate Bernoulli Naïve Bayes (Charu C. Aggarwal and ChengXiang Zhai, 2012) and Multinomial Naïve Bayes (Charu C. Aggarwal and ChengXiang Zhai, 2012). The performance of Naïve Bayes despite of Naïve assumption is surprisingly effective (Ali Danesh, Behzad Moshiri and Omid Fatemi, 2007).

3.2.2 K-Nearest Neighbor

It is a lazy learner classifier because it does not build model (Minakshi Sharma and Suresh Kumar Sharma, 2013). It does not perform any computations on training time and delay all the computational overhead till classification time. It calculates distance/similarity of a test document ‘x’ to all training documents in the training data and predict the class of test document ‘x’ which has the maximum representation in surrounding ‘k’ nearest neighbors. The distance/similarity between a test document ‘x’ and training documents can be calculated via Cosine similarity (Stephen Marsland, 2009), Euclidean distance (Minakshi Sharma and Suresh Kumar Sharma, 2013), Manhattan distance (Minakshi Sharma and Suresh Kumar Sharma, 2013) and Minkowski distance (Minakshi Sharma and Suresh Kumar Sharma, 2013).

It is a robust classifier with small error ratio (Minakshi Sharma and Suresh Kumar Sharma, 2013). The choice of ‘k’ depends on the dataset and may vary with different datasets. In general if value of ‘k’ is too small, the result can be sensitive to noise points (Pradeep Mewada et.al, 2011) whereas if value of ‘k’ is too large than neighborhood may include too many points from other classes (Pradeep Mewada et.al, 2011).
3.2.3 Support Vector Machine

It is a supervised text classification technique; widely used to address text categorization problems. It works by finding a separating hyper plane with maximum margin to separate positive and negative examples from the training data.

![Figure 2. SVM Separating Hyperplan for Binary Problem](image-url)

It is a two class classifier in its standard form, for each topic ‘t’ in a multi-class the problem is converted into multiple un-related binary problems and multiple one-vs-all (or in-topic vs not-in-topic) SVM classifiers are produced with a scoring function.

SVM is not a natural probabilistic classifier and its output is turned into probabilities by calibration. Platt scaling (Alexandru Niculescu-Mizil and Rich Caruana, 2005) and isotonic regressions (Alexandru Niculescu-Mizil and Rich Caruana, 2005) are the calibration methods by which output of SVM is transformed into probabilities. SVM becomes one of the best probabilistic classifier after calibration for generating well calibrated probabilities (Alexandru Niculescu-Mizil and Rich Caruana, 2005).

4 Instance Filtration

In multi-class text categorization, the likelihood of occurring of an instance in each category is not same. The probabilistic classifier in machine learning helps in learning about this likelihood/probability of an instance for each category. The probability for the category which is more relevant for the document/instance is higher than probability for the other categories.

In this study, classification task is divided in two levels. The first level, deals with filtering test instances. In this level, the likelihood for a test instance in each category is calculated from training data. The categories are ranked by the likelihoods calculated and training data is reduced to top ranked ‘t’ categories and their instances.

In second level, classifier is trained on the filtered dataset that we obtain from level-1. This dataset has a small subset of categories and their instances to classify a test instance.

We used SVM as level-1 classifier whereas KNN and Naïve Bayes are used as level-2 classifiers in order to study the impact of filtration on the effectiveness of level-2 classifiers.

Based on experiments result level-2 classifiers achieves better effectiveness with filtered dataset than with original dataset. Fig 3, shows the instance based filtration process.

![Figure 3. Instance Based Filtration Process](image-url)

4.1 Categories Ranking/Pruning

Sometime it is desirable for a classifier to generate score that reflect confidence of classifier that a document ‘d’ belongs to class ‘c’.
In categories pruning, categories are ranked and dataset is filtered to top ‘t’ categories and their instances whereas remaining categories (with instances) are removed from the dataset. In machine learning, probabilistic classifier assigns score to each category for a test instance which shows relevance to that category.

This score/probability is used to rank categories. We need to determine threshold for a dataset to find out the number of categories to keep in the filtered dataset for each instance. This threshold is not fixed and may vary with datasets.

4.2 Threshold Selection

It is one of the most important areas and may affect the performance of the system. Any value ‘t’ for a threshold means that only top ranked ‘t’ categories and their instances are relevant for the classification of an instance. There is not much literature available on threshold selection which is why we employed one of the most commonly used threshold selection algorithm in this research.

The mostly used threshold selection algorithm are Rcut (Benhui Chen, Weifeng Gu and Jinglu Hu, 2010), Pcut (Benhui Chen, Weifeng Gu and Jinglu Hu, 2010), Scut (Benhui Chen, Weifeng Gu and Jinglu Hu, 2010). We used Rcut (Benhui Chen, Weifeng Gu and Jinglu Hu, 2010) and select top ‘t’ categories where the value of ‘t’ is selected as the one that generates best results for validation set.

5 Experiments and Results

5.1 Datasets

5.1.1 20 Newsgroups Dataset

The first dataset we used is publicly available “20 news group”. There are many versions of this dataset available however the one we used is 20newsgroup-18828 dataset (Lang, K., 1995). It is a single label dataset where each document is assigned to exactly one category. We created a dataset D1 250 samples/category where randomly selected 20% (50 instances/category) of the D1 is used as validation set and 80% (200 instances/category) is used to evaluate the system.

5.1.2 Reuters-21578 Dataset

The second dataset is reuters-21578 (David D. Lewis, 2004). It is a multi-label dataset in original however we used its single-label version provided by Ana (Ana Cardoso Cachopo, 2007). We created two datasets D2 with most frequently used 8 categories and D3 with top 24 categories where 20% randomly selected samples of D2 and D3 are used as validation set and 80% are used to evaluate the system.

5.2 Experiments

We applied stratified 10-fold cross validation for D1 whereas standard ModApte split was used for D2 and D3 to keep our experimental settings close to other existing approaches. The validation set of each dataset (D1, D2 and D3) was used separately to determine threshold.

The distribution of samples in reuters-21578 was much skewed with some categories containing very few documents. This distribution made tuning of parameters via validation set impossible which is why we only selected categories with at least 20 training documents in D3.

We pre-processed our documents in D1, D2 and D3 for all the pre-processing steps defined in section 3. We cleaned our datasets from all the special characters, dates and digits. We removed stop words by using SMART list whereas porter stemmer was used to stem words to their root i.e. “asks” and “asked” were converted to “ask”.

We obtained optimum performance for bi-grams in our datasets D1, D2 and D3. In this study text documents were converted into computer recognizable format by vector space model where tf-idf
was used to weight terms. We used classifier implementation from rapid miner as classifier implementation was not the primary objective of this research.

We used accuracy, macro-average precision, macro-average recall, macro-average f-measure as our evaluation measures in this research.

In this study we employed Naïve Bayes Multinomial, K-nearest neighbor with cosine similarity and SVM with linear kernel because we achieved optimum results with these implementations.

We were not able to reproduce the results provided by Ana (Ana Cardoso Cachopo, 2007) in D2. We believe that it is because we handled datasets differently than Ana (Ana Cardoso Cachopo, 2007) i.e. we used 20% of D1, D2 and D3 as validation sets. We provided results obtained on our system for fair evaluations. The standard classification accuracies we obtained on our system for level-1 classifier in datasets D1, D2 and D3 are given in Table 1.

Table 1. Standard level-1 classifier accuracy

<table>
<thead>
<tr>
<th>Dataset</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>89.27%</td>
</tr>
<tr>
<td>D2</td>
<td>97.26%</td>
</tr>
<tr>
<td>D3</td>
<td>95.96%</td>
</tr>
</tbody>
</table>

The values of our evaluation measures obtained for level-2 classifiers in their standard (without instance filtration) form in dataset D1, D2 and D3 are shown in Table 2.

Table 2. Standard level-2 classifier effectiveness

<table>
<thead>
<tr>
<th>Level-2 Classifier</th>
<th>Dataset</th>
<th>Accuracy</th>
<th>Macro Avg. Precision</th>
<th>Macro Avg. Recall</th>
<th>Macro Avg. F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>KNN (K = 30)</td>
<td>D1</td>
<td>79.55%</td>
<td>80.47%</td>
<td>79.55%</td>
<td>79.22%</td>
</tr>
<tr>
<td>KNN (K = 35)</td>
<td>D2</td>
<td>93.66%</td>
<td>92.42%</td>
<td>86.36%</td>
<td>88.37%</td>
</tr>
<tr>
<td>KNN (K = 25)</td>
<td>D3</td>
<td>90.79%</td>
<td>90.91%</td>
<td>74.45%</td>
<td>78.07%</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>D1</td>
<td>88.10%</td>
<td>88.73%</td>
<td>88.10%</td>
<td>88.08%</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>D2</td>
<td>94.52%</td>
<td>80.46%</td>
<td>73.39%</td>
<td>75.74%</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>D3</td>
<td>87.86%</td>
<td>65.47%</td>
<td>42.93%</td>
<td>47.16%</td>
</tr>
</tbody>
</table>

We divided our system into two sub-systems S1 and S2. In S1 we used SVM as level-1 (filtering) classifier and KNN as level-2 classifier whereas S2 is created with SVM as level-1 (filtering) classifier and Naïve Bayes as level-2 classifier.

In all of the sub-systems, level-1 classifier was used to filter each test instance ‘x’ by ranking categories and reducing training dataset to top ranked ‘t’ categories and their instances whereas level-2 classifier is trained with filtered training dataset obtained from level-1 classifier to classify test instance ‘x’. The number of categories (with their instances) to be included in filtered training dataset was determined by threshold (measured by validation set). The effectiveness of S1 and S2 we obtained with instance filtration at their respective threshold (measured from validation sets) value are given in Table 3.
We found from experimental evaluations that reducing the number of categories for a test instance in a dataset has little or no impact on the categorization effectiveness of level-2 classifier. A level-2 classifier in its standard (without instance filtration) form misclassified some of the instances into a category that is relevant to their actual category, we termed such wrong predicted category for an instance as false relevant category.

In instance filtration this false relevant category was filtered out from the filtered dataset for some of the misclassified instances while filtering with level-1 classifier which is why level-2 classifiers achieved improved categorization effectiveness with filtered dataset. We also removed irrelevant categories for each test instance in order to keep the filtered training dataset as small as possible however it has minimal or no impact on the categorization effectiveness of level-2 classifier.

The goal of level-1 classifier is to remove irrelevant categories and false relevant category along with their instances from the filtered dataset. We referred to such instances as “correctly filtered instances”. The categories for each test instance are ranked by level-1 classifier and a filtered dataset is created from the top \( t \) categories.

We studied the impact of instance filtration on Naïve Bayes and KNN as level-2 classifiers to make sure that our approach stands true for multiple machine learning classifiers. Table 4 showed the number of misclassified instances correctly filtered by level-1 classifier at different thresholds and our systems behaviour at those thresholds in validation sets of D1, D2 and D3.

<table>
<thead>
<tr>
<th>Threshold (t)</th>
<th>Dataset</th>
<th>Level-1 Classifier Accuracy</th>
<th>% of misclassified instances correctly filtered</th>
<th>System Behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>KNN Naïve Bayes S1 S2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>D1</td>
<td>74.00%</td>
<td>4.5% 3.00% ++</td>
<td>++</td>
</tr>
<tr>
<td>3</td>
<td>D1</td>
<td>74.00%</td>
<td>3.00% 2.00% --</td>
<td>--</td>
</tr>
<tr>
<td>2</td>
<td>D2</td>
<td>94.97%</td>
<td>2.06% 3.59% ++</td>
<td>++</td>
</tr>
<tr>
<td>3</td>
<td>D2</td>
<td>94.97%</td>
<td>1.75% 2.97% --</td>
<td>++</td>
</tr>
<tr>
<td>2</td>
<td>D3</td>
<td>90.33%</td>
<td>2.47% 5.55% ++</td>
<td>++</td>
</tr>
<tr>
<td>3</td>
<td>D3</td>
<td>90.33%</td>
<td>1.85% 4.53% --</td>
<td>--</td>
</tr>
</tbody>
</table>

Table 4. Correctly filtered instances in validation sets
The “% of misclassified instances correctly filtered” column in above table presents the number of instances misclassified by level-2 classifiers in their standard form but correctly filtered by level-1 classifier in instance filtration. The “system behaviour” columns shows the change in effectiveness of the sub-systems at different threshold ‘t’ where ‘++’ means increase and ‘--’ represents decrease in the effectiveness.

We noticed continuous shrinkage in the effectiveness (all measures) of D1, D2 and D3 in our systems for threshold ‘t’ higher than the value obtained from the validation sets. The reason for this decline is that at higher threshold ‘t’ false relevant category appeared more often in top ‘t’ categories for most of the test instances and thereby many of them were misclassified by level-2 classifier. This behaviour confirmed that the threshold calculated from the validation sets was correct.

Results Comparison and Discussion

The table 5 presents improvements obtained in our approach over standard level-2 classifiers

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Classifier</th>
<th>Accuracy</th>
<th>Macro Avg. Precision</th>
<th>Macro Avg. Recall</th>
<th>Macro Avg. F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>KNN</td>
<td>2.9%</td>
<td>2.86%</td>
<td>2.9%</td>
<td>3.28%</td>
</tr>
<tr>
<td>D1</td>
<td>Naïve Bayes</td>
<td>0.65%</td>
<td>0.56%</td>
<td>0.65%</td>
<td>0.64%</td>
</tr>
<tr>
<td>D2</td>
<td>KNN</td>
<td>0.12%</td>
<td>0.39%</td>
<td>0.67%</td>
<td>0.72%</td>
</tr>
<tr>
<td>D2</td>
<td>Naïve Bayes</td>
<td>1.14%</td>
<td>10.55%</td>
<td>13.07%</td>
<td>12.157%</td>
</tr>
<tr>
<td>D3</td>
<td>KNN</td>
<td>0.97%</td>
<td>-2.27%</td>
<td>3.29%</td>
<td>3.56%</td>
</tr>
<tr>
<td>D3</td>
<td>Naïve Bayes</td>
<td>4.87%</td>
<td>27.59%</td>
<td>30.18%</td>
<td>31.68%</td>
</tr>
</tbody>
</table>

Table 5. Improvements over standard level-2 classifiers

In case of D1 the proposed KNN outperformed standard implementation of KNN. Our results (shown in Table 3) are also better than the results reported by Shi-Zhu Liu et.al (Shi-Zhu Liu and He-Ping Hu, 2007) for KNN.

Some categories in D1 are relevant to each other i.e. ‘alt.atheism’, ‘soc.religion.christian’, ‘talk.religion.misc’ etc. We found from experimental evaluations that standard KNN misclassified instances of these relevant categories into each other. We were able to remove ‘false relevant category’ for most of the misclassified instances that belong to these relevant categories. The KNN in our approach performed better with the filtered dataset obtained through instance filtration.

The Naïve Bayes in its standard form performed poor on small categories and misclassified many instances of the small categories. It is because distribution of samples was very skewed in D2 and D3 where few categories were very large whereas other were very small i.e. ‘aq’, ‘crude’, ‘money-fx’, ‘interest’, ‘trade’ and ‘earn’ contains 84.75% training documents of D3. The Naïve Bayes biases prediction in the favor of large classes in its standard form (Eibe Frank and Remco R. Bouckaert, 2006) in such datasets. We were able to filter out large categories for most of the instances that belong to small categories in D2 and D3 and thereby Naïve Bayes as level-2 classifier achieved improved categorization effectiveness in D2 and D3.

The categorization effectiveness of Naïve Bayes as level-2 classifier in D2 and D3 was increased more than D1 (shown in Table 4) whereas for KNN as level-2 classifier, increase in D1 was
higher than D2 and D3 (shown in Table 4). This behaviour suggested that instance filtration improve
the categorization effectiveness of level-2 classifier however the magnitude of improvement also
depends on dataset.

6 CONCLUSION AND FUTURE WORK

In this paper we purposed a novel approach to filter each test instance via its likelihood of occurring in categories. Naïve Bayes and SVM are used to rank categories and then filter training instance on behalf of threshold calculated for dataset. For evaluation 20newsgroup and reuters-21578 corpuses are used.

The experimental evaluations confirmed that instance filtration improves classifiers effectiveness however the magnitude of improvements depends on dataset. Our future work includes in studying the effect of our approach on other classifiers in machine learning and calculating threshold dynamically.

References

Ali Danesh, Behzad Moshiri and Omid Fatemi.2007. 'Improve Text Classification Accuracy based on Classifier Fusion Methods'. Information Fusion, 2007 10th International Conference,Quebec,Que.
Benhui Chen, Weifeng Gu and Jinglu Hu.2010. 'An improved multi-label classification based on label ranking and delicate boundary SVM'.IEEE World Congress on Computational Intelligence - CCIB, Barcelona, Spain.
Eibe Frank and Remco R. Bouckaert.2006. 'Naive Bayes for Text Classification with Unbalanced Classes'.Published in: Proceeding PKDD'06 Proceedings of the 10th European conference on Principle and Practice of Knowledge Discovery in Databases Pages 503-510 Springer-Verlag Berlin, Heidelberg.


UBIQUITOUS DECISION MAKING AND SUPPORT: A FRAMEWORK AND EVALUATION

Anika Rani, University of Auckland, aran066@aucklanduni.ac.nz
Shahper Vodanovich, Auckland University of Technology, svodanov@aut.ac.nz
David Sundaram, University of Auckland, d.sundaram@auckland.ac.nz

Abstract

While the field of decision making is rich in literature, and the work on ubiquitous technology is expanding, there is little understanding of how ubiquitous information systems (UIS) and ubiquitous devices (UD) support organisational decision making and ubiquitous decision support. This research presents a framework for understanding how ubiquitous decision support has emerged and is extending beyond the requirements that traditional DSS can satisfy. The proposed framework provides a foundation for ubiquitous decision support. The intention of the proposed framework is essentially to communicate how decision making can be applied to the advanced capabilities of ubiquitous computing in order to achieve decision making support through a UDSS. An evaluation of the framework through application indicates that while there is potential for further development of the framework, it is headed in the right direction and has encompassed critical elements of decision making and ubiquitous computing.

Keywords: Ubiquitous Computing, Decision Support System, Ubiquitous Technology, Ubiquitous Decision Support Systems

1 INTRODUCTION

Decision-making constitutes the core of organisational activity (Bonczek et al., 1981). Organisations encounter many decisions to do with various aspects of the business, and most often it is imperative to base such decisions on timely information. Decision making situations fall along a continuum that ranges from a simple sequence of activities to a collection of complex activities. Organisational decision makers may encounter simple and routine tasks, but quite often business decisions can be complex and require a great deal of analysis, where each alternative is attached with a set of consequences (Cyert et al., 1956).

Regardless of where on the continuum the decision falls, decision makers typically engage in some sort of process in order to arrive at a final decision. Decision support systems (DSS) have evolved and become recognised as an important tool in supporting such organisational decision making. In fact, at the emergence of DSS, organisations were considered to be on the verge of an era in computer-based information systems (Sprague Jr and Carlson, 1982). In a similar way, we seem to be entering another era in organisational decision making and support. As organisations are becoming more agile and flexible, mobile systems are providing support for decision making that extends beyond the traditional desktop.

Ubiquitous information systems were first introduced by Weiser (1993), who anticipated that these technologies would become indistinguishable from our everyday lives. Subsequently, ubiquitous technologies have become widely accepted and are continuing to evolve with an increasing body of research. Although the widespread acceptance and use of ubiquitous technologies can be observed in almost every aspect of our daily lives, it is also having a profound effect on organisations. While traditional information systems continue to facilitate decision makers, ubiquitous technology has become seamlessly integrated into our daily lives such that users are now empowered to make decisions whenever and wherever needed.
However, the advances in such technology in combination with their increased mobility remain a challenge for users, developers and researchers alike. The dimensions of different devices can become a constraint by limiting their display capabilities. Different systems also have varying functional capabilities, making it challenging for ubiquitous computing to provide consistent support, for all tasks, to all users. These challenges are particularly evident for tasks such as information seeking, where the results of a detailed information search can be difficult to perform effectively on a small screen device.

The motivation underlying this research was twofold:

1. While ubiquitous technologies are emerging as an important part of information systems research, ubiquitous Information Systems (UIS) remain relatively unexplored in relation to decision making and decision support. Although traditional information systems (TIS) remain a part of organisations, and will continue to do so, the rapid advances in mobile technology are bringing about fundamental changes in the way organisations interact with computers (Lyytinen and Yoo, 2002) and make business decisions.

2. To date, there appears to be limited information systems (IS) and industry research on how ubiquitous devices and systems can that support decision makers in performing tasks. In particular, there is a need to shift our focus from understanding traditional DSS (TDSS) to understanding how users engage in decision making on multiple devices, across multiple systems and across different contexts (Lyytinen and Yoo, 2002). This research lacuna and overall focus of this research is illustrated in Figure 1 below.

![Figure 1. Research lacuna in Ubiquitous Decision Making and Support](image)

This paper begins to address these motivations in more depth and propose a framework from which to understand this domain better. In this paper section 2 begins with an exploration of ubiquitous information systems and ubiquitous devices. This is followed by section 3 which has a focus on ubiquitous decision support followed by section 4 which looks at ubiquitous decision support systems (UDSS). Using these as a foundation we propose a framework for the understanding, design, and implementation of UDSS in section 5. We conclude by discussing how the UDSS framework could potentially support various decision making models (section 6) and an evaluation of the UDSS framework (section 7).

2 **Ubiquitous Technology**

The idea of ubiquity was initially introduced by Weiser (1993), who defined it as the idea that the most profound technologies are those that weave themselves into the fabric of everyday life and are used by people unconsciously to accomplish everyday tasks. The continuing miniaturisation of
computer systems and devices, combined with their increasing processing power, storage capacity and distribution (Lyytinen and Yoo, 2002), has made the vision of ubiquitous computing a reality. The rise of ubiquitous computing nearly a decade ago, has recently catapulted the world into becoming a ubiquitous society. Ubiquitous technology refers to ubiquitous information systems (UIS) and ubiquitous devices (UD). Figure 2 illustrates four dimensions that appear to be relevant in understanding UIS and its use. The context and system dimensions are explained in this section, where the activity and user dimensions have been explored by considering the decision maker and decision processes/tasks. Much of the research on ubiquitous computing has been focused towards the centre of the diagram: traditional information systems used by digital immigrants for professional purposes at the office (Vodanovich et al., 2010).

![Figure 2: Dimensions of UIS](image)

Figure 2. Dimensions of UIS

*Adapted from Vodanovich, Sundaram & Myers (2010)*

Figure 2 illustrates the focus of this research which is exploring UIS used by all users for professional purposes at the home, office and any other spaces.

The System dimension incorporates traditional and ubiquitous IS. Traditional IS have been used to help improve the efficiency and effectiveness of organisations (Vodanovich et al., 2010) and continue to be used today. Traditional IS are mostly restricted to desktop computers, and have much different characteristics to UIS. UIS on the other hand, refers to applications and systems which exist everywhere (Sørensen, 2005), and extend well beyond the desktop. Traditional IS and their users have been well researched, however there is limited research on the usage of UIS (Vodanovich et al., 2010).

The Context dimension is essentially “when” and “where” users engage with information systems (Vodanovich et al., 2010). Information systems are typically used in the home (personal space) and the office (professional space). However, as we move into becoming a ubiquitous society the context for using information systems has extended to other spaces outside the home and office. This can be attributed to the mobility of ubiquitous technology which allows for a seamless transition between the home, office and almost any other space.

### 2.1 Characteristics of Ubiquitous Information Systems

Ubiquitous information systems (UIS) are those systems that exist everywhere (Sørensen, 2005). “System” is used in a broad sense, and implies not only the application but the people, processes and information involved. Mobility (Roman and Campbell, 2002), information sharing, performance, responsiveness and focused purpose, among others, are the hallmarks of good UIS (Roman and Campbell, 2002).

**Mobility:** The same or similar UIS can be accessed through multiple devices. To exemplify mobility, we can consider Apple applications, majority of which are available across the iPhone and iPad. In the
past applications were restricted to a single device and users had to utilise a particular device in order to access a service/ application (Roman, 2001). Fortunately, in the case of UIS, users have increased accessibility to applications, whenever and wherever they are needed (Lyyntinen & Yoo, 2002). In light of the high level of mobility, many UIS are also adaptable to different devices.

**Information Sharing and Transfer:** UIS integrate different types of data from various sources and are capable of capturing real time information. Information sharing is enabled by and depends on the availability of communication and access to data. An example of this characteristic in practise is Apple’s iCloud, which is a data syncing service across multiple devices, from all locations. Data transfer is also in play when one application is used across multiple devices.

**Performance:** An important characteristic of UIS is its ability to serve a purpose well. Many applications today are providing users a rich and fast experience in performing a vast range of functions. There is an application for almost any task a user may want to perform. Performance is particularly critical in regards to mobility; mobility imposes some requirements on UIS where applications need to be capable of migrating between devices without affecting the consistency of the application and its performance (Roman, 2001).

**Responsiveness:** Increased speed and accessibility are two common benefits associated with UIS. It is important that a UIS provide acknowledgment upon a user performing an action.

**Focused Purpose:** UIS tend to be purpose-focused and specific to performing a certain task. Applications have clearly defined way of doing things, and aim to do achieve them with a minimum number of clicks (Salmre, 2005).

### 2.2 Characteristics of Ubiquitous Devices

UIS are characterised by their mobility, flexibility and integration, connectivity and support for applications (Tilvawala et al., 2011). We use a considerable number of devices today, and many of them are aligned with those proposed by Weiser (1993). He envisioned three forms of ubiquitous computers, in different sizes, suited to particular tasks; tabs, pads and boards. Today we see tabs as smartphones (iPhones, Blackberries), pads as tablets (iPad, the Kindle, laptops) and boards as larger touchscreens and interactive whiteboards. Organisations primarily utilise these three forms of devices in addition to traditional desktop computers.

**Mobile:** In order to support mobility, there is a continuous reduction in the size and shape of UD. The smaller and lighter a device, the more ubiquitous it is often considered (Tilvawala et al, 2011). Different tasks are often associated with different sizes and types of devices. For example, conducting a detailed information search is not as feasible on a mobile phone as it is on a laptop or desktop computer.

**Flexibility and Integration:** With an ever increasing number of mobile applications, UD must be able to cater for the requirements of various UIS/ applications.

**Connectivity:** Connection to the global infrastructure enables communication across multiple networks. The majority of UD are Wi-Fi enabled and/or have 3G data access, which supports many applications in performing tasks that dependent on information transfer.

**Input Capabilities:** Different devices offer different input methods. Touch screen display offers a rich environment for navigation and input (Salmre, 2005), whereas the traditional inputs such as keyboard and mouse continue to be very effective.

The basic foundation for any UD is an operating system (OS) or mobile OS that controls the device, and determines the UIS and types of applications that can be used with that particular device. Different types of UD vary in size dimensions from large screen displays on laptops/tablets to smaller displays on mobile and smartphones. A user typically owns or has access to multiple devices, containing numerous applications that span a variety of tasks. The availability of assorted UIS across many UD are accommodated for in a number of ways in an attempt to maintain consistency, reliability and performance.
The content delivery of a UIS is usually adjusted depending on the type of device it will be used on. The larger screen area of laptops and tablets allow for greater content to be displayed, while for mobile phones and smaller handheld devices content is usually reduced. Tailoring content and information presentation is necessary in order to provide multiples modes of access to a UIS, and overcoming the fact that each device varies in how effectively information can be presented. Desktop and laptops offer a rich platform for exploring information in a free-form way, while mobile devices and applications tend to be focused on enabling specific features and information with little navigation (Salmre, 2005).

UIS will generally perform according to the capabilities of the UD that it is being supported by. An application is likely to reach its optimal performance, if it has been configured according to the functionality of the particular device, on which it is being used. Many applications require a wireless or data connection to access information to help in performing tasks, which is dependent not only on the device but also on the context of use.

In some cases, UIS are not configured for the capabilities of different devices (Islam and Fayad, 2003). This gives rise to a number of implications for the user, which in turn affects user acceptance, especially where they are restricted in how they can perform tasks. While ubiquitous computing is better able to support users with their information needs, often there are more options and detail than can easily fit on the smaller screen (Billsus et al., 2002).

UIS must be tailored to the characteristics of various UD. UIS that are designed with the constraints of UD in mind will provide a better experience to users (Billsus et al., 2002). This will not, of course, eliminate all implications, because ultimately the effectiveness and usefulness of a UIS and UD depends on the task for which it is being used.

2.3 Complementary Perspectives

While ubiquitous technologies are an enabler of anytime, anywhere computing as initially proposed by Weiser (1991), there are several other terms which describe a similar phenomenon. Nomadic computing and experiential computing are two dominant ideas that are closely aligned with ubiquitous computing, but take slightly different stance view.

2.3.1 Experiential Computing

Everywhere we go we are surround by information or mobile technology. When we talk, listen to music or drive, we use computers that are not in their traditional form, but rather they are portable music players (MP3 players, iPods), global navigation systems (GPS), mobile phones with cameras and internet access (Blackberry, iPhone). Yoo (2010) has proposed several research directions in relation to computing in everyday life, otherwise known as experiential computing. Many of the experiences we go through while performing routine, everyday activities using involve using computing capabilities that are moving into everyday objects/artifacts. This IT phenomenon remains unexplored as researchers have mostly focused on organisational use of technology. Yoo (2010) suggests that there is a need to expand the IS research domain to embrace the impact of ubiquitous computing in everyday life. There are four dimensions which comprise the experiential computing framework as proposed by Yoo (2010). Artifacts refer to the actual ubiquitous systems and devices and actors refer to the users of those artifacts. Finally, space and time define when and where users engage with artifacts.

2.3.2 Nomadic Computing

Lyytinen & Yoo (2002) elaborate on ubiquitous computing where systems support computing and communication capabilities for nomads as they move from place to place. Handheld computing devices provide access to information and capabilities to perform tasks while constantly on the move. The essential features of nomadic computing are mobility, large infrastructure and digital convergence. Mobility is clearly evident in nomadic computing environments, where systems and services are available to the user whenever and where needed. Furthermore, services are now available through multiple devices, and can move across and between devices to suit the user (Yoo & Lyytinen, 2002). Digital convergence requires that devices can seamlessly share information across devices and
networks. Finally, due to the widespread adoption of such systems and devices, the global information infrastructure supports all services, communication and networking capabilities required by nomadic computing. Lyttinen & Yoo (2002) established eight different research directions for nomadic computing, in the hope that researchers would take the emergence of nomadic computing as an opportunity to learn about how the design and use of ubiquitous technology will take shape in organisations. However, there appears to be little research to show that the research directions introduced by Yoo (2002) have been explored.

3 Ubiquitous Decision Support

An understanding of how DSS evolved and a synthesis of the literature surrounding their definition and characteristics will provide a basis for designing systems with improved ubiquitous decision making support. An historical perspective of DSS shows it originated from two separate fields of research, computer-based information systems and operations research/management science (Silver, 1991). Moreover, Konsynski and Stohr (1992) believe that there is a need to remain focused on the original objectives of DSS with support for organisational decision makers (p. 26).

However, despite the success and acceptance of DSS, there is little consensus as to “What is a DSS”? Generally DSSs have been understood as computer-based information systems that support the decision making activities of people. Researchers and practitioners have proposed a variety of other definitions to extend this basic understanding, however with on-going developments in technology and emerging research, the variance in defining DSS does not appear to be diminishing (Stohr and Konsynski, 1992). New forms of computer based decision support are continually being developed, which is expanding the definition of DSS to a range of sub systems, where DSS is considered an umbrella term and systems are differentiated into expert support systems, executive support systems, business intelligence systems, analytic systems, etc.

The idea of ubiquitous computing as a mechanism for decision support has remained relatively unexplored, but it is emerging as a primary driver in changing how we make decisions (Kwon and Park, 1996). Its widespread implementation and acceptance has highlighted the need for growth in this area. Users working in the ubiquitous computing environment have exploited opportunities of real-time and “on the spot” decision making coupled with the potential to save time and increase productivity (Burstein et al., 2008, Carlsson et al., 2005).

While decision support has typically been associated with traditional desktop systems, ubiquitous computing provides extended support for decision making. Ubiquitous decision support (UDS) addresses the needs of the dynamic and competitive work environment, by providing decision makers with more functionality than what TDSS have. This improves the capability of the decision maker to generate more accurate and agile decisions, particularly in decision making environments where traditional desktop computers are not available (Kwon and Park, 1996). The deployment of such technology is mobilising the work practices of many disciplines, particularly in healthcare (Churchill and Munro, 2001).

Braudel (1981) introduced Braduel’s Rule, which states that if widespread use of a mobile device changes the structure of our everyday routines, it is considered valuable. In the case where a device has the potential to be useful but it has not impacted the way we do things, it is perceived as having little value. Ubiquitous computing has undoubtedly redefined the way in which we do many things. However, to completely understand the value and change introduced by ubiquitous computing for decision making processes, we must consider the UDS delivered by specific ubiquitous applications and devices (Carlsson et al., 2005).

It is interesting to observe that in reality the relationship between decision making and ubiquitous computing is quite strong, in that decision makers engage in ubiquitous decision support almost every day, without realisation. Despite this, IS research has not completely utilised this growing phenomenon as a research opportunity. More recently, there has been recognition to investigate how users are dealing with decision making in a ubiquitous context.
Research on UIS is continuously emerging, but not as rapidly as the development of mobile technology and its penetration in the market. A high degree of mobility is desirable for most decision making environments, particularly in today’s dynamic, fast-paced business environment. Ubiquitous technology undoubtedly fits the profile of a system that can provide support for the need of real time and on the spot decision making (Burstein et al., 2008). Since decision support has typically been associated with traditional information systems and desktop computers, the mobile nature of ubiquitous computing accommodates and addresses limitations of traditional decision support systems (Burstein et al., 2008).

While the benefits of ubiquitous computing are evident, such as increased accessibility to information and speed of communication, the research gap presents an opportunity to review UIS and UD in the context of decision making. It seems imperative to investigate whether the support they provide is effective and appropriate in fulfilling various processes of decision making since we have so readily accepted them into our lives.

4 UBIQUITOUS DECISION SUPPORT SYSTEMS

We propose the idea of a UDSS which seamlessly integrates the concepts of decision making and DSS with characteristics of ubiquitous computing. A ubiquitous decision support system (UDSS) essentially has the same objective as a TDSS, that is to support users in the decision making process. The prominent difference is that a UDSS is more flexible, interactive and geared towards providing support anytime, anywhere.

Understanding UDS requires knowing how people make decisions and the effect of UDSS on the underlying processes of decision making. With the variation in decision making processes, a UDSS is more likely to be useful if it supports multiple processes (Silver, 1991). An important consideration in providing decision making support is identifying the needs of decision makers and delivering DSS with capabilities that address those needs (Fick and Sprague, 1980). Ubiquitous technology has satisfied the need for faster information and mobility but the needs of decisions makers engaging in UDS have not been completely understood. In particular, what are the differences between users of TDSS and UDSS? The challenge remaining in UDS is to identify decision making phases and processes that can be better supported by UDSS. The mobility of UDSSs allows them to extend across numerous decision making environments. Decision makers are provided with decision support even as they move across environmental boundaries.

4.1 Characteristics of UDSS

Identifying a UDSS and differentiating it from traditional DSS require an understanding of its functional capabilities and characteristics. The basic characteristics of a UDSS are:

- Facilitating and supporting all phases of decision making for a variety of purposes and across a range of contexts.
- Intended to provide mobility, flexibility, interaction and a quick response to decision situations.

There is a tremendous need for an UDSS framework. There is a need to make it easier for decision makers to transition from using TIS to UIS and UD. A complementary approach is required to support all phases of the decision making. A portfolio approach will combine the traditional information system (TIS), TDSS, UIS and UD, to offer a complete solution.

5 DEVELOPMENT OF AN UDSS FRAMEWORK

In this section we discuss the development of the UDSS Framework, including all of its components. The intention of the proposed framework (Figure 3) is essentially to communicate how decision making can be supported by the advanced capabilities of ubiquitous computing in order to achieve decision making support through a UDSS. The framework considers the underlying conceptual principles of decision making in combination with the system related components. These components
are considerations for carrying out ubiquitous decision support. While undoubtedly, technological advancements reap benefits, they also introduce complexity into the decision support that was not present in TDSS.

The framework is composed of several components which relate to the key concepts already covered in earlier sections. Figure 3 not only illustrates the components that form the framework, but the way in which they are structured towards achieving ubiquitous decision support. The following sub sections briefly describe each of the components and their interrelationships with one another.

5.1 Decision maker/ stakeholder

The decision maker and/or stakeholder play a fundamental role in the proposed framework. Decision making is typically focused around a person, organisation or group of people making a decision. In this component we differentiate between a decision maker and stakeholder, because some decision situations simply involve a single decision maker, while others involve many stakeholders who are affected by the decision.

![Figure 3. A Framework for the Design of Ubiquitous Decision Support Systems](image)

5.2 Decision Work Flow

Decision work flows constitutes an important part of decision making, and defines the process by which a decision is made. We view decision work flow from two perspectives in this framework, micro and macro. A micro decision work flow is typically a small or set of small decisions that
decision makers typically do without even realising. Whereas a macro decision represents a holistic view of the decision situation, and may be composed of many smaller work flows that form one bigger, higher level decision.

This component of the framework also considers decision activity, namely single decision making (SDM) and group decision making (GDM). As the names suggest, SDM involves one decision maker, whereas GDM can involve multiple decision makers and/ or stakeholders. We consider decision activity because it affects the way in which a UDSS is used to support decision making. TDSS place particular emphasis on their ability to support decision makers with a variety of decision tasks, phases and processes (Sprague, 1980), and we believe a UDSS should also adhere to those guidelines.

5.3 Context
The context component has particular influence over a UDSS. The mobility of a UDSS enables it to be used across a variety of contexts, from the home, office and almost any other spaces in between. A decision maker or user of UDSS also has influence over the context component. Once again the mobility factor comes into play, which essentially gives the user control to use the UDSS anywhere they go. Another concept to consider is that of the same-space any-space, same-time any-time use of computing, which implies that a decision maker has support regardless of space and time. A TDSS cannot support any-space any-time decision making because they are typically restricted to a desktop and any-space same-time is supported to a certain extent.

5.4 Ubiquitous Device
Ubiquitous device is a system related component of the framework. The UD plays an important role in the overall framework because in ubiquitous decision support, the device has an influence on the type of support provided and its overall effectiveness. There is a linear relationship between traditional devices and ubiquitous devices with respect to the types of decisions they support.

5.5 UDSS
The UDSS component is central to the overall framework, which is why it is positioned between all the other components. There are two aspects to be considered in this part of the framework; selection and configuration of a UDSS. The selection of a UDSS determines the actual application to be used, for example a user may select stock market application for to make trading decisions. Subsequently, configuration of the UDSS will account for the context of use, the decision maker and other elements to adjust the UDSS for the designated task.

5.6 Ubiquitous Service Infrastructure
The final component in the framework is the overall infrastructure that supports the transmission of ubiquitous services and information across different devices and networks. Kawahara (2007) proposes a ubiquitous service infrastructure, which represents the overwhelming amount of information flowing through such system infrastructures. The lowest level (Level 1), which is the object level depicts the idea that ubiquitous services to gather information from around the world which is then pushed through several networks (Level 3). The information is then configured in terms of the context, the device and its platform (Level 4) after which it is delivered to the user as a service (Level 5).

5.7 Behaviour of Components
Interrelationships between the components indicate their behaviour and the effect they have on the overall framework. As mentioned in the previous section, the UDSS is the focal point of the framework, while the other components influence the operation and performance of the UDSS.

There are five main connections in the framework represented by the solid lines, explained as follows;
Decision maker / stakeholder and UDSS: The user is the primary component that utilises the system. Context and UDSS: The context influences how the UDSS responds to decision support in different spaces. Ubiquitous Device and Context: The UD in combination with the context, form the system related elements which enable the mobility in this DSS. Ubiquitous Service Infrastructure and Ubiquitous Device: The UD requires support to enable network connectivity, information transmission and delivery of ubiquitous services.

There are three interrelationships in the framework represented by the dotted lines between components, explained as follows:

Decision maker/ Stakeholder and Ubiquitous Device: This represents the decision maker’s access to multiple ubiquitous devices, which they may alternate between.

UDSS and Ubiquitous Device: The performance of the UDSS is dependent on the device on which it is used. This interrelationship is the primary concern of this research.

Decision Work Flow and UDSS: The type of decision work flow impacts the support provided by the UDSS and whether or not it is effective. Decision work flow feeds the decision problem into the DSS, which ultimately enables it to provide support for the problem.

6 UDSS AND ITS SUPPORT OF VARIOUS DECISION MAKING MODELS

Sprague (1982) states that a DSS should support a variety of decision making processes, but not be dependent on any one. He goes on to say that while Simon (1960)’s model is widely accepted, it represents only one model of how decisions are made. While there is no universally accepted model of the decision making process, it is important to ensure that a DSS can provide a decision maker with a set of capabilities to apply in a sequence that fits each person’s style of decision making (Sprague Jr and Carlson, 1982).

For this reason and to determine whether UDSS is process independent, the following examines the support provided by UDSS for various decision making models. It is important to note that because there is no existing research on UDSS and decision process, these beliefs are based on the literature review conducted earlier in the article. The physical size dimensions of a UDSS is an important factor in determining whether a certain phase of the decision making process is supported.

UDSS provides a certain extent of support for the sequential decision process, such as in Simon (1960)’s IDC model. A UDSS is most likely to be effective in supporting the Choice, because this phase of the decision making process requires little exploration, and is very task specific (Salmre, 2005). The intelligence phase and design phases are likely to be least supported by smaller UDSS, such as smartphones, but more supportive on laptops or tablets.

Moreover, UDSS is somewhat likely to provide support for the anarchical decision process, such as Cohen et al (1972)’s “Garbage Can” model. Decision making literature has been consistent in saying that TDSS do not provide enough support for unstructured decision. The mobility and variations in the applications available for UDSS may in fact improve the support for unstructured decisions because decision makers have increased accessibility to other characteristics of UIS. For example, a decision maker can use other applications on their UD that enable them to communicate with other stakeholders. A structured decision that cannot be completely supported by a UDSS, has potential to be supported by other UIS that may be available to the decision maker.

We believe that UDSS provides partial support for the iterative decision process, such as Mintzberg et al (1976)’s model. This idea results from combining our thoughts of the sequential and anarchical processes. The structured part of the iterative process may be supported to a certain extent by an UDSS, while the remaining unstructured has increased chances of support than a TDSS. Once again, size dimensions of the UDSS are important determinants in the effectiveness of the support provided.

Here we address the implications that may arise when a UDSS cannot support all the processes and phases of decision making. Increased accessibility to devices enables users to have a wider collection of systems which they can alternate between, known as a portfolio of DSS. Exercising such a portfolio
has become increasingly important because nowadays as decision makers can seamlessly transition between traditional and ubiquitous information systems (Vodanovich et al., 2010). Also, as noted earlier, users typically have access to multiple devices making it almost effortless to form and make use of a portfolio.

The portfolio of a decision maker who uses both a desktop and mobile device is hypothesised by Salmre (2005); a traditional desktop-bound DSS provides support for tasks that extend over long durations for exploratory activities, such as producing documents or browsing the web. When using a mobile device the user’s activities will tend to be shorter and specifically focused on a simple task such as choice selection.

Imagine a scenario where a user is on a desktop computer, seeking information on the share market, going back and forth through large, detailed tables and dynamic graphs. The desktop very effectively supports this intelligence phase. If the user has to leave the desktop for a particular reason, he may decide that he’ll continue his information search on his iPhone, where he also plans to also buy the shares he is interested in. Purchasing the shares will not be so much of a problem on his mobile, but if he plans to continue with examining the trends of different shares, this will prove difficult because the size constraint of a mobile does not actively support the display of such detailed information.

The transition from using a desktop to a mobile device represents the user’s portfolio. Had he switched to using a tablet that would also have been part of his portfolio, and supported the intelligence phase more effectively than the mobile phone.

The advantage in effectively managing and using a portfolio of DSS is being able to achieve complete support for decision making activities and tasks by utilising the devices and systems available. Often the shortcomings of one application or device can be fulfilled by another, providing a complementary approach to decision support.

7 Evaluation of the UDSS Framework

In this section we first introduce a typology that serves as the basis for identifying and selecting the UIS to be used in the evaluation of the proposed framework; information provision and exchange. UIS based on information provision are simply applications that are focused towards information delivery, and facilitating with the analysis of that information. However, UIS based on information exchange facilitate with the transmission of information. Information can be transferred between different devices, as well as sent and received through connection to the wider network, most often through the internet. To effectively support ubiquitous decision making and support we need both information provision and exchange.

The UIS selected for evaluation of the framework is Roambi (Figures 4 and 5). Roambi aligns well with the information provision category because if allows for easy access to business data, anytime, anywhere. The Roambi Visualiser App allows a user to view and analyse important business data. Furthermore, Roambi also aligns well with information provision, enabling a user to manage, interact with, and share visualisations of their business data.

We evaluate Roambi with respect to each component in the UDSS framework as follows;

7.1 Decision maker/ Stakeholder

The profile of the decision maker using Roambi is likely be a business person who has access to organisational data. Since Roambi offers the ability to share analysed information, the app has potential to involve other stakeholders.
7.2 Decision Work Flow

Roambi appears most likely to support a micro decision work flow. Roambi supports the intelligence and design phases of decision making, but it does not allow a person to actually make a choice and implement a decision. A decision maker is likely to utilise Roambi to gather information and intelligence, and subsequently make a choice in another way. In addition, Roambi provides support for single decision making (SDM), where a single person can gather and analyse information. There are no particular features that would enable or enhance support for group decision making (GDM).
7.3 **Context**
Roambi has been designed to give the decision maker complete mobility. As a result, Roambi can be used in a personal, professional or other space. It is also capable of enabling support at any time, as long as the user can provide data.

7.4 **Ubiquitous Device**
Roambi is available in a tablet and phone configured UIS. It appeared that Roambi is consistent with visualisations across both devices but as expected shows a more detailed view on the larger screen. The iPad also provides more filters for the user to further refine the information, which is likely to be more limited on the iPhone. Both devices deliver reasonably the same content, therefore it seems that they would both support-structured decisions that are based solely on the analysis of business data.

7.5 **UDSS**
Users who want support in analyzing, managing and interpreting business data should select Roambi as a UDSS. Roambi would serve as an effective visualisation tool for smaller businesses that need to gain a snapshot of their business for information purposes, rather than for implementing any decisions.

Roambi is an effective UDSS in terms of providing configuration capabilities. The application can be configured completely for the user, using their personal business data, inputs.

7.6 **Ubiquitous Service Infrastructure**
Like every ubiquitous service, Roambi needs to be supported by the wider infrastructure. It offers certain functionalities that require a connection to the network, such as acquiring the data of the user which may be stored in another system and sharing of visualisations which requires connecting to the wider network to send such information.

8 **CONCLUSION AND RESEARCH CONTRIBUTION**
Decision making and ubiquitous computing are increasingly being adopted and implemented by decision makers, who do not seem to realise that they are engaging in a new phenomenon. Even more surprising is the lack of focus on this area within the Information Systems discipline. Researchers have recognised the potential of ubiquity and have continued to develop the concept since its introduction. However, it seems that ubiquitous computing is developing faster than research can follow, particularly in regards to decision support. Ubiquitous decision support has recently been recognised as an unexplored area, and this research has made an initial attempt to capture the opportunity provided by this research gap.

The main contribution of this research is the development of a ubiquitous decision support systems (UDSS) framework which outlines the key considerations required for ubiquitous decision support. Elements of decision making and ubiquitous computing were extracted and structured to support a UDSS. This framework offers implications for researchers and practitioners alike; it is hoped that this framework will serve as a benchmark for future research on ubiquitous decision making support. It is also developed with the intention of providing a guideline to users to understand the elements of ubiquitous decision support, and understand that each of the components plays a critical role in establish completely effective ubiquitous decision support.

Another contribution of this research, which led to the development of the framework, was the unexplored area between decision making and ubiquitous computing. While the proposed framework seeks to integrate these two areas, it is only one of the many proposals that can arrive as a result of identifying this gap.
9 REFERENCES
Cloud Computing

Chairs: Associate Prof. Ricardo Jimenes Peris
Table of Contents

Track – Cloud Computing

BUILDING ROBUST CLOUD ENABLED E-HEALTH APPLICATIONS

George Pittas: Department of Digital Systems, University of Piraeus, Greece, gpittas@unipi.gr
Marinos Themistocleous: Department of Digital Systems University of Piraeus, Greece, mthemist@unipi.gr

Abstract

The rapid increase of mobile devices usage offers innovative ways for users to access medical data. Advanced e-health applications should be secure, available, scalable and user friendly. From developer’s perspective they should be easy to develop, deploy, test and maintain. The increasing volume of health data and the heterogenous sources create the need to adopt an international standard to facilitate the integration. In this paper, we address concerns about the development of e-health applications, and we analyze the need to decouple the backend from the frontend. We demonstrate a way to do so by using REST API (Representational State Transfer Application Programming Interface) serving through cloud alongside HL7 (Health Level Seven) standard and JavaScript framework for multiple platform compatible user interface.

Keywords: E-health applications, REST API, HL7, Cloud Computing, Web Services

1 INTRODUCTION

The introduction of Cloud-Computing has been some of the biggest changes impacting not only the IT sector but also healthcare. The impact of Cloud-Computing on healthcare can be characterized as a positive change as it provides integration at a manageable cost.

Doctors, hospitals, and patients require fast access to medical data, which is not possible in legacy systems (not cloud based). Additionally, the multiple heterogeneous data sources require an efficient way for medical data transmission across various devices and geographical locations in a fast, efficient and cost effective way without limitations (e.g. errors, high cost). Recently, cloud technology has started replacing legacy healthcare systems and offers easier, undisrupted and faster access to medical data. Literature depicts that Cloud-Computing offers significant benefits to the healthcare sector with its business (e.g. pay-as-you-go) model and integration capability (Kuo, 2011, IBM, 2015). There are examples of global IT companies like Microsoft, Oracle, Amazon that have already invested in more powerful, always available and cost-efficient cloud platforms, providing new offerings for their e-health services, such as Microsoft’s HealthVault, Oracle’s Exalogic Elastic Cloud, and Amazon Web Services (AWS) (Zhang et al., 2010).

Apart from the Cloud Computing adoption, there is an emerging area of mobile devices. The latest generation of smartphones and tablets combine both computing and communication features in a single device, allowing easy access and use at the point of care (Mosa et al., 2012). Some of the advanced features that new mobile device models offer are web searching, global positioning systems (GPS), high-quality cameras, and sound recorders (Boulos et al., 2011). Due to the aforementioned capabilities of mobile devices there is an increase on e-health applications development. According to a recent report e-health applications that are published on the two leading platforms, iOS and Android, has more than doubled in only 2.5 years to reach more than 100,000 apps (e.g. 1st quarter of 2014) with a market revenue of USD 2.4bn in 2013 and projections to grow to USD 26 billions by the end of 2017 (Research2guidance, 2014). Sarasohn-Kahn (2010), identified that a major mobile application vendor provides 5,805 health, medical and fitness applications with 73% of them used by patients and 27% by healthcare professionals (Sarasohn-Kahn, 2010).
Many of these applications are based on Service Oriented Architecture (SOA) as e-health services can be easily delivered to both desktop and mobile computer devices using JavaScript frameworks, HyperText Markup Language (HTML) and server side technologies capable to handle the requests of the clients. Based on the SOA paradigm e-health services can be exposed and run over cloud (in the form of SaaS) (Poulymenopoulou et al., 2012).

In this paper, we explain the challenges when creating new e-health applications that can run on multiple platforms and devices delivered through cloud. We describe the need for decoupling the backend from the user interface. Also we analyse how to build a secure REST API HIPAA (Health Insurance Portability and Accountability Act) compliant and a multiplatform compatible user interface.

2 BACKGROUND THEORY

2.1 REST API

REST is an architectural style, and an approach to communications that is often used in the development of Web services (Fielding and Taylor, 2002). According to IBM, REST defines a set of architectural principles to design Web services that focus on a system's resources, including how resource states are addressed and transferred over HTTP by a wide range of clients written in different languages. The last few years, REST has had such a large impact on the Web that it has mostly displaced SOAP- and WSDL-based interface design as a simpler alternative (IBM, 2015). A REST Web service follows four basic design principles described in the following paragraphs.

First REST design principle establishes mapping between create, read, update, and delete (CRUD) operations and HTTP methods (zur Muehlen et al., 2005). According to this mapping:

- POST is used to create a resource on the server,
- GET is used to retrieve a resource,
- PUT is used to change the state of a resource or to update it,
- DELETE is used to remove or delete a resource,
- PATCH is used to update partial content of a resource,
- OPTIONS is used to get information about the communication options for the request.

A second characteristic is that all REST interactions are stateless and they need to scale to meet increasingly high performance demands. That is, a REST Web service application (or client) includes all of the information necessary for the server to understand the request, independent of any requests that may have preceded it. This removes any need for the server-side components to retain application state between requests, thus reducing consumption of physical resources and improving scalability (Fielding and Taylor, 2002).

A third characteristic of REST Web service is the URIs. REST Web service URIs should be easy to guess and to use. One way to achieve this level of usability is to define directory structure-like URIs (zur Muehlen et al., 2005). This type of URI is hierarchical, rooted at a single path, and branching from it are subpaths that expose the service's main areas. For example, if we need to get the exams of a patient we use HTTP method GET like this:

GET https://api.example.com/patient/{patientid}/exams

The root /patient has /exams beneath it. Within this structure, it's easy to pull up the patient’s exams just by typing the patient’s id after /patient/. URIs should also be static so that when the resource changes or the implementation of the service changes, the link stays the same.
The last set of constraints that goes into a REST Web service design has to do with the format of the data that the application and service exchange in the request/response payload or in the HTTP body. REST can support XML and JSON content types. Both of them have strong points. For instance, JSON is lightweight, thus offers better performance compared to the complexity of XML (zur Muehlen et al., 2005).

2.2 Model View Controller (MVC)

Model View Controller or MVC as it is popularly called, is a software design pattern for developing web applications. It is a widely adopted pattern, being supported by many different programming languages and implementation frameworks. MVC is popular because it isolates the application logic from the user interface layer and supports separation of concerns (Reenskaug and Coplien, 2009). The three components of MVC are the following ones:

- **Model** - It is the lowest level of the pattern responsible for maintaining data. It responds to the request from view and to the instructions from controller to update itself.
- **View** - It is responsible for displaying all or a portion of the data to the user, in other words it handles the user interface (UI)
- **Controller** - It controls the interactions between the Model and the View.

More specific, the controller receives all requests for the application and then works with the model to prepare any data needed by the view. The view then uses the data prepared by the controller to generate a final presentable response. The MVC abstraction can be graphically represented as follows at figure 1.
REST API FOR CLOUD ENABLED E-HEALTH APPLICATIONS

As stated at introduction there is an emerging increase of mobile usage. Today there are many different devices (e.g., Desktops, Laptops, Tablets and Smartphones) each one has different screen size and operating system. The new e-health applications should be compatible with all of them, capable to empower doctors and patients to access medical data. In order to achieve a cost effective and efficient development we need to adopt appropriate technologies and standards.

First of all, Cloud Computing offers virtualization which allows an abstract representation of logical and physical resources including servers, storage devices, networks and software. The basic idea behind it is to pool all physical resources and their management as a whole, meeting the individual demands from these shared resources (Lupse et al., 2012). Using Virtualization has many benefits including: (a) easier replication and cloning a VM than physical server, (b) lower down time in case of failure, (c) lower power consumption and saving resources by running multiple Virtual Machines (VM) within the same physical server (Chang, 2013).

Another requirement that we need to take under consideration is the interoperability which is the ability of two or more systems or components to exchange data and use the information that has been exchanged (Lupse et al., 2012). The Integrating the Healthcare Enterprise (IHE) organization provides standards (e.g. HL7, DICOM, etc.) to enhance the interoperability and information sharing (IHE, 2012). A way to enable those standards is the REST API. So, there is the need to decouple backend and the frontend. In that way, we loose the dependence to the implementation providing easy integration of applications. Another important factor to achieve interoperability is the adoption of HL7 CDA (Clinical Document Architecture). The HL7 CDA standard is a document markup standard that specifies the structure and semantics of “clinical documents” for the purpose of data exchange (Lupse et al., 2012).

Another important feature that should be taken under consideration is the scalability. Due to the type (e.g. numerous different users and real time data streaming) of the e-health application we need to ensure its availability regardless of the number of the concurrent connections. REST along with Cloud Computing can provide solution for scalability. REST takes advantage of JavaScript Object Notation (JSON) which offers better performance compared to the complexity of XML. With REST there is no need to create complex request to the server. For example, when the doctor needs to retrieve the patient’s medical data, we need to create a GET request, whereas when he/she needs to add medical data (e.g., exams results), we use a POST method and the sensitive information passed in the body of the request as a JSON object over an SSL connection.

As HIPAA states the security of e-health applications is crucial, so we need to adopt a secure user authentication model to ensure protection of information against unauthorized access. As different users access the system, sensitive information may be provided only to authorized users (Narayanan and Giine, 2011). To meet the legislation standards we need to create a secure mechanism for user authorization. REST API provides us an easy, secure and effective way to authenticate users. The concept behind that is simple. The user to access the patients’ health records should authenticate. They enter their credentials to access the patients’ health records. These credentials are sent to the server via the API. In case of successful authentication, the API returns as a response a security token. The returned token, provides to user access to the health data based on his/her authorization. That token must be resent to the API for each subsequent request. The server identifies the user and his/her permissions based on that token and provides conditionally access to resources. To be keener, as it is depicted at figure 2, a doctor can have access to the medical data of the patients as long as the patients have given consent to
the specific doctor. Moreover, we need to take under consideration the fact that all the data at REST API are being transmitted through HTTP protocol. To secure the data transmission and prevent eye dropping we need to send the request over Secure Sockets Layer (SSL).

![Diagram](image)

*Figure 2. Token based Authentication*

4 **FRONTEND OF E-HEALTH APPLICATIONS USING MVC**

At the previous chapter, we talked about building REST API for e-health applications and its advantages. In the current chapter, we analyze ways to build user interfaces that can be interact with the REST API. One of the important tasks when developing user interfaces is to insure that the system will provide the best user experience on all the devices. This may provide higher adoption rates of the applications.

The recent advances in client-side technologies have changed the way that we develop web applications. Android, IOS and purely browser-based clients are powerful computing platforms. Their success has led to a reevaluation of ‘classic’ web applications. We notice that more and more application are trying to loose the bond between the client and the server, that led to an emergent increase in usage of MVC frameworks that are based on JavaScript. Those frameworks enable developers to build powerful and compelling UIs. By implementing a majority of the MVC functionality at client-side, it reduces the complexity of the server and results in more applications that are maintainable and multiple platform compatible. A recent and significant change in web application frameworks has been the shift from Model-View-Controller (MVC) on the server to the MVC on the client. Advancements in client-side
technology are driving this change (Google, 2015b). Android, iOS and browser-based clients are fully compatible with JavaScript MVC frameworks. Nowadays, there exist many JavaScript MVC frameworks (Osmani, 2012) such as AngularJS, Backbone, Ember.js and Knockout. The most promising one is AngularJS. It is developed by Google and has a large developer community (Google, 2015a, Tilde, 2015). It supports composite views, data-binding between model and view, and custom vocabulary definition by directives. The following Figure 4 shows an overview of high level AngularJS’s architecture. AngularJS has a template system for reusing the same types of views. With directives, developers can declaratively specify data-bindings between views and scopes and set up controller association to views. Scopes manage the state of views, expose data form models to the corresponding view, and operate models according to the message form the view. We can also define custom directives as element names, CSS classes, or element attributes to reuse user-defined functionalities or behaviors in a declarative way (Fujima, 2013).

The aforementioned features of AngularJS, provides us the ability to achieve easy development, multiplatform compatible and user friendly user interface. It coorparates with latest web technologies such as HTML5 and CSS3. Thus, it allows us to create responsive user interfaces which is the optimal viewing experience (e.g. easy reading and navigation with a minimum of resizing, panning, and scrolling) across a wide range of devices (Marcotte, 2011). There are various AngularJS compatible frameworks that support responsive design but the most promising one and widely used is Bootstrap (Bootstrap, 2015). Bootstrap has been developed by Twitter and it has a large support community.

5 EXAMPLE OF A REAL E-HEALTH APPLICATION

In this section we present an example of the application of the aforementioned techniques in a real e-health project. The project is called Providing Integrated eHealth Services for Personalized Medicine utilizing Cloud Infrastructure (PINCLOUD) and it seeks to integrate different application components, leading to the provision of an end-to-end personalized disease monitoring and medical data service “anytime, anywhere”, which ensures independent living regardless of age (Lab of Medical Informatics, 2014).

PINCLOUD has various components such as: PHR platform, e-Prescribing, e-Referral and homecare. As we stated at previous chapter PINCLOUD leverages the REST API for integrating those different components. This gives us the ability to run the application in multiple clients with different operating systems. The following Figure 4 depicts, how the clients consuming the web services serving through cloud.
One provided service of PINCLOUD is the E-Referral. After a patient’s visit, the doctor needs to retrieve the exams of the patient using the PINCLOUD application. Yet, before the doctor gain access to the E-Referral service, he/she needs to be authenticated using his/her credentials through the login page of the PINCLOUD. For the scope of this example, we will not focus on the business process but what is happening behind the scenes (e.g. background). The client (e.g. desktop, tablet, smartphone) sends the credentials to the API server and is waiting for the response. In case of successful authentication the server responds returning the user’s specified token (Endpoint URI 1). The client redirects the user to the main screen of the application where he/she selects the E-referral and locates the patient using the patient’s Social Security Number (S.S.N.) which is unique for each patient. To retrieve the profile of the patient, client sends another request using HTTP method GET (Endpoint URI 2). Before the doctor can execute the exam, it is required by the system to retrieve the pending exams. To retrieve those exams a request is sent to the server containing the security token (Endpoint URI 3). It insures that access to the exams and in general to all the medical data only the authorized users can have. After doctor finishes the examination he/she updates the system that it is executed (Endpoint URI 4). That action is being translated to a PUT request to the server as it is depicted in Table 1. The whole communication between the client and the server takes place over a SSL secure connection to make sure compliance with HIPAA.

<table>
<thead>
<tr>
<th>Endpoint URI</th>
<th>Operation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. POST user/login</td>
<td>Authentication for user</td>
</tr>
<tr>
<td>2. GET patient/{ssn}</td>
<td>Find the patient based on S.S.N.</td>
</tr>
<tr>
<td>3. GET patient/{patient-id}/labexams</td>
<td>Lab exams of a patient</td>
</tr>
<tr>
<td>4. PUT patient/{patient-id}/labexams/{labexams-id}/execute</td>
<td>Execution of the doctor specified lab exam</td>
</tr>
</tbody>
</table>

*Table 1. Consuming Web Services serving through Cloud*
6 TESTING REST API

The REST API is very easy to test using HTTP requests. To send those HTTP requests we can use cURL from bash or to use a web browser. There are free tools available as browser’s add-ons for API testing such as postman for Chrome.

During development of PINCLOUD we were able to prototype the API’s endpoints during the first days and we imported all the endpoints into the postman and we were able to easily test the created operations. Furthermore grace to AngularJS MVC framework that we have used, the configuration of the host, port and base URL was easy on switching between the development and the production server. Furthermore, the testing with AngularJS was as easy as possible since it has been designed in that way. It even comes with an end-to-end and unit test runner setup.

7 CONCLUSION

In this paper, we have presented the advantages of using a REST architecture for designing secure, available, scalable, multiplatform compatible and user friendly e-health applications. We were able to create a working API prototype and develop a full functioning set of sophisticated health record web services accessible by our all the components of PINCLOUD. We gathered user feedback from a large group of users (doctors and patients) and statistics about usage that were very valuable to improve the performance of our health record REST API and the user experience of the application. Also the audit of server logs for the REST API and the use of data mining tools will be valuable to see performance bottlenecks and usage trends. As far the user interface is concerned, we have run tests in various platforms and screen sizes in order to eliminate compatibility issues (e.g. not optimal display).

Acknowledgement

The work reported here has been carried out in the framework of national project Providing Integrated eHealth Services for Personalized Medicine utilizing Cloud Infrastructure (Pincloud), led by the University of Piraeus, conducted in the context of the National Strategic Reference Framework NSRF 2007-2013, Cooperation 2011 and co-funded by the European Commission.

References

BOULOS, M. N., WHEELER, S., TAVARES, C. & JONES, R. 2011. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomedical engineering online, 10, 24.


RESEARCH2GUIDANCE 2014. *m-Health Application Developer Economics 2014: The state of the art of m-health app publishing.*


DYNAMIC PREDICTION ON AN E-HEALTH APPLICATION BASED ON VARYING SCALABILITY PARAMETERS

Avraam Giogourtsooglou: Department of Digital Systems University of Piraeus, Greece agiog@unipi.gr

Marinos Themistocleous: Department of Digital Systems University of Piraeus, Greece mthemist@unipi.gr

Dimosthenis Kyriazis: Department of Digital Systems University of Piraeus, Greece dimos@unipi.gr

Abstract

This paper focuses on the creation of a model for dynamic resource prediction. This is achieved by analysing the behaviour of a service, which is installed in cloud computing infrastructures. As a result, we aim to create a service performance prediction model. Based on this model, we expect that the performance of a service can be predicted, when the values of parameters that influence it are changing. Thus, specifying the desired performance of a service and the changes depending on the parameters, one can increase or decrease the resources used. We used a representative case study of an e-health application that has been developed in PINCLOUD project.

Keywords: prediction, cloud computing, performance, cloud services, prediction model, resource elasticity.

1 INTRODUCTION

Cloud Computing infrastructures provide us a dynamic resource scaling. In this way, using the appropriate tools at platform level, it can provide the necessary resources depending on the requirements for each service even if they change from time to time. Cloud provide us tools to manage the resource needs of every service, but the preparation that is needed in order to offer these resources dynamically is a time consuming operation and can even take hours to complete. The permanent reservation of resources so as they are available when needed is not a solution, however, as it does not allow for optimal use of the total resources. Thus a non-negligible need is created, for a mechanism that will predict the service performance, based on effecting parameters (e.g. number of parallel requests to the service) and forecast the resource needs. These predicted needs would then be used to start preparing them the needed resources at the right time, so as to be ready when needed. Similarly, in case the prediction reports reduction on needed resources, then the procedure of gradual release would start at the right time.

This paper focuses on the creation of such a model for dynamic resource prediction for a specific e-health cloud application. This is achieved by analyzing the behavior of a service,
which is installed in cloud computing infrastructures, in order to create a prediction model of its performance. Based on this model, the aim is to predict performance of the service when the values of parameters that influence it are changing. Thus, specifying the desired performance of the service and the changes depending on the parameters, one can increase or decrease the used resources.

In previous studies (Giogourtsoglou et al., 2014) we have worked on the creation of a similar prediction model, analyzing the service’s behavior of the encyclopedic website Wikipedia.org. The selection of that service was based on specific criteria sets. These criteria relate to: a) the service must use cloud computing infrastructure, b) the service must mostly use the resources for serving the request, and not for its lifecycle, c) the service must widely be used by a large part of the population, to have representative data, and d) the requests for Europe must be served from the same infrastructure. More details about related work presented in Section 2.

In this case, we focus on a e-health application, that was developed in PINCLOUD project. This application provides integrated e-health services for personalized medicine like e-PHR, e-prescription and e-referral. PINCLOUD seeks to integrate different application components, leading to the provision of an end-to-end personalized disease monitoring and medical data service “anytime, anywhere”, which ensures independent living regardless of age.

The prediction model that we focus to create, it will predict the performance of the application based on effecting parameters. The parameter that determines the performance of this service is the response time of every request, and the critical effecting parameters are: a) the number of parallel requests from users, b) the location from where the requests are executed, and c) the time at which the requests are performed to the service.

To develop a performance prediction model, we have to analyze the behavior of the aforementioned service. To create such a model, it is necessary to collect experimental data of the service, about the critical effecting parameters. After the data collection, it is necessary to analyze them using statistical tools (like R language). The aim of this process is to evaluate the data collection, and to decrease the volume of data for the next processes. Next step of the procedure, to create the prediction model, is to determine the shape of the data with mathematical terms. Those terms will help us to decide the mathematical models that we will use in development of prediction model. Last step of the procedure includes the development of the model, which we use the Eureqa tool, and the its evaluation.

In section 1, of this paper, is introduced the purpose of the study, which described before. Literature Review including background theory about the key terms of Cloud Computing, and related work are presented in section 2. The next section the proposed solution, about the creation of a prediction model for the service that is studied. In section 4, are presented the future plans of this study, including the creation and the evaluation of the model and also, suggesting some ideas to make it more accurate. Finally, in section 5, the conclusion about the study that has been done is presented.
2 LITERATURE REVIEW

In this section will be presented the background theory and the literature review analysis of the related work. Some the key terms will be presented such as a) definition and characteristics of Cloud Computing, b) elasticity in Cloud Computing, and c) enabling technologies behind Cloud Computing. Related work includes other studies and a similar study that have been done about prediction model creation.

2.1 Background Theory

2.1.1 Definition and Characteristics of Cloud Computing

Cloud Computing refers to both, applications and hardware infrastructures that delivered as services over the internet and the hardware systems in the data centers, which provide those services. The National Institute of Standards and Technology (NIST) recommends the following definition of Cloud Computing:

“Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models.” (Mell and Grance, 2009)

Also, NIST definition (Mell and Grance, 2009) describes the essential characteristics of Cloud Computing, which are on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service. There are two kinds of models about Cloud Computing. The first one related with the services that are offered to consumers. Those models are: a) Software as a Service model (SaaS), in which the consumer has the capability to use the provider’s applications running on a cloud infrastructure, b) Platform as a Service model (PaaS), in which provided to the consumer the capability to deploy onto cloud infrastructure his applications, and c) Infrastructure as a Service model (IaaS), in which provided to the consumer the capability to provision processing, storage, networks, and other fundamental computing resources, where he is able to deploy and run software. The deployment models describe the organizational structure and the provisioning location of cloud infrastructure. Four deployment models are usually distinguished, namely public, private, community and hybrid cloud service usage. In public model, the cloud infrastructure is provisioned for open use by the general public, on the contrary in private model is provisioned for exclusive use by a single organization comprising multiple consumers. In community model is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns, and the hybrid model which is a composition of two or more distinct cloud infrastructures (private, community, or public).
2.1.2 Elasticity in Cloud Computing

The concept of elasticity has been transferred from physics and economic theories (in which has originally been defined (Chiang and Wainwright, 2005)), to the context of cloud computing and is commonly considered as one of the central attributes of the cloud paradigm. Cloud providers use the term of elasticity for marketing purposes, even in the naming of specific products or services. There are many definitions about what is elasticity, but it is very difficult to quantify and compare elastic behavior. The paper (Herbst et al., 2013), is trying to make a composite definition by presenting five well known definitions of elasticity(Inc, 2014) (NIST, IBM, ODCA, Rich Wolski and Reuven Cohen definitions). So the definition of elasticity, as it presents is:

"Elasticity is the degree to which a system is able to adapt to workload changes by provisioning and deprovisioning resources in an autonomic manner, such that at each point in time the available resources match the current demand as closely as possible."

(Herbst et al., 2013)

2.1.3 Enabling Technologies behind Cloud Computing

There are enabling technologies that contribute to Cloud Computing(Wang et al., 2010). The most significant of them are presenting here:

- **Virtualization technology**: Virtualization technologies partition hardware and thus provide flexible and scalable computing platforms. Except computers’ hardware, virtualization includes operating systems and network resources. Virtual machine techniques provide virtualized computer infrastructure on demand. Virtualization techniques are the bases of the Cloud Computing, since they render flexible and scalable hardware services.

- **Orchestration of service flow and workflow and workflow - choreography**: Cloud Computing provides an amount of services, which consumers can combine them and compose some other services. Service orchestration describes the automated arrangement and management of services that are deployed on different locations. Services exchange messages between them and choreography describes the workflow.

- **Web service and Service-oriented architecture (SOA)**: Services on Cloud Computing are commonly exposed as Web services, follow standards as WSDL (Chinnici et al., 2007, Cerami, 2002), UDDI(Cerami, 2002) and SOAP(Cerami, 2002, Chinnici et al., 2007). Cloud Computing uses a Service-oriented architecture to organize, manage and orchestrate the web services. With this architecture, all web services offered on various distributed platforms can be further accessed through a single entry point from anywhere in the Internet.
2.2 Related Work

We focused on studies about provision of resources from applications that are installed on cloud computer’s infrastructure. These studies are usually performed on specific structures of the cloud computing, defining applications’ parameters that influence their performance. Some other studies focus on models to predict the behavior of the applications. This section presents the studies that have been taken in consideration to analyze the service’s behavior and create a performance prediction model.

An interesting study has been done in paper (Kousiouris et al., 2011), in which the aim was to study and predict the effect of a number of critical parameters on the performance of virtual machines. These parameters include allocation percentages, real-time scheduling decisions and co-placement of VMs when these are deployed concurrently on the same physical node, as dictated by the server consolidation trend and the recent advances in the Cloud Computing systems. To measure the effect of parameters, they use the score of specific benchmarks running inside the VMs. So they produced a Generic Algorithm-optimized artificial neural network that can be used to model, quantify and accurately predict the performance of the applications for a given configuration.

Zhu and Agrawal (2010), focuses on an automated and dynamic resource allocation problem associated with the execution of adaptive applications in a cloud environment. To solve that problem is presented the design, implementation, and evaluation of a framework that can support adaptive applications in a cloud computing environment. The key component of the framework is a dynamic resource provisioning algorithm that created and integrated into the proposed framework.

Islam et al. (2012) presented the develop of a prediction-based resource measurement and provisioning strategies using Neural Network and Linear Regression to satisfy upcoming resource demands. While the initialization of a new virtual instance in a cloud is not instantaneous, cause of delay in the hardware resource allocation, the solution that presented aimed to start the allocation process, several minutes. This approach, combined Neural Network and Linear Regression on experimental data, to create the prediction models, it is very suitable for similar issues.

Giogourtsoglou et al. (2014) focused on the creation of a performance prediction model, analyzing the service’s behavior of the encyclopedic website Wikipedia.org. The selected parameters that influence the performance of the service were: a) the number of parallel requests from users, b) the location from where the requests are executed, and c) the time at which the requests are performed to the service. Similar critical parameters we used and in this study, because as it presented in study (Giogourtsoglou et al., 2014), the e-health application of PINCLOUD meets the same criteria as Wikipedia.org, about the parameters that influence the service’s performance.
3 PROPOSED SOLUTION

Cloud Computing allows dynamic resource scaling for enterprise online systems. That is one of the key characteristics that differentiate the cloud systems from the traditional computing paradigm. The main issue of that feature is that the resource allocation is not instantaneous. The preparation that is needed in order to offer these resources dynamically is a time consuming operation and can even take hours to complete. This study focuses on a model that will predict the service performance, based on effecting parameters (e.g. number of parallel requests to the service) and forecast the resource needs.

To create such a model, it is necessary to collect experimental data, about the critical effecting parameters (Giogourtsoglou et al., 2014). For this collection is needed an application, which collects the necessary data, and stores them to a specific format that will be easy to analyze them by the next steps. The service that its behavior is analyzed is the e-health application of PINCLOUD project. PINCLOUD provides integrated e-health services for personalized medicine like e-PHR, e-prescription and e-referral. Also, seeks to integrate different application components, leading to the provision of an end-to-end personalized disease monitoring and medical data service “anytime, anywhere”, which ensures independent living regardless of age (Koufi et al., 2014).

The parameter that determines the performance of this service is the response time of every request, and the critical effecting parameters are a) the number of parallel requests from users, b) the location from where the requests are executed, and c) the time at which the requests are performed to the service. So the application must collect data for those four parameters.

For this study, an appropriate application has to be created, to collect the necessary data, about service’s behavior. This application is separated into two parts, acting like two stand-alone applications. The first one managed the whole procedure of data collection. The user inserts the arguments that are: a) the target service, in which will execute the requests, b) the number of parallel requests that will execute every time and c) the public IP and TCP port of Figure 1: PINCLOUD application components (Koufi et al., 2014)
remote computers that will execute those requests. The second application is installed on remote computers, which are located in different locations. The PINCLOUD project aims to provide e-Health applications to Greek residents. So, the second application should be located in different locations in Greece. When the second application is executed, it is waiting to get the arguments about the requests to service, which are sent from the first application to remote computers. When the whole process ends, all the data that have been collected from computers that made the requests to service, are send back to computer, which execute the first application. In this paper, we propose the computers that execute the second application, to be located in different cities in Greece like Athens, Thessaloniki, Patra, etc. The first application will be installed in the computer that user manages the data collection procedure.

In this procedure it is important to record the distance between the cloud infrastructure of the e-health application and the remote computers where the second application is executed. The reason that the distance should be recorded is because is a critical parameter about service’s behavior serving requests from different distances.

Next step, after the data collection, is the data processing. In order to process them a statistical tool should be used. This tool would be, for example, a program that uses R language commands (http://www.r-project.org/, 2014). The main process that have to be done is to decrease the volume of data, which probably be several MBs, and will be used at the next steps. When that process will be finished, all the data should be in the right format and the right size (probably KBs) to be easier to use them to produce the prediction model.

In order to create the prediction model, it is necessary to determine the shape of the data with mathematical terms. This step is important because according to shape of the model depicted in XY axes, the next process of data will change. For example, a shape of the depicted function of data that is a multiple nonlinear model.

Next procedure includes the selection of the appropriate mathematical functions. According to previous example (multiple nonlinear model), to find the model that fits to data, we need to use a combination of nonlinear mathematical functions like polynomial functions. The selected functions combined with the data used in application Eureqa(Nutonian, 2014), to produce some prediction models. The form of the model would be like the follow:

\[ \text{responseTime} = f(\text{time}, \text{requests}, \text{distance}) \]

This procedure is a time consuming operation that probably take several weeks to complete. The results will be more than twenty models that fit with the data. Those models must be compared to choose this that would fit better to data. To choose the best model, we will use some mathematical terms. Those terms are: a) R Squared Goodness of Fit, b) Correlation Coefficient, c) Maximum Error, d) Mean Squared Error, and e) Mean Absolute Error. According to previous terms we will select the best prediction model.

Last step is the evaluation of the produced performance prediction model. In order to do that, we have to collect data again with different user distribution (e.g. 15 different cities on
Greece. Then we will compare the results of the produced prediction model with the real data from new data collection. This process will help to evaluate the model and make it more correct according with the real results.

The following figure sums up the whole process to create the prediction model. In the next section will be presented some future plans about the aforementioned process.

![Research flow to produce the prediction model](image)

**Figure 2: Research flow to produce the prediction model**

4 **FUTURE PLANS**

Future plans include the execution of the aforementioned process to create the performance prediction model for the PINCLOUD e-health application. In the Data Collection process it is necessary to install the data collection’s application to computers that are located in different locations in Greece, aiming to simulate the real requests from around the country. Also, Data Collection process should be considered to be executed in different time periods and with different consumption patterns (e.g. 4 cities burst – 11 cities sleeping).

Related future work includes improvement of the produced prediction model analyzing more parameters that affect the service’s performance, like cloud infrastructure hardware parameters and other application parameters. The next steps of data processing and model production should be the same as described before.

5 **CONCLUSION**

This paper presents a proposal about the creation of a model for dynamic resource prediction of a service (PINCLOUD e-health application). This model will help ensure that the preparation of the necessary resources, to offer to a service when they are needed, will start on time. To create that model, we presented a number of steps starting with the collection of experimental data that depicted the behavior of a service. The main parameter that reflects the
performance of the service is the response time of requests. The other parameters that studied were: a) the number of parallel requests from users, b) the location from where the requests are executed, and c) the time at which the requests are performed to the service. Analyzing and processing the collected data by the appropriate tools produced some models which were evaluated in order to find the best fit to the data. The model comparison includes mathematical terms like: a) R Squared Goodness of Fit, b) Correlation Coefficient, c) Maximum Error, d) Mean Squared Error, and e) Mean Absolute Error.

Finally, we presented the related future work about the creation of the performance prediction model for the PINCLOUD e-health application. Also, we presented some ideas about the future plans which include studies regarding more parameters.

Acknowledgement

The work reported here has been carried out in the framework of national project Providing Integrated eHealth Services for Personalized Medicine utilizing Cloud Infrastructure (Pincloud), led by the University of Piraeus, conducted in the context of the National Strategic Reference Framework NSRF 2007-2013, Cooperation 2011 and co-funded by the European Commission.

References


DRIFTING CLOUDS: A CASE STUDY OF IT TRANSFORMATION AT eni

Giorgio De Michelis: Department of Informatics Systems and Communication (DISCO) - University of Milano Bicocca, Italy. email: gdemich@disco.unimib.it (Corresponding author)

Gianluigi Viscusi: College of Management of Technology (CDM) - École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. email: gianluigi.viscusi@epfl.ch (Corresponding author)

Tiziano Salmi, Direzione Information & Communication Technology, IT Infrastructure, Architecture & Application Operations, eni s.p.a, via Fabiani 1, 20097, San Donato Milanese, Milan, Italy. email: Tiziano.Salmi@eni.com

Abstract

In this paper, we investigate how cloud computing can be seen as a planned result or perhaps a consequence of IT transformation initiatives. To this end we discuss the case of eni, a multinational utility corporation based in Italy. The analysis has been carried out following an interpretative perspective, on the basis of the observation of the current structuring of the infrastructure in relation to IT transformation strategy. The latter has been considered in general, thus, independently from the existence or lack of a clear cloud strategy at business level rather than at IT level. Contrary to other contributions we first focus on IT transformation, showing how in the considered case the principles of simplification and virtualization create shared understandings suitable to facilitate the idea of cloud computing as “the next step”. Thus, the main contribution of the paper is to show how cloud computing initiatives may rise from drifting of other unrelated IT transformation strategies rather than from a clear top down cloud computing strategy. To this end, state of the art dimensions are used to identify different steps of a path for the emerging of cloud computing infrastructure, pointing out the role of openness and modularity.

Keywords: cloud computing, IT transformation, bricolage, drift, information infrastructure

1. INTRODUCTION

The growth and rapid diffusion of cloud computing as a new paradigm for hosting and delivering services over the Internet seems to have brought new opportunities and challenges for businesses and academics (Zhang, Cheng, and Boutaba 2010; Willcocks, Venters, and Whitley 2011). However, at the state of the art, few case studies have yet allowed investigations on how cloud computing initiatives relate to long-term transformation of IT in both public and private organisations (Murugesan 2013; Kagermann, Osterle, and Jordan 2010; Chang and Wills 2013). In this paper, we

1 “Drifting Clouds” (Ukigumo) is a Japanese novel written in 1887 by Futabatei Shimei (Ryan 1965). The novel has no ending and prioritises the development of characters over plot. However, for further comments we refer the reader to (Rubin 1970).
discuss cloud computing at eni corporation (a multinational utility corporation based in Italy) as the result of a long term IT transformation initiative.

The main goal is to describe a case study where cloud emerges from drifting of other unrelated IT transformation strategies rather than from a clear top down cloud computing strategy. Thus, the analysis has been carried out on the basis of the observation of the current structuring of the infrastructure in relation to IT transformation in general. As a consequence, the paper aims to provide an emergent perspective on cloud as an evolving information infrastructure compared to current top down rhetoric of cloud adoption as a result of alignment of IT with business strategy, see as an example the discussion in Oracle (2010).

In the following sections we first discuss background literature used to build the theoretical framework guiding the case study. Then, having introduced the research method adopted, we detail the case study, and discuss results and limitations, before some concluding remarks on the analysis carried out.

2. Background and Motivations

Cloud computing is a complex construct, which covers various aspects of information systems engineering, requiring different business models and bringing interesting research questions for computing, software engineering and information systems, likewise (Venters & Whitley, 2012). As a consequence, several definitions of cloud computing exist in literature as pointed out by Vaquero et al. (2009), which provides the following definition: “Clouds are a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services). These resources can be dynamically reconfigured to adjust to a variable load (scale), allowing also for an optimum resource utilisation. This pool of resources is typically exploited by a pay-per-use model in which guarantees are offered by the Infrastructure Provider by means of customized SLAs” (Vaquero et al., 2009). The definition emphasizes the service nature of cloud computing and virtualisation as the issue characterising the different resources which make it up. Considering the different types of clouds on the basis of the resources for the information systems layers they impact, such as, e.g., Infrastructure as a service (IaaS), Platform as a service (PaaS), Software as a Service (SaaS), Business process as a service (BPaaS), it is worth noting that each one may represent various degrees of evolution and innovation of the infrastructure of a given organisation, encompassing not only information technology, but also organisation and work practices, services and linkages to environment and social context (Ciborra et al., 2000).

Furthermore, as pointed out by Willcocks et al. (2011), cloud computing may be seen as the infrastructure for business services innovation following different paths, depending on the strategy proposition. Consequently, the different paths span from incremental innovation through cost control and virtualisation, architectural innovation through improvement in business processes and agility, to radical innovation through collaboration and consumerisation. Thus, a clear understanding of the characteristics of each type of cloud is relevant for identifying the appropriate evolution path suitable for an organisation to be followed. As an example, IaaS is a core type when cost control and consolidation are the strategy focus, whereas it must be coupled with the PaaS type when the interest is in agility of business processes. SaaS may be decoupled or built on the previous types when the organisation focus is i) as a provider, on simplified software installation, maintenance and centralized control over versioning; ii) as an end user to access service “anytime, anywhere”, share data and collaborate, without necessarily building or having a own data centre (Armbrust et al., 2010). However, it is worth noting that each path has to be further framed considering another relevant dimension of cloud computing, being the current adoption strategy oriented or based on a private or public cloud perspective (Armbrust et al. 2009; Armbrust et al. 2010; Buyya et al. 2013; Marston et al. 2011).

The characteristics supporting the identification of the evolution paths enabled by different types of cloud may be related to three clusters of dimensions resulting from the analysis of the literature on the benefits and the advantages of cloud: cost/rationalisation, agility, and consumerisation (see Table 1). We consider these dimensions as clusters that include most of the relevant characteristics of cloud
computing discussed in literature. However, for a critical review we refer to Venters & Whitley (2012) who also provide a classification scheme based on the cloud “desires” at technological and service level, emerging from results of a large survey on users. As for the technological dimension, Venters and Whitley (2012) consider abstraction, equivalence (in terms of security, latency, and availability), scalability, and variety; whereas the service dimension is characterized by efficiency, creativity (as innovation capability often in contrast with agility), and simplicity (Venters and Whitley 2012). Thus, in Table 1, we adopt and adapt the distinction between technological and service level, associating them to the different dimensions of the three higher-level cloud advantage types.

Cost/rationalisation is related to the desire of abstraction from unnecessary complexity in technical services provision, ranging from the least abstracted computing hardware to the most abstracted service (Venters and Whitley 2012). Furthermore, cost/rationalisation relates with characteristics of infrastructure and platform as a service, such as an effective budget allocation based on the use of calculation power or applications (pay per use), the latter being allowed by control over available resources at service level (resource optimisation), and technological level as for their interoperability and customisation (security and standardisation). Moreover, agility asks for rationalisation at infrastructure and platform level, in order to have a simplified portfolio of applications (simplification) exploitable by business processes with the appropriate efficiency in terms of response time and calculation power. Thus, the efficient exploitation of the portfolio is possible due to the abstraction (virtualisation) of logical resources away from their underlying physical resources (Buyya et al. 2013; Rimal, Choi, and Lumb 2009), being consequently adaptable to different business model needs (scalability).

<table>
<thead>
<tr>
<th>Advantage Dimensions</th>
<th>Reference</th>
<th>Type of dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstraction</td>
<td>Venters &amp; Whitley (2012)</td>
<td>Technological</td>
</tr>
<tr>
<td>Pay per Use (Efficiency)</td>
<td>Armbrust et al. (2010); Vaquero et al (2009); Venters &amp; Whitley (2012)</td>
<td>Service</td>
</tr>
<tr>
<td>Resource Optimisation (Efficiency)</td>
<td>Armbrust et al. (2010); Venters &amp; Whitley (2012)</td>
<td>Service</td>
</tr>
<tr>
<td>Security (Equivalence)</td>
<td>Venters &amp; Whitley (2012)</td>
<td>Technological</td>
</tr>
<tr>
<td>Standardisation (Simplicity)</td>
<td>Vaquero et al (2009); Venters &amp; Whitley (2012)</td>
<td>Technological</td>
</tr>
<tr>
<td><strong>Agility</strong> (Willcocks, et al., 2011)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scalability</td>
<td>Armbrust et al. (2010); Vaquero et al (2009); Venters &amp; Whitley (2012)</td>
<td>Technological</td>
</tr>
<tr>
<td>Simplification (Simplicity)</td>
<td>Armbrust et al. (2010); Vaquero et al (2009); Venters &amp; Whitley (2012)</td>
<td>Service/Technological</td>
</tr>
<tr>
<td><strong>Consumerisation</strong> (Willcocks, et al., 2011)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Availability (Equivalence)</td>
<td>Venters &amp; Whitley (2012)</td>
<td>Technological</td>
</tr>
<tr>
<td>Creativity</td>
<td>Venters &amp; Whitley (2012)</td>
<td>Service</td>
</tr>
<tr>
<td>Internet Centric</td>
<td>Vaquero et al (2009)</td>
<td>Service/Technological</td>
</tr>
<tr>
<td>Latency (Equivalence)</td>
<td>Venters &amp; Whitley (2012)</td>
<td>Technological</td>
</tr>
<tr>
<td>Variety</td>
<td>Armbrust et al. (2010); Vaquero et al (2009); Venters &amp; Whitley (2012)</td>
<td>Technological</td>
</tr>
</tbody>
</table>

Table 1. Cloud advantage dimensions considered at the state of the art.
Finally, cloud computing is related to the **consumerisation** of IT and the promise of computing resources delivered as utility, and purchased off-the-shelf (Willcocks, et al., 2011). Moreover, **consumerisation** characteristics are related to the availability as services (Sasikala 2011) over the internet (**availability** and **internet centric**) of a wide range of equivalent IT resources (**variety of resources**), allowing a reliable, rapid, and flexible adaptation of them to business needs, without a human intervention (**latency** and **automatic adaptation**).

As a consequence, cloud computing impacts on the role and the capabilities of IT units in organisation (**creativity**), in terms of their capacity to provide scalable IT enabled services (Lewis, Mathiassen, and Rai 2011) and to exploit digital options (Overby, Bharadwaj, and Sambamurthy 2006; Sambamurthy, Bharadwaj, and Grover 2003). These perspectives consider the exploitation of digital options as part of planning and management of the corporate infrastructure, traditionally emphasizing the standardisation of systems, data and processes as a way to reconcile or align them with (IT/Business) strategies (Ciborra 2000; Melville, Kraemer, and Gurbaxani 2004; Luftman 1996). Whereas in this paper we want to pursue the perspective proposed by Ciborra (1997, 2004) that planning and implementation show technology having a certain degree of autonomy and inner dynamics, both as a drifting system and as an organism to be cultivated. Consequently, we adopt the definition of drift as “**the outcome of the match between two agents: technology-possessing affordances; and humans in their roles of sponsor, user, and designer**” (Ciborra 2004, p.91). Furthermore, it is worth noting that to combine consumerisation and agility characteristics, cloud computing paradigm requires highly modular information systems. Modularity allows for components to be separated and recombined, depending on the actual business needs, often not planned ex ante, and also enabling new (unexpected) configurations through what has been called a situated **bricolage** rather than adoption of formalized methods (Ciborra 1992).

The above perspective, little considered in the literature on cloud computing, allows framing the development of cloud computing initiatives in the broader concept of **information infrastructure** (Star and Ruhleder 1996) and to problematize the characteristics in Table 1 in terms of information infrastructure tension between **standardisation** and **flexibility** (Hanseth, Monteiro, and Hatling 1996; Star and Ruhleder 1996). Moreover, the characteristics in Table 1 may be compared with information infrastructure as characterised by **openness to users, interconnections of numerous modules/systems** (i.e. multiplicity of purposes, agendas, strategies), **dynamically evolving portfolios of systems**, **embeddedness**, being **shaped by an installed base** of existing systems and practices (Monteiro et al. 2012; Star and Ruhleder 1996). Taking these issues into account, cloud computing can be considered a digital information infrastructure resulting from a process of digitalization (Tilson, Lyttinen, and Sørensen 2010). Consequently, it is worth noting the composite nature of digital infrastructure mechanisms. As pointed out by Henfridsson & Bygstad (2013), the latter are actually connecting three types of mechanisms: **situational mechanisms** (macro–micro level, explaining how the infrastructure enables and constrains its various components), **action-formation mechanisms** (socio-technical action, explaining how desires and beliefs generate a specific action), and **transformational mechanisms** (micro–macro level, explaining how different components interact to produce an outcome) (Henfridsson & Bygstad 2013). In what follows, the proposed case study aims to observe the above issues, in particular, considering how cloud computing may emerge from the drifting of an IT transformation not contemplating it as its original explicit target outcome.

### 3. Theoretical Framework

As also emphasized at the state of the art, a relevant distinction between technological and service **desires** guides cloud computing adoption by businesses (Venters & Whitley, 2012). In the following we rely on this distinction and we focus on the technological desire related to infrastructure and platform as a service, characterized by a extension of the role of users at the configuration level (Papazoglou 2012; Nguyen et al. 2012). This desire provides a basis for investigating cloud as a basis for an emergent information infrastructure and identifying potential growing paths for its evolution, likewise. In particular, we focus on agility dimensions. Notwithstanding agility is recognized as being a key business priority (Luftman et al. 2012), it is suitable here to consider practitioner's distinction.
between agility as engineering resource availability, and agility as business response to changing conditions or opportunity (Golden 2012). Indeed, as pointed out by Golden (2012): "if cloud computing comes to be seen as an internal IT optimisation with little effect on how quickly computing capability rolls out into mainline business processes, the potential exists for IT to never receive the business unit support it requires to fund the shift to cloud computing. If the move to cloud computing is presented as 'helps our programmers program faster,' necessary funding will probably never materialize."

Thus, agility dimensions of cloud computing refer either to IT agility (in the case of infrastructure and platform) or Business agility (in the case of service oriented applications and platforms); whereas virtualisation is often considered both the foundation for cloud and the main source of cost reduction and saving advantages (Vmware 2012). Furthermore, focusing on agility as the ability of firms to sense environmental change and respond readily (Overby, Bharadwaj, and Sambamurthy 2006), the role of a cloud enabled IT agility is related to the enforcement of an IT architecture suitable to i) avoid limiting the range of responses available to a firm, ii) simplifying the firm to adjust processes to changing conditions, and iii) enabling information visibility and compatibility with systems adopted by customers and suppliers (Overby, Bharadwaj, and Sambamurthy 2006). As for these issues, we consider agility under a not clear-cut perspective of the boundaries between IT and Business agility.

This perspective results from a relational and emergent definition of information infrastructure as a complex bundle of evolving dimensions (Ciborra, 2000). In line with this perspective we consider the infrastructure role as a bundle of utility (reducing costs of processing and communicating information throughout the organisation), dependence (improving with modularity the performance of key business processes), enablement (supporting evolution with architectures and platforms for new applications and new businesses).

Taking these issues into account, Figure 1 shows the theoretical framework resulting from literature analysis, whose aim is to provide a description of cloud adoption from an information infrastructure perspective. The focus of the framework is on virtualisation and simplification (encompassing scalability) seen as agility (mainly technological) dimensions of cloud computing at infrastructure and applications level; whereas we consider again simplification as related to the platform level as being at the top of infrastructure or else at the basis of the application evolution path. These dimensions are then investigated in their relationship with information infrastructure characteristics (embeddedness, modularity, openness, evolutionarity) to describe the type of adoption of cloud by an organisation (see quadrants in Figure 1). In the following Sections we first discuss the research method adopted before introducing the case study, and providing the details for the analysis carried out.

**Figure 1. Framework for evaluating the type of cloud adoption.**
4. RESEARCH METHOD

In this paper we conduct qualitative research as an inquiry process of understanding based on distinct methodological traditions (Creswell 1998, 15). Thus, we discuss a single case study, whose rationale is to represent an exploratory case study (Yin 2009), based on an interpretive approach to research in information systems (Klein and Myers 1999; Walsham 1995), involving both researchers and practitioners. In particular, the study aims to produce an understanding of the context of cloud computing adoption, and the process whereby cloud computing influences and is influenced by the context (Walsham 1993). To this end we adopt the framework in Figure 1 as an interpretative tool to investigate in a real case the correspondence between cloud types, advantage dimensions considered at the state of the art (see Table 1) and, on the one hand, the meaning provided to cloud computing by IT people involved in transformation initiatives; on the other hand, the framework is going to support a mapping to a temporal evolution of those IT transformation initiatives. As for sources of evidence we have considered:

- **Documentation**: memoranda and formal reports;
- **Archival records**: organisation charts and service, personnel or financial records;
- **Interviews**: eight semi-structured interviews to four middle managers and one corporate manager.

Starting with a first round on May 2012 and concluding with a final session on March 2013, each interview has had an average duration of 40 minutes, following guidelines focused on cloud computing with respect to IT transformation in eni, focusing on IT services and demand management, applications portfolio, consolidation and virtualisation initiatives. Interviews were mainly open-ended or made up of a limited number of questions (max 5) in order to obtain reach material, avoiding the imposition of preconceived concept on it at the data collection step of the research activity (Charmaz 2006). It is worth noting that one of the authors has been directly involved in the IT transformation at eni, having a direct experience of the overall process. As a consequence, in the course of the research its commentaries and notes have been considered as part of the overall documentation, and analysed from a critical perspective, both independently and in common sessions, by the other authors not directly involved in eni organisation. The goal of this procedures being both grounding theory on data and managing reflexivity through its critical acknowledgment (Strauss 1987; Bryant and Charmaz 2007; Charmaz 2006).

In the following we first provide an overview of the case, detailing the main insights and understanding on IT transformation and cloud computing at eni as emerging from the collected materials. Subsequently, we discuss these results in terms of contribution and limitation of the research.

5. IT TRANSFORMATION AT ENI

Established in 1953, as **eni (Ente Nazionale Idrocarburi)**, today eni is one of the most important integrated energy companies operating at global level in the oil and gas, electricity generation and sale, petrochemicals, oilfield services construction and engineering industries. Based in Italy, at the time of the research presented in this paper eni is active in 85 countries with almost 80,000 employees. The IT unit has a staff of 675 people (supported yearly by an almost equivalent number of people from service providers), a yearly budget of 475 million euros and a dedicated budget for other projects of 250 million euros.

In 2006 eni launched an IT transformation program aiming to redesign the entire infrastructure and operational management of applications. As the current Chief Information Officer (CIO) pointed out during one of the meetings, the design of IT transformation in eni began due to “blocks of the operations for failure of the central data centre”, thus, electromechanical accidents rather than information systems or digital innovation needs. As for IT infrastructure, in 2006 each application had its own server or its own server reference. Likewise, as argued again by current CIO: “there was no clear strategy on how to build the infrastructure, and we had a high fragmentation: different and heterogeneous servers, with high operating costs and low flexibility. The continuity was affected too:
apart from critical applications, full redundancy was not economically sustainable. When a machine broke, then, there was no chance to migrate its applications on another one in a short time” (Gianni 2012 our translation). Indeed, in 2007 eni started an initiative of consolidation, following a server farm logic, having a small number of big machines that could be partitioned.

The consequence of early actions has been a reduced operational complexity, but a high heterogeneity. In parallel, eni initiated an extensive program of virtualisation, leading in the case of Intel machines to replacing a few hundred small servers by thirty very large fully virtualized servers (see Figure 2). These initiatives implemented an IT strategy focused on costs efficiency, as also noticed by one of the interviewees: “since 2007 we started to consolidate through virtualisation. We virtualised X86 platforms aiming at efficiency of cost-savings”.

The consolidation and virtualisation initiatives concluded their course with the building of a new data centre infrastructure totally homogeneous and integrated in October 2013. It is worth noting that the vision of the CIO sees a data center “not merely as a building where to stock machines, but an ‘industrial plant’ as a foundry”, further claiming the old-economy nature of eni business, “which is not actually a digital enterprise” (interview with CIO).

Indeed, the new data centre has been actually thought and designed as an industrial plant, disposing large amount of heat in an energy-efficient manner, using free cooling (air circulation) enabled by chimneys above the engine rooms, aspiring 8 million m³ of air per hour. Considering the plant produces computing power and heat, with a high energy density of 10KW per m² (30 megawatts), the goal stated by the CIO has been an Objective Power Usage Effectiveness (PUE) 1.2 (in Italy, average 2.7) better than Google PUE 1.28. As for this benchmark, the CIO in one interview pointed out the difference with Google being in its use of ad hoc machines; whereas eni focused on modularity and standardisation enabled by the data centre architecture and structure: “We buy them at the supermarket, they are not ad hoc. Our focus is on the plant and on the ‘continuity of service’: if the machines (BLADE) are broken, the replacement cost is balanced by the energetic saving guarantee by the data centre infrastructure”. It is worth noting that the off-the-shelf solution for the “hardware-as-a-service” perspective adopted for the data centre requires standardisation and modularity of the software portfolio.

Besides the above mentioned infrastructure initiatives, in 2007 eni launched an efficiency program for the IT unit which produced, in 2011, a reduction in staff from 1045 to 675 people and a saving of almost 100 million euros a year on external spending. The budget stabilisation was followed by a constant attention to cost optimisation, complemented by precise indicators of quality of outsourcing. As pointed out by current CIO “in other words it is just putting suppliers under pressure, but there are thresholds below which it is not reasonable to go down because, as I often say ‘if you pay peanuts you get monkeys’. We are aware of what are the tariffs under which it becomes extremely dangerous to go down, either because it pollutes the market - because they are not real - and because it leads to a deterioration of quality.” (Gianni 2012 our translation). As a consequence, the IT transformation
program aimed to obtain efficiency gains through a focused insourcing strategy of core activities and services, whose principles are simplification, timeliness, central role of operational resources, and alignment with the business (consolidation led to 105 outsourcer in 2008 compared to 173 in 2005).

This strategy has completed the earlier focus on infrastructure with a rationalisation of applications. Some have been eliminated, other incorporated or replaced in order to be compliant with the current platform, the latter resulting from the consolidation and virtualisation program started in 2007 (see Figure 2). The platform based on Linux aimed to satisfy the strategic goal of dismissing all proprietary UNIX platforms in favour of open alternatives. Thus, the rationalisation program has led to a reduction of the number of applications in portfolio from 540 to 400. Among them nearly 250 are compliant with the consolidated standardised platform. However, a number of specific industrial applications continue to run on dedicated servers, because the supplier does not make them available on the current eni consolidated standard platform. The program was guided by the business needs and did not consider the case requiring the rewriting of the code or software packages linked to specific requirements. Moreover, due to the alignment with the business needs the complexity of the program resided in the use of many applications in countries other than Italy (where the three core data centres and the IT direction are based) and must ensure that the service is not interrupted.

As pointed out by one of the interviewees: “we are defining the standard level of our applications, on top of which we will design and develop new services compliant with this interoperability framework”. In the case of enterprise systems such as SAP, the result of the rationalisation has not been a unique integrated system, but a network of modules each dedicated to one and only one specific cluster of processes (accounting, human resources, etc.): “we have a unique kernel, to whom we say ‘this is your SAP’ and this SAP is integrated in a centralised way” (interview with middle manager). Indeed, in eni there is a unique central virtual SAP that provides services to the overall corporation at global level, where the consolidation and virtualisation has been produced on the basis of common workflow of resource usage (“the goal was to generate agility in resource allocation…we have taken a traditional SAP, we have putted it on top of the new infrastructure in order to test its compliance” - interview with middle manager). As said above, the adopted platform is open and use solely industry standard blade servers based on Linux operating system, pushed with a level of virtualisation in order to have a greater saturation of the CPU, a better ability to absorb peak loads, enabling at any time the allocation to applications which need their computational power, and a better continuity of service. As for the management of applications on servers of the new data centre completed in 2013, integrating two other existing data centres, eni has developed an operational plan (Dynamic Infrastructure eni). Thus, 2013 has been dedicated to the gradual shutdown of the old server and start applications on the new structure.

6. CLOUD COMPUTING AT ENI

The current outputs of the IT transformation program at eni may be considered as resulting in two of the current recognized cloud types: infrastructure as a service (IaaS) and platform as a service (PaaS). Nevertheless, cloud computing was not part of the initial IT transformation strategy where the focus was initially mainly on consolidation and virtualisation of the infrastructure. Only in 2009, with the strategic planning of a new integrated big data centre, the translation of the infrastructure as potentially being perceived or conceived in terms of cloud began: “it has been the discontinuity, producing other ideas that we have then connected in a new perspective. We have said each other: ‘if we do this, we can do this too, etc.” (interview with middle manager). Furthermore, as emerging from the discussion of the IT transformation program outlined in the previous section, the results of the earlier initiatives in eni comply with at least four out of the seven cloud capabilities identified by lyer and Henderson (2012; 2010), namely, controlled interface (eni infrastructure and platform aim to be organic and responsive to changing in business requirements), location independence (the eni infrastructure aims to enable control and access to services regardless of the global location), sourcing independence (as we have seen the eni infrastructure and platform aim to consolidate and control the quality levels of sourcing), virtual business environment (eni platform aims to be a suite of integrated applications that support business needs). Nevertheless, there is little agreement among the people
interviewed at eni in considering cloud as the core or at least the most appropriate strategic perspective under which seeing the IT transformation strategy. As pointed out by the current CIO: “there is too much emphasis around the word ‘Cloud’, but it is true that at the infrastructure level we can have an Infrastructure as a Service, and at platform level we have a good percentage of Platform as a Service.”

Furthermore, it is worth noting that the consideration of cloud as a drift of other not related initiatives emerges also from an evaluation of the current platform by middle managers: “the final result is a set of parallelisms even I would say ‘serendipitous’...in the beginning we might not even call it cloud”. The identification of cloud related concepts such as flexibility, continuity, improved efficiency in the management of the resource calculation, occurs only in a second time while consolidating and virtualising the infrastructure and platforms (“but we do not need to create a true private cloud”). Finally, it is worth noting that while cloud emerges as a drift of other initiatives, the consolidated infrastructure and the standardised platform allow for strategic consideration of cloud opportunities for eni business process management, as pointed out by the current CIO: “instead, we are investigating the opportunity to create a catalogue of business processes addressable as Business Processes as a Service (BPaaS), for which, together with colleagues from the business lines, deciding whether to purchase integrated services and not just IT services. Of course, ensuring proper interoperability with other applications in the Company.” (Gianni, 2012 - our translation).

In the following section we discuss findings resulting from the analysis of the case study presented above, providing further details about the different initiatives carried out at eni for work practice and IT organisation transformation at eni.

7. DISCUSSION OF THE FINDINGS

The case study reports the evolution of the IT transformation at eni, where cloud computing is understood by informants as an emergent perspective from the implementation of a general strategy aiming to carry out an overall IT transformation and not as a consequence of a strategy explicitly focused on cloud computing. Figure 2 shows how the above discussed theoretical framework may be adopted to map this evolution or drifting path. It is worth noting that the consolidation and virtualisation process started in 2006 at technological level (information infrastructure), since 2007 has being carried out in parallel to a specular program at organisational level to support better governance, cooperation, and computer supported knowledge sharing (work infrastructure). Furthermore, the years 2007-2008 represent a key period for the choice of adopting an open perspective on IT infrastructure and platforms, while the focus on work infrastructure has been combined with investments on the learning infrastructure, as pointed out by the CIO: “We had constant meetings and sent daily e-mails to align [staff] expectations. This was combined with an education programme” (CW 2010).

It is worth noting that eni, formerly a state owned institution acting in an oligopolistic national market, being a stock quoted corporation since 1995, inherited from its origins a centralised organisation. From the turn of the century, due to global competition constraints and the need for business agility to improve market orientation, eni moved its organisational model towards a decentralisation of responsibility for business processes and projects management. To this end, a new Chief Executive Officer (CEO), who joined eni in 2005, considered IT as a key driver and a strategic asset for change. Notwithstanding the dissatisfaction of senior management, the new CEO did not want to outsource IT, but charged the current CIO with the task of improving internal IT (CW 2010). Indeed, a set of IT competences were “lost” as a consequence of their outsourcing policies to IT vendors (see Dunleavy, Margetts, Simon, & Tinkler, 2006 for an interesting comparison with, e.g., similar actions taken by public agency and state owned organisations). As a consequence, insourcing IT competences, empowering an emerging community of practice and a modularisation and process oriented work organisation were among the strategy goals of the IT transformation initiative.

Thus, in 2009, when the IT transformation initiatives encountered cloud computing perspectives, the IT at eni looked at the convergence between these latter perspectives and their own experience rather
than at a change in the IT strategy. Furthermore, the idea of a translation of cloud into their own perspective, and not the opposite, had been reinforced during the years 2009-2010 by the further evolution of learning ("In 2009 our training budget was slightly more than 2008." - CW, 2010) and work infrastructure towards open and social paradigms. Indeed, in 2010 the IT rolled out a multimedia collaboration system called eniWave to encourage innovation, reaching 30,000 employees at global level (services being accessed by an average of 17,000 users a day) and a beta version of a corporate social network called moka. Thus, the path shown in Figure 2 supports the idea of embeddedness of information infrastructure with other infrastructures (Star and Ruhleder 1996; Pipek and Wulf 2009), pointing out the role of openness for the emerging of cloud computing infrastructure. Also, the above insights point out the findings of state art contributions arguing the role of innovation oriented strategy and the degree of sophistication of firm’s information infrastructure have a positive effect on the propensity to adopt cloud computing (Loukis and Kyriakou, 2015).

These points can be considered the contribution of the research presented in this paper together with the question left to a future field study as for the role of the different infrastructures guiding a path toward modularity (dotted line between the years 2012-2013 in Figure 2) rather than a straight evolution that see modularity as a side effect rather than a necessary requirement. Considering the qualitative research perspective, the contribution of the study aimed to build a holistic picture (Creswell 1998) of the cloud computing as emergent phenomenon from IT transformation initiatives.

We are aware of the limitation of this research in terms of both the difficulty to generalise the results of the case study and of the limited number of interviewees under the constraints to the access to the field posed by a global corporate context; nevertheless, the information provided by the research participants seem to be appropriate to support the interpretative work being integrated with sources. Finally, it is worth noting that a small sample and limited cases are considered as not posing a problem when the interviews are intensive and the aims of data collection is directed to illuminate properties and relations between categories (Charmaz 2006). In future work, we would try to continue the observation of the IT transformation at eni; depending on the type of access to the field we will increase the number of interviews in order to engage in developing a grounded theory approach, which we consider as more suitable to elicit a theory from the data collected for the interpretative but mainly exploratory aims of the case study research discussed in this paper.

8. Conclusion

As we have pointed out in the previous sections, IT transformation at eni can be considered a very large example of a cloud computing project in a multinational company. Its main feature has been the realisation of a well performing private cloud, where sustainability, standardisation, and efficiency have been increased. From this viewpoint, IT transformation at eni is mainly a IaaS project, where benefits come from efficiency in computing with respect to energy consumption, and from a radical simplification of the software both at the infrastructure and application level (from a number of operating systems to 2; from over 540 applicative systems to almost 400). It has also to be mentioned that IT transformation in eni is a project paying itself over the years, on the basis of the plans defined by the IT unit. It has also to be recalled that IT transformation at eni is also a PaaS project, reducing the number of operating systems and network protocols and services adopted at eni and concentrating their management and maintenance at the corporate IT unit. It is expected that the benefits of the cloud, from the PaaS point of view, will also be quite relevant, increasing the quality of the exercise of the information services of all its companies and reducing the costs of their ordinary maintenance. It is not easy to quantify the estimate of these benefits, but informants at eni reported that savings are the 9% of the overall IT annual spending. However, even if it is a very good result of five years of hard work and even if it seems to have a very good return on investment, is also a very challenging occasion for further innovation of eni’s information system. How can eni exploit the potential for innovation of its new Information System? Let us disregard the simplest actions they can do along the line on which they developed IT transformation in eni, looking where they can do relevant steps towards a radically new scalable, flexible and evolutionary system.
Claudio Ciborra (1992) has pointed out that innovation, at the Information Systems level, is frequently performed in terms of ‘bricolage’. Bricolage, in Ciborra’s thinking, can be seen as the constant re-ordering of people and resources to fulfil organisation needs, performed by “trying out” and experimenting with the available resources. For this reason, bricolage is not a random trying out, Ciborra emphasises that it is a trying out based on leveraging the world “as defined by the situation”. At the IT level, bricolage is difficult for the complexity and rigidity of its applications with respect to “trying out” new ways of accessing and using their data and functions. At eni as well as actually many large corporations this was due to both the heterogeneity of the platforms sustaining their Information Systems and to the strict integration of the functional units of the Enterprise-Resource-Planning (ERP) systems its companies adopted. IT transformation at eni solved the first problem, and only at a small degree the second one. Among the simplifications it has already performed at the application level, there is, in fact, the unification of some ERP components used by different companies within the group. The reason why this unification has been limited is that it was possible to reconcile different applications to a sole model, only when the differences among the existing applications were small or null, and this was quite rare due to, both, the inner complexity of the components of existing ERP platforms and the diversity of the eni group companies.

It is evident, that extending the reconciliation of existing applications to simple common models in a larger set of functions and data can be done only, if applications themselves become simpler, or better, composed by simpler units (modules), so that they can evolve substituting some modules, whenever requested and/or are needed. Modular Information Systems (systems built putting together a set of modules, selected within a library containing several versions of functional units, respecting a clear separation of data and functions) are the reference architecture needed to make this ‘bricolage’ approach to ICT innovation feasible and effective, but today’s ERP systems are far from modular. But also a promising opportunity that cloud computing offers a homogenous and assisted platform to Information System components and creates, for the first time, the conditions for effective modular (bottom-up) design. In other words, on the one hand, the most diffused ERP platforms are not modular and are trying to innovate without radical redesign and there is not, for what we know, any serious and relevant project willing to develop a new modular ERP platform, on the other hand, the Information Systems of the companies of eni group are not at all modular. The design and development of a modular ERP platform is not a duty for a company like eni, that is not active in the software market; nevertheless, this should not prevent eni from exploiting the potential of cloud computing for strengthening its bricolage capability, making slow but important steps towards a modular architecture of its Information Systems, and extending the SaaS dimension of its cloud. Thus, what companies like eni can do is experimenting a new, updated version of a ‘Migration of Legacy Systems’ (Brodie and Stonebraker 1995) approach that emulates, step by step, the dismantling of existing ERP systems. This will increase its capability to react timely and effectively to the requests coming from the business units and to increase the flexibility exhibited by its IT applications. It will not be a surprise, if, contextually, initiatives willing to develop modular ERP platforms will take form. Indeed, as shown in the case presented in this paper, principles of simplification and virtualization create shared understandings on modularity and openness suitable to facilitate the idea of cloud computing as “the next step”.

Thus, the main insights for managers as well as scholars concern the fact that cloud computing initiatives may rise from drifting of other unrelated IT transformation strategies rather than from a clear top down cloud computing strategy. Consequently, a specific attention and further research should be carried out to elicit the appropriate dimensions to identify different steps of a path for the emerging of an appropriate infrastructure from IT transformation strategies, thus enforcing effective cloud computing initiatives.

References


Ciborra, Claudio, Kristin Braa, Antonio Cordella, Bo Dahlbom, Angelo Failla, and Ole Hanseth. 2000. From Control to Drift: The Dynamics of Corporate Information Infrastructures. Oxford University Press.


STRATEGIES FOR CLOUD COMPUTING ADOPTION: INSIGHTS FROM THE NORWEGIAN PUBLIC SECTOR

Rania Fahim El-Gazzar, Department of Information Systems, University of Agder, Kristiansand, Norway rania.f.el-gazzar@uia.no

Fathul Wahid, Department of Informatics, Universitas Islam Indonesia, Yogyakarta, Indonesia, and Department of Information Systems, University of Agder, Kristiansand, Norway fathul.wahid@uii.ac.id

Abstract

This study aims to understand the adoption of cloud computing in the Norwegian public sector. Although considered a well-informed adopter of cloud computing, Norway’s adoption rate in the public sector is still low. Using a case study method guided by institutional theory, we defined field level changes, which created various isomorphic pressures: coercive, normative, and mimetic. The pressures have been responded differently by government agencies, which manifested in their strategy for adopting cloud computing, from adopting a public cloud to developing their own private cloud infrastructure. The study went further by identifying factors that either hinder or foster the adoption. Reasons behind non-adoption related to unsolved complexity, organizational conflicts, and specific business needs, but not risks of the cloud computing model itself. Furthermore, the study found that a certain strategic response could trigger a new pressure, which might be more aggressive. It revealed that previous unlikely incidents reported in the media had influenced the Norwegian public organization’s adoption strategy.

Keywords: cloud computing, public sector, institutional theory, isomorphic pressure, strategic response, developed country, Norway.

1 INTRODUCTION

The prevailing global market trends toward Cloud Computing (CC) and the enticing benefits it offers are inevitable for businesses (Marston et al. 2011; CA Technologies 2013). CC has shifted responsibilities, roles, and the way organizations run their businesses and manage their IT resources (Ragowsky et al. 2014; Schneider & Sunyaev 2014). CC brings desired agility, scalability, economies of scale, and strategic values to organizations; yet, it poses concerns and uncertainties (Venters & Whitley 2012; Schneider & Sunyaev 2014; Garrison et al. 2012; Malladi & Krishnan 2012). Concerns related to the adoption of CC range from technical and organizational to environmental (i.e., both business and legal) (Marston et al. 2011; Morgan & Conboy 2013; Owens 2010). Public organizations perceive CC as a steering wheel that drives innovation in delivering better public services (EN 2012; Frelle-Petersen et al. 2012). However, public organizations also perceive the risks associated with adopting cloud services, which mainly relate to technical, security, and legal issues (Paquette et al. 2010; Seddon & Currie 2013). To the best of our knowledge, limited research efforts focused on legal and business issues still exist (Yang & Tate 2012).

Despite previous research focused on identifying and addressing those risks, CC adoption rates in the public sector increase more slowly than in the private sector (Hawkins 2013). Likewise, although Norway is reported to be a promising market for CC, both in the private and public sectors (Haeberlen et al. 2013), the adoption rates in the Norwegian public sector (33%) are lower than the Norwegian private sector (38%) (Kristiansen 2011). The Norwegian public sector (i.e., municipalities, public organizations, and public authorities) began considering CC in their strategy (Ministry of Local
Government and Modernization 2013; Devoteam daVinci 2011). Accordingly, CC providers began investing in a more reliable infrastructure such as the capacity of data networks (Telenor 2013).

However, Norwegian public organizations’ experience with CC was confronted with several challenges, and their adoption did not always go smoothly (O’Donoghue 2012; Veum & Thoreid 2012; Jørgenrud 2011). Hence, it is important to understand their motivation and the strategy behind their adoption. Against this backdrop, this study sought to answer the following questions: How is CC adoption in the Norwegian public sector influenced by the surrounding institutional environment? Why did they respond differently to these influences?

In seeking these answers, we conducted an interpretive case study in the context of the Norwegian public sector. We used neo-institutional theory to guide our data analysis. This paper begins with the background of CC definition and CC research stream in the public sector in Section 2, followed by the background of borrowed concepts from the neo-institutional theory. Section 3 presents a description of the contextual setting and data collection for the case study method used in Section 4. Section 5 shows the findings framed within the neo-institutional theory concepts. The findings discussed in Section 6 by highlighting the insights gained from the case study, conclusions, and limitations.

2 CLOUD COMPUTING IN THE PUBLIC SECTOR

There is considerable interest in CC from public sector organizations; for them, CC “represents a fundamentally different way for government to architect computing resources, allowing governments to leverage powerful IT infrastructures in a fraction of the time it takes to provision, develop, and deploy similar assets in-house” (Shin 2013, p.194). With CC services, public organizations can reduce IT capital expenditures through pay-per-use subscription and improve their services to the public as well as overcome difficult financial crises (Haebgelin et al. 2013; Sultan 2010). However, public organizations are still lag behind private organizations in adopting CC services (Hawkins 2013; Baldwin 2012); this slow adoption has three reasons.

Firstly, the meaning of CC is not clear despite the literature contributed to defining the CC model and differentiating it from previous IT models (i.e., clusters and grids (Buuya et al. 2009)) and business models (i.e., IT outsourcing (Schneider & Sunyaev 2014)). Hence, CC definition states, “computing services are delivered over the Internet, on demand, from a remote location, rather than residing on one’s own desktop, laptop, mobile device, or even on an organization’s servers. For an organization, this would mean that, for a set or variable, usage-based fee—or even possibly for free—it would contract with a provider to deliver applications, computing power, and storage via the web” (Wyld 2009, p.6). CC services are offered as Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS), and Infrastructure-as-a-Service (IaaS) (Mell & Grance 2011). These services are deployed as public, private, or hybrid, based on the criticality of the data and applications (Mell & Grance 2011).

Secondly, data security risks regarding citizens’ records are further serious concerns that require adequate risk management approaches (Paquette et al. 2010) to assure that data confidentiality, integrity, and availability can address a wide range of security risks (Zissis & Lekkas 2011) to promote better experiences using CC services in the public sector (Shin 2013). In this regard, the literature contributed to introducing models for guiding the selection of CC services based on technical (i.e., security and performance) and economic (i.e., cost) criteria (Broberg et al. 2009; Garg et al. 2013).

Thirdly, laws and regulations related to data protection and privacy in the cloud, which constrain the exchange of public data through the cloud in highly regulated sectors, such as healthcare, are not mature or ready to regulate the CC industry (Schneider & Sunyaev 2014). These legal struggles pose enormous challenges to policy-makers regarding how to sustain the adoption of CC in the public sector and clarify the responsibilities of both public organizations and cloud providers (Seddon & Currie 2013; Wyld 2009). Hence, both public organizations and cloud providers should cooperate on complying with strict regulations “to avoid penalties for non-compliance and possible reputational damage if exposed by the media” (Seddon & Currie 2013, p.230). Thus, governments have to address a set of challenges to facilitate using a variety of CC services in public organizations (Wyld 2009).
The literature covered areas related to addressing security issues, providing adoption guidelines, and explaining legal struggles in the public sector (Shin 2013; Paquette et al. 2010; Zhao et al. 2013; Seddon & Currie 2013). However, the literature did not explain why public organizations follow different adoption strategies toward CC services. Hence, we aimed at exploring the institutional influences that affect the adoption of CC services in the Norwegian public sector and the reasons for different strategic responses to influences exerted by Norwegian public organizations.

3 THEORETICAL UNDERPINNINGS

Neo-institutional theory has been used to study Information Systems (IS)/IT-related phenomena in terms of adoption and use of IS/IT innovations in organizations in general (Currie 2009; Mignerat & Rivard 2009; Weerakkody et al. 2009) and in the public sector in particular (Wiredu 2010; Currie & Guah 2007). CC is an innovative IS/IT outsourcing model; its idea of ‘shared services’ enables enhancing organizations’ architectural agility and improving organizational learning and innovation (Su et al. 2009). Neo-institutional theory is a better lens to understand, in the context of IS/IT outsourcing, “how social choices are shaped, mediated, and channelled by institutional arrangements ...and how actors and their interests are institutionally constructed” (Currie 2004, p.238).

Organizations are required to conform to a set of rules and requirements at the organizational field level to legitimize themselves (Wooten & Hoffman 2008). The organizational field is defined as “a community of organizations that partakes of a common meaning system and whose participants interact more frequently and fatefully with one another than with actors outside the field” (Scott 2001, p.84). This may include regulatory bodies, business partners (i.e., customers and suppliers), peer organizations, competitors, and professional and trade associations.

Orlikowski and Barley (2001) argued that neo-institutional theory is helpful for understanding ‘how technologies are embedded in complex interdependent social, economic, and political networks, and how they are consequently shaped by broader institutional influences’ (p.154). These institutional influences enable or constrain organizations’ behaviour (Orlikowski & Barley 2001), and they are imposed by the community of organizations that create changes at the organizational field level in which the adopting organization operates. These isomorphic pressures are (DiMaggio & Powell 1983): coercive, normative, and mimetic (see Table 1).

<table>
<thead>
<tr>
<th>Isomorphic pressure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coercive</td>
<td>Imposed by the legal environment and can be either formal (i.e., laws and standards) or informal (i.e., cultural)</td>
</tr>
<tr>
<td>Normative</td>
<td>Imposed by professional associations, universities, and training institutions to produce guidelines for professional behaviour or individuals with similar orientation</td>
</tr>
<tr>
<td>Mimetic</td>
<td>Imposed by environment uncertainty and goal ambiguity so that organizations model themselves on other successful organizations in their field</td>
</tr>
</tbody>
</table>

*Table 1. Isomorphic pressures*

Organizations can respond differently to the three isomorphic pressures through five strategic responses to legitimize themselves within their organizational field (Oliver 1991). These strategic responses are (see Table 2): (1) acquiescence: when the organization conforms to institutional norms through imitate, habit, or comply tactics; (2) compromise: when the organization accommodates multiple institutional stakeholders through balance, pacify, or bargain tactics; (3) avoidance: when the organization precludes the need for conformity to institutional norms through conceal, buffer, or escape tactics; (4) defiance: when the organization resists the institutional norms through dismiss, challenge, or attack tactics; and (5) manipulation: when the organization actively exerts power over the sources of the pressure by applying co-opt, influence, or control tactics on institutional constituents.
<table>
<thead>
<tr>
<th>Strategies</th>
<th>Tactics</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquiescence</td>
<td>Habit</td>
<td>Following invisible, taken-for-granted norms</td>
</tr>
<tr>
<td></td>
<td>Imitate</td>
<td>Mimicking institutional models</td>
</tr>
<tr>
<td></td>
<td>Comply</td>
<td>Obeying rules and accepting norms</td>
</tr>
<tr>
<td>Compromise</td>
<td>Balance</td>
<td>Balancing the expectations of multiple constituents</td>
</tr>
<tr>
<td></td>
<td>Pacify</td>
<td>Placating and accommodating institutional elements</td>
</tr>
<tr>
<td></td>
<td>Bargain</td>
<td>Negotiating with institutional stakeholders</td>
</tr>
<tr>
<td>Avoidance</td>
<td>Conceal</td>
<td>Disguising nonconformity</td>
</tr>
<tr>
<td></td>
<td>Buffer</td>
<td>Loosening institutional attachments</td>
</tr>
<tr>
<td></td>
<td>Escape</td>
<td>Changing goals, activities, or domains</td>
</tr>
<tr>
<td>Defiance</td>
<td>Dismiss</td>
<td>Ignoring explicit norms and values</td>
</tr>
<tr>
<td></td>
<td>Challenge</td>
<td>Contesting rules and requirements</td>
</tr>
<tr>
<td></td>
<td>Attack</td>
<td>Assailing the sources of institutional pressure</td>
</tr>
<tr>
<td>Manipulation</td>
<td>Co-opt</td>
<td>Importing influential constituents</td>
</tr>
<tr>
<td></td>
<td>Influence</td>
<td>Shaping values and criteria</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>Dominating institutional constituents and processes</td>
</tr>
</tbody>
</table>

Table 2. Strategic responses and tactics (Oliver 1991)

4 RESEARCH SETTING AND METHOD

4.1 The setting

According to Global Information Technology (GIT) reports from three consecutive years 2012, 2013, and 2014, Norway ranked among the top 10 advanced economies being the 7th, 5th, and 5th respectively (Schwab & Greenhill 2012; Bilbao-Osorio et al. 2013; Schwab & Eide 2014). Furthermore, GIT reports recorded Norway’s substantial positions in those three years as being a ready society for Information and Communication Technology (ICT) as measured by the Networked Readiness Index (NRI) in terms of environment friendliness for ICT, readiness to use ICT, actual usage of ICT by all stakeholders, and the impact of ICT on the economy and society. As demonstrated in Table 3, Norway recorded greater progress in 2014 than 2012 and 2013 in most of the readiness indices; however, the economic and social impact of ICT in Norway is not yet stable. In terms of readiness for cloud services, Norway is among the cloud prepared countries, according to a recent report by Cisco (Cisco 2013). Norway meets the minimum requirements for advanced cloud services in terms of key cloud infrastructure indicators are in place (i.e., hardware, network security, and Internet speed) (Information Economy Report 2013).

<table>
<thead>
<tr>
<th>Readiness index</th>
<th>Rankings over years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2012</td>
</tr>
<tr>
<td>Environment friendliness for ICT (i.e., political and regulatory, business and innovation)</td>
<td>10</td>
</tr>
<tr>
<td>Readiness for making good use of ICT (i.e., ICT infrastructure and accessibility of digital content, cost of accessing ICT, and basic educational skills for effective use of ICT)</td>
<td>6</td>
</tr>
<tr>
<td>Actual usage of ICT (i.e., individuals, businesses, and government)</td>
<td>6</td>
</tr>
<tr>
<td>Impact of ICT in gaining economic and social competitiveness</td>
<td>13</td>
</tr>
</tbody>
</table>

Table 3. The rankings of Norway across ICT readiness indices of the NRI. Sources: Schwab and Eide (2014), Schwab and Greenhill (2012), and Bilbao-Osorio et al. (2013)
4.2 The method

To explore how the adoption of CC in Norway is influenced and why organizations respond differently to institutional influences, the researchers conducted an exploratory case study to address the how and why questions (Yin 2009). Case study strategy is useful in understanding the dynamics of the context of the phenomenon under investigation (Darke et al. 1998; Yin 2009) and the human/technology interaction in the natural social setting (Orlikowski & Baroudi 1991). Two main sources provided the basis for this paper’s data: (1) 9 semi-structured interviews conducted through various communication modes; and (2) document analysis (i.e., research articles, reports, and official documents generated by Datatilsynet, the Norwegian government authority, and news articles). Interviewees represented various stakeholders (i.e., cloud service providers, consultancy, and customer organizations from the public sector). We chose the informants purposefully based on their knowledge about and/or experience with CC (Patton 1990; Palinkas et al. 2013) as demonstrated in their profiles, shown below in Table 4.

<table>
<thead>
<tr>
<th>Code</th>
<th>Position</th>
<th>Institution</th>
<th>Interview mode</th>
<th>Interview duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer 1</td>
<td>CIO</td>
<td>Municipality</td>
<td>Phone</td>
<td>25 min</td>
</tr>
<tr>
<td>Customer 2</td>
<td>Enterprise architect</td>
<td>Government regional authority</td>
<td>Microsoft Lync</td>
<td>44 min</td>
</tr>
<tr>
<td>Customer 3</td>
<td>IT advisor</td>
<td>Municipality</td>
<td>Face-to-face</td>
<td>44 min</td>
</tr>
<tr>
<td>Customer 4</td>
<td>CIO</td>
<td>Municipality</td>
<td>Phone</td>
<td>23 min</td>
</tr>
<tr>
<td>Customer 5</td>
<td>Assistant director</td>
<td>Government agency</td>
<td>Face-to-face</td>
<td>42 min</td>
</tr>
<tr>
<td>Customer 6</td>
<td>Technical project manager</td>
<td>Government agency</td>
<td>Face-to-face</td>
<td></td>
</tr>
<tr>
<td>Consultant 1</td>
<td>Senior security consultant</td>
<td>Public IT consulting company</td>
<td>Face-to-face</td>
<td>89 min</td>
</tr>
<tr>
<td>Consultant 2</td>
<td>Cloud advisor</td>
<td>Public IT consulting company</td>
<td>Skype</td>
<td>32 min</td>
</tr>
<tr>
<td>Provider 1</td>
<td>Product manager</td>
<td>Private IT services company</td>
<td>Face-to-face</td>
<td>56 min</td>
</tr>
</tbody>
</table>

Table 4. Informants’ profiles

5 FINDINGS AND ANALYSIS

5.1 Changes at the organizational field

This study aimed to explain the institutional influences that facilitate and/or constrain the adoption of CC in the Norwegian public sector, how public organizations strategically balance their need to move their IT environment to the cloud, and the requirements from both the legal and business environments. The defined organizational field for this study included all stakeholders (i.e., Ministry of Government Administration, Norwegian Data Protection Authority [Datatilsynet], professional consultancy, customer organizations from the public sector, and cloud providers). The findings from this study were framed within the isomorphic pressures (i.e., coercive, normative, and mimetic) to identify factors that influence the adoption of CC in Norway, followed by the identification of various adoption strategies by public organizations framed within the strategic responses (i.e., acquiescence, compromise, avoidance, defiance, and manipulation) with given rationales.

5.2 Isomorphic pressures

5.2.1 Coercive

After experiences with Narvik and Moss municipalities (Veum & Nymoen 2012; Veum & Thoreid 2012; O’Donoghue 2012), opinions about CC changed. These two municipalities used public SaaS cloud models; Narvik used Google Apps (i.e., e-mail and file sharing) and Moss used Microsoft Office 365 (Datatilsynet 2012). Moving to the public cloud was an expensive experience for Narvik, and an alert for Datatilsynet to safeguard the use of public clouds in municipalities. Consultant2 stated,
“two municipalities in Norway that moved to the public cloud last year; one to Microsoft and one to Google and they spent millions of kroners on lawyers in order to sort out the laws and regulations that were affected by moving to the public cloud...but of course they are early adopters so they had to use some money on it, and they got some good feedback and guidelines on how to approach the cloud, which was good after all!”.

Datatilsynet banned the use of Google Apps in Narvik after finding that Google’s data processor agreement did not fulfill information security requirements, according to the Norwegian law, regarding the location of the data, and Narvik did not conduct proper risk assessments (Veum & Thoreid 2012; Thon & Vetland 2012; Gould 2012). Datatilsynet had a concern about the U.S. Patriot Act and required Narvik to conduct risk assessment to get more information about data location, data segregation, and access controls (Gould 2012).

Nine months later, Datatilsynet decided to let Narvik municipality continue to use Google Apps (Essers 2012), but conditionally upon certain guidelines (Datatilsynet 2012). Moss municipality requested a review and guidelines from Datatilsynet for using Microsoft Office 365 (Veum & Nymoen 2012). These guidelines are explained by Consultant1,

“When they were finished with going through these cases [Narvik and Moss], we discovered three things that were very important: (1) it is required by the Norwegian law, risk analyses [before moving to the cloud]; (2) have an agreement [called databehandleravtale that is available on Datatilsynet’s home pages] with the cloud service provider with specific needs according to the Norwegian law. So you cannot use the agreement directly from Microsoft…and Microsoft knows about it...; (3) implement regular security audits of Microsoft if you choose Microsoft, and of Google if you choose Google. So Moss and Narvik, they now get information about security audits on Microsoft and Google”.

Furthermore, Consultant1 stressed that municipalities cannot use CC services without these three guidelines and that the responsibility is on the municipalities to ensure that the cloud provider processes the personal data according to the Norwegian data privacy regulations. If something went wrong with the data, the first entity will be questioned is the customer organization or the municipality. Therefore, for municipalities, CC is neither fully prohibited nor fully allowed, but they must be cautious as Provider1 stated,

“Datatilsynet, they say it is not by itself illegal to store data on a server outside Norway. But, you have to consider each case and make sure it is not sensitive data. So they don’t give a green flag for all types of scenarios”.

In addition to the guidelines offered to municipalities regarding the use of CC services, there were lessons learnt by the Norwegian legal system to upgrade and address the challenges brought by using CC services. In this regard, Consultant1 contended,

“These two cases [Narvik and Moss] are what we need to look at to see and compare with for future cases because every detail of the law is not regulated because cloud services systems are new things and they are a lot between different countries [laws] and there are so many new things for the Norwegian law and the court systems to go through, so they mainly take a few cases and they build next cases on laws”.

5.2.2 Normative

CC adoption in Norway is influenced by learning processes of exchanging knowledge about professional guidelines to deploy CC services. Norway established its own strategic initiative called ‘Digital Agenda for Norway’, putting CC among other digital goals; cloud services offer flexible and low-cost solutions for public and private sectors, as well as create jobs and businesses in Norway rapidly (Ministry of Local Government and Modernization 2013; Etro 2009). Accordingly, the Ministry of Government Administration, Reform, and Church Affairs encourages public agencies to procure cloud services (Ministry of Local Government and Modernization 2013).

CC adoption remains slow in the Norwegian public sector due to lack of experience and knowledge about CC among Norwegian organizations. This led the 10 largest Norwegian municipalities to form a
cooperation forum called K10\(^1\) in collaboration with the Norwegian Association of Local and Regional Authorities (KS) to speed up the adoption of cloud services in Norway (Hustad & Olsen 2012; Devoteam daVinci 2011). The purpose of the K10 forum is to exchange knowledge about CC, establish common municipal ICT architecture, and influence ICT policies of public sector agencies regarding procurement, standards, and cloud service models (Hustad & Olsen 2012; Devoteam daVinci 2011). Furthermore, Hustad and Olsen (2012) argued that the adoption of CC services in Norway will be driven strategically by business value rather than economically by low-cost decisions.

5.2.3 Mimetic

One study claimed that demonstrating successful stories of pioneer Norwegian organizations adopting CC solutions and growing significantly is a key motive to adopting CC in Norway (Hustad & Olsen 2012). Cloud providers can show their specialities in various CC service models (e.g., SaaS, PaaS, IaaS), references, and successful stories on their websites (Microsoft 2013; Xledger 2014) to gain credibility and convince potential customers to use their cloud services, especially when their reference customers are from the public sector or the field of heavy industries. Pure Norwegian cloud providers, such as JottaCloud and Evry, also demonstrate their compliance with the Norwegian legislation and privacy regulations, and their physical storage of data in Norway (Andino 2013; Evry 2014). This conveys their independence from the Patriot Act and makes it a competitive advantage for them to grow in the cloud market. Thus, customer organizations do not have to worry about the location of the data or if the foreign international cloud provider is complying with the Safe Harbor agreement. Regarding local providers, Consultant2 said that he understands the customer organizations’ fear of dealing with foreign international cloud providers. He asserted,

> “I think that goes more back to the rules and regulations and where data are physically stored. I have seen that Norwegian cloud providers, especially backup services and those kinds of things have got a lot of attention lately because then you don’t store data anywhere near the NSA”.

Thus, local cloud providers represent a safe option for some customer organizations to avoid inspection by Datatilsynet as happened with Narvik municipality, and ensure adherence to the Norwegian regulations.

5.3 Strategic responses

5.3.1 Acquiescence

At first, Narvik municipality used Google Apps without a thorough risk analysis and without a clear statement about the data location (Gould 2012); this implied that the municipality used the imitation tactic to unconsciously mimic the model of Google’s public CC services for the following reasons (Jørgenrud 2011): (1) efficiency: reducing maintenance costs to focus more on strategic issues, desiring mobile solutions, and seeking new functionality after their previous software environment, that is IBM Lotus Notes, became outdated and they lacked the right skills to maintain it; and (2) limited alternatives: Narvik sent requests to IBM, Microsoft, and Google, but only Google responded to their request. Therefore, Narvik chose to enter into an agreement with Google Apps.

After banning Narvik from using Google Apps, the municipality used the comply tactic to gain legitimacy by adhering to Datatilsynet’s restrictions through changing their contract with Google and obtaining additional assurances and information about location of the data from Google (Gould 2012). Moss municipality used the comply tactic by following consciously the guidelines they requested from Datatilsynet before beginning use of Microsoft Office 365 cloud to gain legitimacy (Veum & Nymoen 2012; Datatilsynet 2012). Furthermore, Moss chose Microsoft Office 365 public cloud for the sake of efficiency: improving work quality of their public employees, seeking flexibility in administering users, and avoiding the hassle of IT maintenance operations (Moss Kommune 2013).

---

\(^1\) The ten largest municipalities in Norway (i.e., Oslo, Bergen, Trondheim, Stavanger, Kristiansand, Drammen, Fredrikstad, Tromso, Sandnes, and Bærum) created K10 in 2010.
Another municipality used Microsoft Office 365; they used the comply tactic to gain legitimacy by following the legal requirements set by Datatilsynet, and they hired external companies to ensure this compliance. In this regard, Customer1 stated,

“We have some services in the cloud, we have Office 365. But, all employees store data on-premise. We have our own guidelines for risk evaluation [set by Datatilsynet for municipalities], but we have an external company that handles this for us today, but the most important thing is that we understand different risks like, for example, loss of data. And, the most important is how the data are stored”.

Another municipality used Microsoft cloud services at length and used the comply tactic, through which they unconsciously used the data processing agreement provided by Datatilsynet to gain legitimacy. In this regard, Customer4 asserted,

“My organization uses quite a few cloud services, especially, from Microsoft [Office 365] platform, and so we have been adopting cloud services for quite long time. Norwegian government organizations need, in Norwegian term, Databehandleravtale [data processing agreements]. Without them, we cannot use cloud services. I feel, at least, that we have to accept what we have got!”

Furthermore, they used CC services for efficiency purposes: automation of IT resources and serving the public efficiently. Customer4 said,

“We need to find a way [to use] public clouds to provide the service to the public in a new and more efficient way with focus on automation and to a larger extent than what we do on-premise”.

5.3.2 Compromise

Another government regional authority currently considering use of CC services has two trajectories planned toward adopting CC. In the first trajectory, they used the bargain tactic to negotiate sourcing strategies with sourcing partners to deliver PaaS and IaaS service models to their associated enterprise units. Their reason for doing so is efficiency: to simplify and regionalize their IT architecture. They are now describing the specifications of the future IT architecture to transition from a more distributed one to a more simplified and consolidated regionalized one. Customer2 explained,

“We haven’t moved to utilize cloud services yet. At the moment, we have a shared service provider [SSP] and we are in the phase of transitioning SSP to become more focused on the application services side, and the business process side…..we have a program at the moment called infrastructure modernization, so a part of its mandate is to look at sourcing strategies and sourcing partners to be able to deliver platform as a service and infrastructure as a service”.

Customer2 asserted that laws and regulations constrain their choice of CC sourcing partners, and they need to follow those regulations to gain legitimacy. The second trajectory towards adopting CC in this organization follows in the next sub-section on avoidance strategy.

5.3.3 Avoidance

In the abovementioned government regional authority, they did not transition to and use CC services; nonetheless; they used the conceal tactic. They aimed at efficiency by testing Microsoft Azure solution to evaluate its functionality for building mobile solutions and regional electronic public record solutions. Additionally, they are testing the CC solution to avoid facing complexity with the regional electronic public record because of its connection to many existing systems. In this regard, Customer2 stated,

“We have at least two evaluation initiatives, and proof-of-concept initiatives are ongoing. One of them is concerned with testing or proofing concept of the environment for mobility solutions where some of the storage of the data is done in the cloud…[where] some of the documentation which is generated by personnel…using iPads and the service we are using is Microsoft Azure in that environment, and also the other initiative is more in early stages; we are looking at how
can we utilize the cloud in the context of establishing a regional electronic [public] record solution which is centralized, consolidated, and standardized [across associated enterprise units]. And, that is too complicated because the electronic [public] record is integrated with many systems and that is why integration can pose a concern”.

5.3.4 Defiance

Another municipality did not initiate the adoption strategy and used the dismiss tactic by ignoring the idea of adopting CC services that are offered and discussed at a small conference, which was held to discuss opportunities in relation to the initiative of common municipal ICT architecture. The reasons for not reaching the proper strategy for using cloud services in this municipality was threefold: (1) loss of control: data sensitivity concern regarding data location; (2) conflicts: with small municipalities joining a big municipality, each has different interests and knowledge about the use of IT; and (3) bureaucracy: the multi-levelled structure of each municipality makes it slow to reach an agreement on any proposed strategy and the gains are not yet clear for them. Customer3 explained, when he was asked if his municipality is using cloud services or planning to do so,

“We haven’t a proper process around that yet. We have discussed, but we haven’t made any proper strategy around the cloud yet, mostly because of the data sensitivity. Where the data are stored…. If the data are stored in the USA, then we don’t have the confidentiality. In our municipality, the main issue is to have a wider cooperation with all [small] municipalities to gather the main components in one place. When a decision matter or the IT department works on a proposal for a strategy, then it must be delivered to the administrative and then the executive, and then it is sent upward to the politicians to be agreed or to be decided whether to do that or not. Those small municipalities have also the same structure and they have different interests, goals, and knowledge about IT and use of IT systems. So our job here is quite a challenge….Two years ago, I put up a small conference for all municipalities to discuss about CC and the vendors were invited two or three of them. We had a full day of discussion and presentations and none of the municipalities decided anything”.

Another reason for defiance is goal ambiguity; they are not sure about the exact benefit from CC services. When asked about whether the municipalities decided to make another meeting afterwards or not, Customer3 answered,

“No, because it is not clear what are the gains because it requires a lot and there is this issue of confidentiality. I have a kind of difficulty to give you a clear enough answer because I suppose we are still unclear ourselves”.

5.3.5 Manipulation

A Norwegian government agency used the influence tactic by choosing to implement a private PaaS environment to meet the specific nature of their applications that they develop and deploy to serve the public. They decided to implement their own development and testing environment for the following reasons: (1) lack of reference: current PaaS implementations are immature and not widely used in Norway; (2) independence: to gain full control over the testing environment and eliminate the dependency on external suppliers; and (3) image preservation: thorough testing of their type-specific applications is important for them to avoid headlines of newspapers and preserve their reputation. Customer6 explained that strategy,

“We are implementing now both infrastructure as a service layer and on top of that PaaS….Because we have looked at some commodity implementations of PaaS, no real resources with experiences locally in Norway at least. The current PaaS implementations, they are very App-specific. Ok I can deploy one App, but I have 50 [Apps]. That is not going to help that much, so we actually need either to script or automate against the [current] PaaS implementation”.

Customer5 highlighted the importance of having internal competence rather than depending on external suppliers and the matter of reputation. She said,
“It is not what we want. We get more restrictions. So, we need to make our own PaaS. To have people know what they are talking about, that is always a good solution. To be very dependent on external suppliers...it is really hard...So, you need to have the competence internally. The rules change every year in the [domain we operate in], so it is not a stable system. So, we have a focus on not facing the front page of the newspapers and not making any mistakes. Testing is very important, because we have an issue with reputation”.

In the same line, Customer6 agreed on the matter of reputation. He asserted,

“If there is a bug in the system, the receipt may be garbled. And, when we have sent it, it is too late to fix the system. It goes externally and in the worst case on the front page of the newspapers. So, we can’t risk it. So, we need to test very thoroughly. That is an important goal of cloud strategy or automation strategy”.

6 DISCUSSION AND CONCLUSIONS

By revisiting the two research questions from the onset, we discuss the corresponding answers, which are framed using concepts from the neo-institutional theory (i.e., isomorphic pressures and strategic responses to institutional processes). This paper presented a case study to explore the isomorphic pressures imposed on Norwegian public organizations by the constellation of various external actors (See Table 5). The adopting organizations interpreted these pressures differently, as mirrored in various chosen strategies to adopt CC services (See Table 6). The chosen strategic responses ranged from willingness to resistance against conformity to pressures. Furthermore, the paper explicated reasons behind each strategic response chosen. Efficiency stood out as a predominant reason among acquiescence, compromise, and avoidance strategies; this indicates rational choice. However, aside from the imitation tactic, this rationality is balanced with irrationality in the acquiescence and compromise strategies by seeking legitimacy.

<table>
<thead>
<tr>
<th>Pressure type</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
</table>
| Coercive      | - Banning the use of CC services in Narvik municipality based on legal assessment  
- Introducing guidelines for using CC services in municipalities (i.e., conducting risk assessment, having a data processing agreement with the cloud provider according to Norwegian law, implementing regular security audits) | Datatilsynet |
| Normative     | - Encouraging Norwegian public agencies to procure cloud services  
- Developing specifications for procuring CC services  
- Exchanging knowledge and establishing common Norwegian municipal ICT architecture with a focus on appropriate sourcing policies (i.e., procurement, standards, cloud SaaS, PaaS, IaaS) | Ministry of Government Administration, Reform and Church Affairs  
- Ten large Norwegian municipalities with the Norwegian Association of Local and Regional Authorities |
| Mimetic       | - Offering Various CC service models (i.e., SaaS, PaaS, IaaS)  
- Demonstrating successful stories online of pioneer organizations adopting CC services  
- Demonstrating compliance with the Norwegian legislation and privacy regulations | Cloud providers  
- Peer organizations  
- Pure Norwegian cloud providers |
<table>
<thead>
<tr>
<th>Adoption strategy</th>
<th>Tactic</th>
<th>Definition</th>
<th>Empirical evidence</th>
<th>Reasons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquiescence</td>
<td>Imitate</td>
<td>Either conscious or unconscious mimicry of institutional models</td>
<td>- When Narvik used Google Apps unconsciously without doing proper risk assessment and having a clear statement in the agreement about the location of the data</td>
<td>- Efficiency - Limited resources</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A conscious obedience to or incorporation of values, norms, or institutional requirements</td>
<td>- After being banned, Narvik changed their contract with Google to get more assurances so that Datatilsynet allowed them to continue using Google Apps</td>
<td>- Efficiency - Legitimacy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Moss followed Datatilsynet’s guidelines before using Microsoft Office 365</td>
<td>- Efficiency - Legitimacy</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Another municipality hired an external company to perform risk evaluation, as set by Datatilsynet, for using Microsoft Office 365</td>
<td>- Legitimacy</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Another municipality used the data processing agreement by Datatilsynet to use a few cloud services from Microsoft</td>
<td>- Efficiency - Legitimacy</td>
<td></td>
</tr>
<tr>
<td>Compromise</td>
<td>Bargain</td>
<td>The effort of the organization to exact some concessions from an external constituent in its demands or expectations</td>
<td>- In a Norwegian public authority, negotiating sourcing strategies to deliver IaaS and PaaS services to associated enterprise units</td>
<td>- Efficiency - Legitimacy</td>
</tr>
<tr>
<td>Avoidance</td>
<td>Conceal</td>
<td>Disguising nonconformity behind a facade of acquiescence</td>
<td>- In a Norwegian public authority, two initiatives for testing cloud solutions to evaluate their functionality: - Mobility solutions - Establish a regional electronic public record solution</td>
<td>- Efficiency - Complexity</td>
</tr>
<tr>
<td>Defiance</td>
<td>Dismiss</td>
<td>Ignoring institutional rules and values, when the external institutional rules are perceived to be low or when internal objectives conflict with institutional values or requirements</td>
<td>- Ignoring the idea of adopting cloud services despite a municipality held a conference to discuss opportunities</td>
<td>- Loss of control - Conflicts - Bureaucracy - Goals ambiguity</td>
</tr>
<tr>
<td>Manipulation</td>
<td>Influence</td>
<td>Shaping values and criteria of acceptable practices or performance</td>
<td>- A public administration authority chose to implement a private PaaS to meet the specific nature of their applications that they develop and deploy to serve the public</td>
<td>- Lack of references - Independence - Image preservation</td>
</tr>
</tbody>
</table>

Table 6. CC adoption strategies and tactics used by Norwegian public organizations

As seen in Table 5, the legal coercive pressure had a dominant influence on the adoption of cloud computing in the Norwegian public sector. The findings from this study indicated the direct effect of the coercive pressure exerted on the previous case of Narvik as public organizations became cautious regarding their strategies to adopt CC services. This manifested in the acquiescence strategy by using the standard agreement provided by Datatilsynet to procure public CC services along with conducting a thorough risk assessment. The compromise strategy resulted from interpreting coercive pressure and mimetic pressure. It enfolds bargaining between regulations and cloud providers who can meet those regulations to attain legitimacy.

Complexity appeared to be one of the key reasons for the avoidance strategy of full adoption of CC services offered by various cloud providers through the mimetic pressure. This manifested in proofing concept of CC solutions to avoid complexity issues that may arise from the interdependency of many...
system modules on public records. Mimetic pressures manifest in looking at reference peer organizations that are pioneers in the field, various service models offered by major cloud providers, and using pure Norwegian cloud providers because they demonstrate their adherence to Norwegian legislation. The mimetic pressure is responded by the influence tactic in the manipulation strategy. Despite the wide variety of CC services offered by cloud providers, they may not be widely used within a certain region or country. This happened with the public organization that chose the manipulation strategy; they decided to build their own cloud environment. Having internal competence allows them cope with their rapidly changing business rules and become less dependent on external providers. Apparently, the indirect effect of Narvik’s case spreading in the news headlines—is yet another reason for manipulation strategy.

The normative pressure aimed at exchanging knowledge and developing specifications for procuring CC services by public agencies. Although it is dismissed by some municipalities in a defiance strategy, the fear of losing control over data secrecy was a reason. However, the findings indicate that the dismissal was mainly caused by the bureaucracy, conflict of interests, and goal ambiguity. Hence, the defiance had no direct relation to the risks of the CC model per se.

According to Oliver (1991), the responses are triggered towards the isomorphic pressures as one-way interaction. However, the findings from this study showed that an organization’s response to some pressures could trigger further pressures on it. This happened with Narvik when it imitated Google’s Apps services without risk assessment. It triggered a new coercive pressure on it from Datatilsynet.

The insights offered by this study may be further compared with CC adoption pressures and strategies in other contexts. This can also extend to soliciting lessons from Norway as a developed-world country for use by those developing-world countries that are still immature to some extent, yet who are willing to innovate with CC in their public sector.

References


A USER-CENTRIC, PRIVACY-AWARE PLATFORM FOR NEW BUSINESS MODELS IN A DATA-DRIVEN ERA

Michael Petychakis, National Technical University of Athens, mpetyx@epu.ntua.gr
Iosif Alvertis, National Technical University of Athens, alvertisjo@epu.ntua.gr
Romanos Tsouroplis, National Technical University of Athens, rtsouroplis@epu.ntua.gr
Evmorfia Biliri, National Technical University of Athens, ebiliri@epu.ntua.gr
Fenareti Lampathaki, National Technical University of Athens, flamp@epu.ntua.gr
Dimitrios Askounis, National Technical University of Athens, askous@epu.ntua.gr
Timotheos Kastrinogiannis, VELTI S.A., tkastrinogiannis@velti.com
Andreas Daskalopoulos, VELTI S.A., adaskalopoulos@velti.com
Theodoros Michalareas, VELTI S.A., tmichalareas@velti.com
Eric Robson, TSSG Waterford Institute of Technology, erobson@tssg.org
Donal MacCarthy, TSSG Waterford Institute of Technology, dmccarthy@tssg.org
Christine O'Meara, TSSG Waterford Institute of Technology, comeara@tssg.org
Łukasz Radziwonowicz, Fraunhofer FOKUS, lukasz.radziwonowicz@fokus.fraunhofer.de
Robert Kleinfeld, Fraunhofer FOKUS, robert.kleinfeld@fokus.fraunhofer.de

Abstract

Today’s mobile application consumers’ rapidly increasing demand for innovative tools and services to support their personal and business needs, has amplified the use of cloud-based services which, through their publicly available APIs, disrupt traditional industries and provide a vast number of features indicatively spanning from location-based services, photo sharing, video/music streaming and recommendations to data storage, data syncing and social networking. To address the multiplicity complexity and sustainability of the emerging cloud-based mobile apps ecosystem, from both consumers’ and application developers’ point of view, OPENi project developed an innovative open-source Graph API platform that enables mobile application consumers to store data and contextual metadata from their mobile application usage in their own space in the cloud, namely the "Cloudlet". This information, along with fruitful dynamic contextual data, can be then shared (under the control of the consumers) securely among their applications, services and across connected devices, towards optimizing end-users’ overall quality of experience (QoE). This platform enables new business models for service providers and builds trustful relationships among data owners-users, application developers and prospective service providers, via an enhanced privacy mechanism.

Keywords: Cloud-based Mobile Applications, Generic Graph API, Context, Personal Data Privacy, Privacy-by-design.

1 INTRODUCTION

In recent years, the proliferation of cloud-based, social-driven and mobile applications has disrupted the traditional enterprise of mobile, web and desktop applications with two major phenomena. First of all, the fragmentation of users’ data across different data silos that hinder mobile applications’ development and sustainability. Equally important is the migration of part of the generated value
outside applications, onto multiple Cloud-based Services (CBS) and social networks, where API-based connectivity is limited and with no efficient support for user context. What is more, the control and use of end users’ personal digital data has migrated to multiple CBSs outside the realm of control of the traditional enterprises (e.g. media), application developers and most crucially, end-consumers.

Motivated by the latter emerging drawbacks, OPENi (Git Repository 2015) has delivered a novel consumer-centric, privacy-by-design, open source, cloud-based development platform, serving as a catalyst for new applications era. This paper aims at pointing out a new business model where end-users own and control their data, developers build applications on a distributed authorization mechanism (i.e. no central authority, thus less business risks) and enterprises can host such services in bundled telecommunication organisations, e.g. cloud hosting provided by carriers.

2 DISRUPTION IN PERSONAL INFORMATION INDUSTRY

Personal information is a cornerstone of economic activity required to fulfill commercial transactions but also having an intangible value in itself. The purchase and sale of data is a well-established and profitable industry. For example, the US data brokerage industry is worth in the region of $15bn (Datacoup - Reclaim your personal data. 2015). Traditional data brokers have been selling and renting customer data, such as contact details, for many decades. This activity is foundation for the phone promotion, market research and direct mail marketing business, but also supplies information for other industries such as background checks and credit ratings. Data exchanges perform data collection & selling activities on a much larger and more sophisticated scale. The players of note operate globally and include (Credit Report & FICO® Score powered by Experian. 2015) and Bluekai (BlueKai | Big Data for Marketing | Oracle Marketing Cloud. 2015). Data is collected from multiple sources and includes significant amounts of information amassed electronically (e.g. cookie tracking). Typically, neither data brokers nor data exchanges remunerate the data subject. The large Internet companies, such as Facebook and Google, typify a more recent data entrepreneur. These organisations offer free services to subscribers and in return drive profits through leveraging access to consumers and consumer information. As with the data brokers, typically there is no direct value exchange between consumer and organization in return for consumer data.

However, these legacy business models may be under threat as the personal digital data sector is experiencing significant disruption, with the emergence of innovative new propositions (TNW - Aral Balkan - Free Is A Lie | The Next Web - YouTube. 2015). The forces for change can be loosely categorized under the headings consumer, regulator and enterprise (Figure 1).

![Figure 1. Forces changing typical business models dealing with personal data.](image_url)

Specifically related to added consumers, they are increasingly aware of their loss of privacy in the Internet era; In a recently published PEW report (Public Perceptions of Privacy and Security in the Post-Snowden Era | Pew Research Center's Internet & American Life Project. 2015), 91% of adults agreed or strongly agreed that consumers had lost control over personal information collected and used by companies. Moreover, 80% of those who use social networking sites say they are concerned about third parties -like advertisers- accessing data they share on these services. The emergence of new tools, such as the Epic browser (Epic Privacy Browser, a more secure and private chromium-based
web browser. 2015) or DuchDuckGo (DuckDuckGo. 2015) search engine, is a response to the increasing demand for privacy. However, individuals are aware of the value of their data and often willing to compromise on privacy in return for benefits, i.e. they expect to derive some value from their own data. 55% of respondents in the aforementioned PEW research indicated that they would be willing to share some information about themselves with companies in order to use online services for free. Other pain-points experienced by individuals, as they transact and interact increasingly online, include headaches associated with digital identity and asset management. As an example though individuals may be concerned about privacy, it is neither practical nor possible for them to examine the privacy policy of every site or service that they use. Furthermore, online fraud is on the increase and individuals are increasingly concerned about cybercrime and identity theft. However, though numerous surveys confirm that there is a growing unease among consumers about their data, mass adoption of new services is still some way off. The 2013 UK MiiData Pilot (The midata Innovation Opportunity / Research / Ctrl-Shift. 2015) in this area identified some critical factors for adoption, which are: (a) data must flow, thus incorporate multiple compliant sources, (b) consumer participation is essential, (c) needs-driven approach is better than data-driven focus at early market stage, (d) the proposition should focus on value, not on data, (e) convenience is important, (f) customer control is a benefit in its own right, and (g) trust and safety are of fundamental importance. Relatively to active regulators, EU-based data regulation is in the final part of its journey to implementation and for the first time all member states will be bound by a common legal instrument, designed to safe guard the interests of individuals. Some key features include the requirement of consent, high levels of transparency with explicit communication on all uses/sharing of data, need to keep data within EU jurisdiction, right to be forgotten and hefty fines for non-compliance (up to 2% of annual turnover). In the US this year, the FTC published its report summarizing a major investigation into the activities of 9 data brokers and made its strongest recommendations yet in the realm of data protection. At US governmental, level Smart Disclosure (Smart Disclosure Policy - Data.gov. 2015) is a policy initiative designed to help individuals access personal information in formats they can use and consequently enabling them to make better decisions in areas such as health.

The newly emerging PIMS (Personal Information Management Services) seeks to rebalance the personal data value chain more in favour of data producers (i.e. individuals). New World Players recognise the shifting dynamics in the market place, particularly more active and informed consumers (data producer) that seek to protect, manage and leverage their digital identities and growing enterprise demand for new ways to build trusted relationships with customers around their data (Feuerlicht, G., & Tran, H. T. 2014). There are many players entering the sector with ctrl-shift estimating that one new entity a week is launching (Mapping the Market for Personal Data Management / Research / Ctrl-Shift. 2015). The myriad of services on offer is extremely varied and can be loosely categorised according to a few types of proposition: (a) Storage & Utility, (b) Transparency & Trust, (c) Marketing & CRM Tools and (d) New World Data Traders. However, there is much activity in the space, the market is still extremely fragmented and there is no clear market leader among the market entrants as organisations look to solve the multiple challenges including disparate views of the individuals and legacy platform integration challenges.

3 PERSONAL CLOUDLETS EXPOSED UNDER A UNIFIED API PLATFORM

OPENi (see Figure 2) is built on personal cloudlets, where users can store data, either on their own cloud hosting service or through service provided by carriers. In that way, mobile users are the ultimate owners of their personal data generated, accessed and managed by their mobile applications, either directly or indirectly (i.e., via the integration of an app with multiple users’ cloud based services, like Facebook, twitter, etc.), as well as the controllers of the access permissions rules ruling the use of the latter. On top of that, a set of personal cloudlets are accessible through a unified Graph API framework (Alvertis, I., Petchakis, M., Lampathaki, F., Askounis, D., & Kastrinogiannis, T. 2014), which is responsible to handle API calls, manage data and connect with existing cloud-based services; this platform is managed by a central, trusted authority that may be a user community or the service provider again. Data models available in a personal cloudlet can be further be extended by platform developers, in order to enhance new applications, based on models not available in advance
(Petychakis, M., Alvertis, I., Biliri, E., Tsouroplis, R., Lampathaki, F., & Askounis, D. 2014); again every additional model allows users to own and control new models handled by applications and developers easily design and integrate existing cloud-based features into their applications. This approach reduces data and services fragmentation, development and maintenance effort and time to deliver through a purely web compatible open-source framework based on the REST paradigm (Fielding, R. T. 2000). Then, service enablers (see Figure 3) offer native, common functionalities to developers to reduce development time and effort, and increase intelligence of their applications; a service enabler is a trusted agent running on the OPENi framework that allows a group of services be developed that otherwise could not, by providing it with adequate power, means, opportunity, or authority to function (e.g. recommendations, advertising targeting, application analytics etc.).

Everything is managed through a common user interface and authorization mechanism, which allows developers to ask and get specific permissions for specific resources, while application users and data owners can overview and manage all authorized applications, cloud-based services and service enablers in a central interface; in that way, the trust among the data owners and the service providers is enhanced through common sense and transparent rules.

Figure 2. OPENi layering of services and stakeholders.

To that end, OPENi delivers all the necessary cloud components to allow the users to create, deploy and manage their Cloudlets, providing qualities such as data storage, discoverability-addressability-access by applications and a user-controlled privacy and security framework. Developers can find an important community where they can build applications fast and easy, by using the Graph API framework, extend models with the API builder, combine new features through service enablers and use other cloud-based services through the same API calls, just by changing some properties in the API calls. The enabler of all of them can be a service provider, like telecommunication carriers or public organizations, who hosts multiple personal cloudlets under the API framework, builds new business models with their customers (i.e. platforms users and developers).

Figure 3. Forces changing typical business models dealing with personal data
4 DEMONSTRATION DESCRIPTION AND OPERATION

The key-feature of the developed OPENi architecture is the enhanced privacy with user-centric authorization mechanism. As part of the demonstration, OPENi demo applications include OPENi Privacy Control Visualization Framework (see Figure 4) and demonstrate the following innovative user-centric privacy-aware features: (I) A permissions dialog and visualization interface that enables users to visualize, monitor and get useful insights on the use of their personal data from OPENi enabled applications. OPENi Privacy Control Visualization Framework allows users to navigate easily and efficiently in the permission visualization interface, hiding the complexity and multiplicity of the privacy framework end-user options. (II) A personal data management (opt-in & opt-out) interface that enables users to control in a transparent way the use of the personal data via clear and easy to understand opt-in and opt-out methods in a layered and structured way. (III) A fine-grained privacy control interface, that enable users to control the use of their personal data in user-friendly and fine-grained per-application access control model, being able to determine the user/access of a) groups of data/object types, b) groups of data/objects, c) exact data/objects and d) data/objects specific attribute, all stored in the Cloudlet.

![Figure 4. OPENi Privacy Control Visualization Framework](image)

See OPENi demonstrators (Figure 5) acts as both i) end-user privacy-awareness educational tools for end-consumers and application developers, and ii) real-time usability trials that allow useful insights on how users interact with OPENi visualization and privacy framework, leading to its fine-tuning and the extraction of useful results with respect to potential business models for its exploitation.

![Figure 5. Indicative OPENi demo applications](image)

5 CONCLUSIONS AND FUTURE CHALLENGES

The OPENi platform gives users a personal cloud repository, which is transferable and expandable over cloud (i.e. dynamic data support) with full control and ownership to their data, without disabling new offerings by third party developers or external cloud-based services. Application developers can
find new, stable communities of users, through a unified, simple and interoperable Graph API Framework, they can expand its functionality via a community-oriented builder, build contextually empowered applications and get advanced capabilities though Service Enablers and integrated third-party cloud-based services. Prospective OPENi service providers can become a platform for applications developers, and thus enable additional business models; innovative, strong multi-segment business models are now available to personal cloudllet providers, as strategic advantage against powerful monopolies, like Facebook or Google, and a key feature to maintain users in their existing businesses. The great challenge of course is the effort to build a community of users that starts using such services, and then support through training platforms, courses and disseminations actions a community of developers who will build useful, innovative applications; or vice versa.

Acknowledgements

This work has been created in the context of the EU-funded project OPENi (Open-Source, Web-Based, Framework for Integrating Applications with Social Media Services and Personal Cloudlets), Contract No: FP7-ICT-317883.

References

Alvertis, I., Petychakis, M., Lampathaki, F., Askounis, D., & Kastrinogiannis, T. 2014. 'A community-based, Graph API framework to integrate and orchestrate cloud-based services'. In Computer Systems and Applications (AICCSA), 2014 IEEE/ACS 11th International Conference on (pp. 485-492). IEEE.


Digital Services, Social Media and Digital Collaboration

Chairs: Dr. Gianluigi Viscusi
**Table of Contents**

**Track – Digital Services, Social Media and Digital Collaboration**

2. [PID-27] Young Jin Lee, Arvind Tripathi, Amit Basu and Ananth Srinivasan, “AN EMPIRICAL ANALYSIS OF INTRA-TRANSACTIONAL DISCLOSURE IN INTERNET AUCTIONS”
3. [PID-57] Brad McKenna, Michael Myers and Lesley Gardner, “ANALYSING QUALITATIVE DATA FROM VIRTUAL WORLDS: USING IMAGES AND TEXT MINING”
4. [PID-58] Shang Gao, Zhe Zang and Hong Guo, “A STUDY ON USERS’ ATTITUDE TO PERVERSIVE GAMES IN CHINA”
9. [PID-105] Nikolaos Basias, Marinos Themistocleous and Vincenzo Morabito, “AN INNOVATIVE DECISION MAKING FRAMEWORK FOR E-BANKING INTEGRATION”
The Role of Intensity of Facebook Usage in Social Capital Development: An Example with Arab Students

Kathy Shen, University of Wollongong in Dubai, United Arab Emirates, kathyshen@uowdubai.ac.ae
Shahper Vodanovich, Auckland University of Technology, New Zealand, svodanov@aut.ac.nz
Mohammed Khalifa, Al-Ghurair University, United Arab Emirates, khalifa@agu.ac.ae

Abstract

This study extends prior literature on social networking sites, e.g., Facebook, to an under researched context, Arab countries. In particular, this research compares the Arab and non-Arab youth in the UAE in terms of motivations, intensity of Facebook usage and their social capital development. With a survey study involving 123 students in the UAE, the results suggest that the need for socialization is a driving force for intensity of Facebook usage among the youth in the UAE, implying that the driving forces of Facebook usage are regional specific. Moreover, intensity of Facebook usage contributes more to bonding social capital rather than bridging social capital. Significant gender and ethnic differences are reported in intensity of Facebook usage and resulting bonding social capital.

Keywords

Youth, Social Networking Sites, Social Capital, Facebook Usage, Arab Countries

INTRODUCTION

Social network sites (SNS) constitute an important research area for scholars interested in examining the social impact of online technologies on the youth of today (Khedo et al., 2013, Brandtzaeg, 2012, Pempek et al., 2009, Valkenburg et al., 2006). Social networking sites such as, Facebook, MySpace and Bebo are member-based communities that allow users to post profile information and to communicate with others in innovative ways such as sending public and private online messages and sharing photos online. In 2008, Facebook reported having 67 million active users, with more than half of them returning daily and spending an average of 20 minutes per day on the site (Park et al., 2009).

Pempek et al (2009) found that Facebook, in particular, provides new avenues for young adults to express themselves and to interact with one another. Among young adults, relationships with peers are important for generating offline benefits, commonly referred to as social capital. Social capital is a construct used to describe the benefits one receives from one’s relationships with other people (Lin, 1999). Ellison et al (2007) suggest that intense Facebook use is closely related to the formation and maintenance of social capital. In their survey of undergraduates at a university, Facebook use was found to be associated with a distinct measure of social capital, including bridging social capital and bonding social capital.

The popularity of social networking sites is a global phenomenon, but the research on this phenomenon has centered in developed countries. Among Arab countries, youth are the fastest growing segment of populations. With a median age of 22 compared to a global average age of 28, the Middle East is one of the most youthful regions in the world (UNDP, 2009). Recently, with the “Arab Spring” (Aday et al., 2013), SNS usage, and its implications in the Arab world, has started attracting international research interest (AlSayed and Guvenc, 2013, Aday et al., 2013, Axford, 2011). However, most research has centered on the media consumption and communications for political reasons (Tufekci and Wilson, 2012, Aday et al., 2013). The area of social media use for social capital purposes by the youth of the Arab World is under-researched.

By examining the young adults’ use of social networking sites and social capital development in Arab countries, the present study contributes to prior work in three key areas. It not only reexamines and confirms existing research on the relationship between intensity of Facebook usage and social capital (Valenzuela et al., 2009c) in a new context but also demonstrates the relative contribution of Facebook usage on different types of social capital development among young adults in Arab countries. Secondly, it extends existing research by elaborating on the main driving forces for Facebook usage (Park et al., 2009) that are specific for the young adults in Arab countries. And finally it reveals...
interesting gender and ethnic differences/similarities in motivations, intensity of Facebook usage and resulting social capital development, which to a great extent are counter intuitive. For example, Valenzuela et al (2009c) found ethnic differences in the relationship between intensity of Facebook usage and social capital building between white and non-white users. This research endeavors to extend this line of research to the Arab context and compare the difference between Arab vs. non-Arab users.

In the following sections, we will review the prior literature and develop hypotheses. This is followed by a description of the empirical study. Next, we report and discuss the results. The paper is concluded by a discussion about implications, limitations and future research.

THEORETICAL DEVELOPMENT AND HYPOTHESES

This section reviews existing literature in this area and outlines the motivation for hypothesis development. These hypotheses seek to confirm, extend and challenge existing notions about Facebook use by Young people in Arab countries. Facebook has a number of features that allow individuals to communicate with others. This particular site for example, allows users to construct an identity for themselves through the information they post about themselves in terms of pictures, videos and other media for their friends to view and comment on. Research is needed to examine these friend-networking sites for numerous reasons. One of the main reasons is to understand the personal and social motivations that youth have for their intense use of Facebook. Intensity of Facebook usage in this research describes the extent to which a user considers Facebook as a part of his/her life and implies the psychological engagement in Facebook interactions.

The uses and gratifications approach (Katz and Blumler, 1974, Ruggiero, 2000) provides a useful framework for drawing these inferences. This approach is a useful framework from which to understand Internet usage and user’s needs. The uses and gratifications theory is concerned with how individuals use the media (in this case, Facebook), and therefore it emphasizes the importance of the individual. This theory is particularly useful as it draws a distinction between concepts that are antecedents to behavior (e.g., uses and gratifications sought) and those that are consequents of behavior (e.g., gratifications obtained) (Bonds-Raacke, 2008).

Numerous studies have been done that support the notion that Facebook is used primarily for social connections (Pempek et al., 2009, Sheldon, 2008). For example, Sheldon (2008) reported that students used Facebook to maintain relationships with people they already knew and only a small number of students used Facebook to meet new people. In addition, a survey of over 2,000 students by Lampe et al (2006), found evidence that the primary use of Facebook was for ‘social searching’ – that is to use Facebook to find out more information about someone they met offline, in class, a social setting or an acquaintance. The use of Facebook for ‘social browsing’, for instance, to meet someone via the site with the intention of a later offline meeting, or to attend an event organized online, scored relatively low amongst their sample.

Moreover, Facebook can fulfill the informational needs of users, young users are facing the period of adolescence in a digital environment, this is typically a period of their lives are dominated by many questions, questions about oneself and questions about changing relationships with the outside world. In this phase of life there is a struggle for independence from parents and an increased reliance on peers for support (Buhrmester and Furman 1987; Larson and Richards 1991). However, there are some areas of a young person’s life that they may not feel comfortable sharing with even their closest friends. For these reasons the Internet in general has become an example of an important informational source that has become increasingly popular among young users (Tapscott 1997; Prensky 2001), it would be interesting to explore Facebook as an information resource for youth, given its ubiquity in this age group. In order to keep users updated about their social circles, Facebook has two features: “News Feed”, which appears on each user’s homepage, and “Mini-Feed”, which appears in each individual’s profile. “News Feed” updates a personalized list of news stories throughout the day generated by the activity of friends (Valenzuela et al., 2009a). Thus, each time users log in, they get the latest updates about their contacts. “Mini-Feed” is similar, except that it centers around one individual. Thus, Facebook use can reinforce existing ties and communities by keeping users constantly updated about what is going on with their contacts (Hargittai, 2007). On the other hand, Facebook allows users to create and to join groups based around common interests and activities by incorporating their profiles into the Facebook Groups application. The Groups application displays each individual’s group memberships as well as groups their friends have joined.

Users can log in to Facebook to satisfy needs of pure entertainment and recreation, for example a popular application for Facebook is “FunWall,” which allows users to post a much broader range of content than the traditional “The Wall,” such as games, videos, and music. Posting links to YouTube on the FunWall or spending time creating Facebook Groups can foster a sense of customization and enjoyment but drive attention away from the real world (Valenzuela et al., 2009a). As Shah and colleagues (2001) have argued, “in such cases, recreation and socializing may become privatized
Social capital particularly in the context of social networking sites like Facebook is about the resources available to people through their social interactions (Lin and Lu, 2000, Putnam, 2001). Generally it is perceived that people derive many benefits of a diverse range of social interaction, indicating that the more interaction a person has the more they benefit. Although people often accumulate social capital as a result of their daily interactions with friends, coworkers, and strangers, it is also possible to make conscious investments in social interaction (Resnick et al., 1993). Many researchers have alluded to the fact that this is the main reason that users join sites like Facebook; is to maintain and increase their social networks (Ellison et al., 2007, Joinson, 2008, Ellison et al., 2011).

The types of interactions have also become increasingly important, which is to say that it is not necessarily the technology but the way the technology is used that is important. There seems to be a general bias that social interactions based around informational needs and community building are positively associated with individual-level production of social capital, while use of social media for purely entertainment purposes are negatively associated with social capital (Wellman et al., 2001, Ellison et al., 2011). This explains why online activities have been found to both reduce and increase social capital.

Social capital has broadly been described as two constructs - bridging and bonding social capital. Bridging capital occurs between individuals who are weakly tied together, this type of capital provides access to novel information or new perspectives but not typically emotional support (Putnam, 2001). Alternatively, bonding social capital is found between individuals in tightly-knit, emotionally close relationships, such as family and close friends. The Internet has been linked both to increases and decreases in social capital (Nie et al., 2002, Wellman et al., 2001). As mentioned before bridging social capital could be enhanced by sites that support loose social ties which allow members to create large diverse networks, while bonding capital could be enhanced by sites that allow more meaningful exchanges to occur (Boyd and Ellison, 2008, Ellison et al., 2011). Facebook in our view could usefully support both bridging and bonding capital. Therefore we examine the following hypotheses:

H4: Intensity of Facebook usage has positive impact on bridging social capital
H5: Intensity of Facebook usage has positive impact on bonding social capital

With the rise of popular SNS like Facebook, comes a vital new area of research examining the relationship between gender and online communication. Social networking sites are one particular avenue for self-expression that is popular among young adults (Hargittai, 2010). Therefore, the ways in which individuals communicate and interact through social networking sites are inevitably a reflection of their identity, including gender.

Cooper (2006) found that males were more comfortable with the use of technologies than their female counterparts, Lin and Overbaugh (2009) and Caspi, Chajut et al. (2008) assert that females are more prolific, productive and effective in online environments. Interestingly, Johnson (2011) concluded that in the short term females might be in a better position to benefit from the Internet due to their cognitive orientation to ubiquitous spaces, while in the long term males will begin to use the Internet more for communication purposes and thereby enhance their developmental outcomes as well. Research that has been conducted in this area supports this assumption that some Face to Face gender norms are replicated in online communication. For instance, Valkenburg, Schouten, and Peter (2005) concluded that both males and females often use gender stereotyping in their presentation of self online. Because of this, there have been well-established findings that suggest that female online behaviour is more interpersonally oriented, while males are more task and information oriented (Haferkamp et al., 2012, Muniz and O’Guinn, 2001). Such findings suggest that motivation to use Facebook may vary between males and females.

Moreover, Agarwai (2000) found that within social networks, females form stronger kinship and friendship relations than males. Therefore, social capital theory would expect that although males may accumulate more friends to widen their network and improve their “status,” the average time spent on Facebook per friend would be higher for females because
of their tendency to establish closer bonds with in a smaller group.

**H6: There is a gender difference in motivations, intensity of Facebook usage and social capital development**

While there has been a recent interest in the use of various social media platforms in the Arab world for political purposes (Ghannam, 2011, Harb, 2011) the area of social media use for social capital purposes by the youth of the Arab World is under-researched. The Arab world has always been an early adopter of “mass media” technology. Egypt, in particular, is perceived as a leader in the adoption of technology and also as the country with the largest number of Internet users in the Arab world (Abdulla, 2007). The UAE continues to lead the Arab world in adoption of information and communication technology, according to a World Economic Forum report, and is expected to spend about $3.3 billion on information technology and communications hardware for schools, hospitals and other civil projects between 2008 and 2011. According to the Internet Telecommunication Union, Internet penetration in the UAE has increased from 36 percent in 2006 to an estimated 64 percent in 2008 (Initiative, 2009). UAE has approximately 5.8 million Internet users and just over 3.4 million Facebook users (Statistics, 2012).

Although the youth have more access to the Internet than their older counterparts, Internet access for youth in the region is still quite low in comparative perspective (WorldBank, 2010). Moreover, a recent UNICEF report found that the youth (18-24) tend to identify closely with traditional, religious, and familial ties. For example, the youth in Jordan still identify strongly with the army and the police over independent media, while the Lebanese youth demonstrate tight sectarian attachments. Given the lack of research done for this important cohort, our research will focus on the intensity of Facebook usage for social capital development between Arab and non-Arab youth. It would be interesting to explore whether or not the online uses of the Internet by these two cohorts reflect the offline differences in their cultures.

**H7: There is difference in motivations, intensity of Facebook usage and social capital development between Arab and non-Arab youth**

**RESEARCH METHOD**

We conducted a survey in the UAE to examine the following above mentioned hypotheses, i.e. motivations in using Facebook, intensity of using Facebook, and resulting social capital. According to the UAE National Bureau of Statistics, UAE has about 88% expatriates coming from more than 200 countries. Given the small amount of UAE nationals in the total population, expatriates are less likely influenced by the nationals but rather keep their distinctive characteristics. Thus, UAE offers a good opportunity to understand the diversity among Facebook users. The survey was distributed in two rounds involving undergraduate students from two universities. The first round was conducted in a private university where the majority of students were from an international background, from more than 50 nationalities. With limited respondents from Arab countries, we had the second round of data collection in a national university. Participation was voluntary and without any motivation. In total, we collected 123 valid responses with 57 participants from Arab countries and 65 participants from non-Arab countries. Table 1 summarizes the demographic information of the participants. About 90% of respondents were adolescents or emerging adults, that is, below 30 years old, according to Erikson's Stages of Psychosocial Development. This is consistent with the other statistics about the social media users. On average, respondents reported 363 “friends” on Facebook with a great standard deviation of 447. In this research, we use Arab countries to refer to 22 Arabic-speaking countries of the Arab League. Particularly, respondents from Arab countries had significant fewer “friends” (Mean=164) than those from non-Arab countries (Mean=538), although both groups had similar Facebook membership duration.

<table>
<thead>
<tr>
<th>Age</th>
<th>Count (%)</th>
<th>Origin of Respondents</th>
<th>Count (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>13-20</td>
<td>25.3</td>
<td>Arab countries</td>
<td>46.3</td>
</tr>
<tr>
<td>21-30</td>
<td>65</td>
<td>Non-Arab countries</td>
<td>53.7</td>
</tr>
<tr>
<td>&gt;30</td>
<td>9.7</td>
<td>Gender</td>
<td>Count (%)</td>
</tr>
</tbody>
</table>
### Frequency of Login

<table>
<thead>
<tr>
<th>Frequency of Login</th>
<th>Count (%)</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>More than once daily</td>
<td>43.1</td>
<td>61</td>
<td>39</td>
</tr>
<tr>
<td>Once daily</td>
<td>16.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Several times a week</td>
<td>17.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Once a week</td>
<td>4.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than weekly, but more than</td>
<td>8.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Once a month</td>
<td>11.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Membership with Facebook</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Count (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>More than one but under two</td>
<td>1.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between two and six months</td>
<td>1.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between six months and one year</td>
<td>4.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>More than one year, less than two</td>
<td>14.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>More than two years</td>
<td>78</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Measures

All measures were adapted from prior literature. In particular, bonding social capital was measured by adapting the existing scales from (Ellison et al., 2007) to capture Internet-specific social capital (Greenhow and Burton, 2011; Steinfield et al., 2008a). A sample question was “There are several people on Facebook I trust to solve my problems” (1=strongly disagree; 7=strongly agree). Bridging social capital was measured with the scale adopted from (Brandtzæg, 2012). A sample question was “My Facebook friends have a different cultural background or race from me” (1=strongly disagree; 7=strongly agree).

Various motivations of using Facebook were measured by the items developed by a previous scale (Vasaloua et al., 2010). A sample item was “I use Facebook to meet interesting people” (1=strongly disagree; 7=strongly agree). Facebook usage intensity was measured with the scale from (Steinfield et al., 2008b). A sample item was “Facebook is part of my everyday activity” (1=strongly disagree; 7=strongly agree).

### Data Analysis

Measurement validation and model testing were done in a holistic manner using Partial Least Squares (PLS). An advantage of using PLS is that the resampling technique establishes confidence intervals based on repeated samples from the researcher's own data rather than on assumptions, such as multivariate normal distributions. Thus, the normality of the survey data will not influence the PLS results. Tests of significance were conducted for all paths using the bootstrap re-sampling procedure and the standard approach for evaluation that requires path loadings from construct to measures to exceed 0.70. Internal consistency of reflective measures was checked with composite reliability measures ($\rho$) and average variance extracted (AVE), as suggested by Fornell and Larcker (1987). The discriminant validity was examined by comparing the square root of the AVE for a particular construct to its correlations with the other constructs (Fornell and Larcker, 1987) and by examining cross-loadings of the constructs.

This study adopted a cross-sectional design and both independent and dependent variables were measured at the same time point, implying that the common method variance could be a major threat for the validity. The control for the common method variance was done first through instrument design by using different scales and randomizing the sequence of the questions, both of which have been shown to effectively reduce the common method variance. In addition, according to Harman’s single-factor test (Podsakoff and Organ, 1986), common method variance is present if a single factor accounts for the majority of the covariance in the dependent and independent variables. The principle component analysis with all variables results in six distinct factors, and the first factor only accounted for 17.5% of the variance, implying no substantial common method bias in our survey data.

---


2 [http://www.dsg.ae/en/Publication/Pdf_En/424201311017185100000.pdf](http://www.dsg.ae/en/Publication/Pdf_En/424201311017185100000.pdf)

RESULTS AND DISCUSSION

Table 2 presents the loadings of the reflective measures to their respective constructs along with AVE value, composite reliability scores, and t-statistics from the PLS analysis. All reflective items are significant at the 99% level with high loadings (all except for two items above 0.70 and most above 0.80), thereby demonstrating convergent scores validity of all. The latent constructs are higher than the recommended value of 0.80 (Nunnally, 1978), demonstrating internal consistency.

Table 2: Measurement Model

<table>
<thead>
<tr>
<th></th>
<th>Loading</th>
<th>T-Test</th>
<th></th>
<th>Loading</th>
<th>T-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensity of Facebook Usage (AVE=0.83)</td>
<td>Item 1</td>
<td>-0.72</td>
<td>9.77</td>
<td>Item 1</td>
<td>0.80</td>
</tr>
<tr>
<td>Item 2</td>
<td>0.88</td>
<td>35.44</td>
<td>Item 2</td>
<td>0.79</td>
<td>11.06</td>
</tr>
<tr>
<td>Item 3</td>
<td>0.64</td>
<td>7.92</td>
<td>Item 3</td>
<td>0.84</td>
<td>13.46</td>
</tr>
<tr>
<td>Item 4</td>
<td>0.91</td>
<td>43.60</td>
<td>Item 4</td>
<td>0.79</td>
<td>12.18</td>
</tr>
<tr>
<td>Item 5</td>
<td>0.87</td>
<td>29.06</td>
<td>Bridging Social Capital (AVE=0.88)</td>
<td>Item 1</td>
<td>0.71</td>
</tr>
<tr>
<td>Item 6</td>
<td>0.82</td>
<td>16.28</td>
<td>Item 1</td>
<td>0.78</td>
<td>3.71</td>
</tr>
<tr>
<td>Item 7</td>
<td>0.81</td>
<td>17.40</td>
<td>Item 2</td>
<td>0.77</td>
<td>4.05</td>
</tr>
<tr>
<td>Item 1</td>
<td>0.92</td>
<td>40.81</td>
<td>Item 4</td>
<td>0.78</td>
<td>3.95</td>
</tr>
<tr>
<td>Item 2</td>
<td>0.92</td>
<td>33.22</td>
<td>Item 1</td>
<td>0.83</td>
<td>3.27</td>
</tr>
<tr>
<td>Item 3</td>
<td>0.88</td>
<td>20.76</td>
<td>Motivation-Socializing (AVE=0.84)</td>
<td>Item 3</td>
<td>0.83</td>
</tr>
<tr>
<td>Motivation-Information Seeking</td>
<td>Item 1</td>
<td>0.95</td>
<td>7.13</td>
<td>Item 2</td>
<td>0.65</td>
</tr>
<tr>
<td>Item 2</td>
<td>0.97</td>
<td>7.11</td>
<td>Item 3</td>
<td>0.83</td>
<td>12.98</td>
</tr>
<tr>
<td>Item 3</td>
<td>0.75</td>
<td>4.26</td>
<td>Item 4</td>
<td>0.84</td>
<td>17.03</td>
</tr>
</tbody>
</table>

Table 3 presents the discriminant validity statistics. The square roots of the AVE scores (diagonal elements of Table 3) are all higher than the correlations among the constructs, demonstrating discriminant validity. Furthermore, all items loaded higher on their respective constructs than on others, providing additional support for discriminant validity.

Table 3: Discriminant Validity

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation-Entertainment (1)</td>
<td>0.91</td>
<td>0.39</td>
<td>0.39</td>
<td>0.39</td>
<td>0.39</td>
<td>0.39</td>
</tr>
<tr>
<td>Motivation-Information seeking (2)</td>
<td>0.91</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
</tr>
<tr>
<td>Intensity of Facebook usage (3)</td>
<td>0.39</td>
<td>0.39</td>
<td>0.76</td>
<td>0.76</td>
<td>0.76</td>
<td>0.76</td>
</tr>
<tr>
<td>Bonding social capital (4)</td>
<td>0.17</td>
<td>0.20</td>
<td>0.34</td>
<td>0.34</td>
<td>0.34</td>
<td>0.34</td>
</tr>
<tr>
<td>Bridging social capital (5)</td>
<td>0.06</td>
<td>0.06</td>
<td>0.23</td>
<td>0.23</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td>Motivation-Socializing (6)</td>
<td>0.60</td>
<td>0.48</td>
<td>0.42</td>
<td>0.42</td>
<td>0.42</td>
<td>0.42</td>
</tr>
</tbody>
</table>

Figure 1 presents the results of the PLS analysis of the structural model, including the overall explanatory
power (R2) and path coefficients (for relationships between latent variables). The research model provides good explanatory power for intensity of Facebook usage (20.9%). Among all three different needs, i.e., socialization (H1), information search (H2) and entertainment (H3), only the need for socialization was found to have a significant effect $\beta=0.244$; on $p<0.01$. It seems that UAE students mainly use Facebook for socialization, rather than entertainment or searching for information.

As for the consequences of Facebook usage, our results showed that intensity of Facebook usage has a significant $\beta=0.349$; effect $p<0.01$ on bonding social capital (12.1%) than bridging social capital (5.2%). The coefficient between intensity of Facebook usage and bridging social capital was significant $\beta=0.227$; $p<0.05$. Thus, our results in general are consistent with prior research regarding the positive impact of Facebook usage on social capital development, supporting H4 and H5. However, the usage of Facebook for social capital development is not balanced between bonding and bridging social capital. In this research, we found that the UAE students mainly construct bonding social capital from using Facebook rather than bridging social capital. The difference in the relative importance of intensity of Facebook usage in social capital development reflects the underlying usage pattern of specific populations.

In this study, intensity of Facebook usage mediates the effect of various motivations on social capital development. Generally speaking, there are criteria used to informally judge whether mediation is occurring: (1) the independent variable has a significant effect on the mediator, (2) the independent variable also significantly affects the dependent variable in the absence of the mediator, (3) the mediator has a significant unique effect on the dependent variable, and (4) the effect of the independent variable on the dependent variable is reduced after the mediator is added to the model. Based on the above criteria, we only tested the potential mediating effect of intensity of Facebook usage on the linkage between need for socialization and bonding social capital development, since the direct effect of the need for socialization on bridging social capital was not significant. Following the procedure suggested by (Sobel, 1982) and (Baron and Kenny, 1986), it was found the effect of need for socialization on bonding social capital development is completely mediated through intensity of Facebook usage (Sobel test=2; $p<0.05$).
To test H6 and H7, we first performed an independent t-test to examine the gender and nationality (Arab vs. non-Arab) differences. We did not find significant gender difference in all three motivations, which is different from prior research that suggests female users are more interpersonal oriented and male users are information oriented (Jackson et al., 2001). Although female users (Mean=2.89) tend to logon to Facebook more frequently than male users (Mean=1.94), female users do not feel strongly affiliated with Facebook or consider Facebook as a part of their daily life. On contrary, male users reported significantly more intensive Facebook usage (Mean=4.37; p<0.01) and bonding social capital (Mean=4.39; p<0.01) than female users (Mean for intensity of Facebook usage=3.76; Mean for bonding social capital=3.51). This result is quite different from existing findings that suggest females tend to form stronger kinship and friendship relations than males in social networking sites (e.g., Agrawal 2000), suggesting that gender difference is not universal.

One plausible explanation for male users having more intensity of Facebook usage and developing more bonding social capital is that males users may find Facebook interaction is more reciprocal than offline face-to-face interaction (Peter and Valkenburg, 2006) whereas female users in the UAE are more keen on face-to-face interaction. Finally, there was no significant gender difference in bridging social capital. Considering the prevalence of religious and traditional cultural influence in the UAE, this result is also counter-intuitive in that both male and female users developed social networks with a similar level of diversity. Particularly, female users enjoy the same openness as male users in connecting with different people. This finding suggests that cultural differences in gender are not reflected in an online environment. To further illustrate the gender impact, we also reported the structural model with gender differences (see Figure 2). The insignificant linkages are not included.
As for the difference between Arab and non-Arab users, similar results as gender differences were reported in that there was no significant difference between Arab and non-Arab users in motivations, suggesting the universal needs for using Facebook. But we found significant difference in intensity of Facebook usage and resulting bonding social capital in that non-Arab users reported less frequency in login Facebook (Mean=1.89; p<0.01) but stronger intensity of Facebook usage (Mean=4.37; p<0.01) than Arab users (Mean for login frequency=3.25; Mean for intensity of Facebook usage=3.57). This could be due to the fact that most non-Arab users in the UAE have been using Facebook for much longer time than Arab users in the UAE. Moreover, non-Arab users in the UAE are more likely to rely on Facebook to maintain their existing social networks than Arab ones. Subsequently, non-Arab users also reported higher level of bonding social capital (Mean=4.16; p<0.01) than Arab users (Mean=3.5) on Facebook. This may be related to non-Arab users' overall long history of using Facebook to keep their existing connections with whom they can trust and count on; while for Arab users, relatively speaking, they are more inclined to use face-to-face for close social connections. Finally, there was no significant difference between Arab and non-Arab users in bridging social capital, although both groups of users reported significantly higher level of bridging social capital (Mean=4.88) than bonding social capital (Mean=3.85). To further illustrate the impact of nationality, we also reported the structural model with nationality differences (see Figure 3). The insignificant linkages are not included.

![Structural Model Results and the Gender Differences](image-url)
Need for Socialization 0.253*** Intensity of Facebook Usage $R^2=0.253$

Need for Information 0.051ns Bridging Social Capital $R^2=0.052$

*: $p<0.1$; **: $p<0.05$; ***: $p<0.01$

Figure 3: Structural Model Results and the Effect of Nationality
CONCLUSION, LIMITATIONS AND FUTURE RESEARCH

In this research, we examine Facebook usage for social capital development in the UAE by young adults, a context that was under studied in prior research on social media. Our results entail theoretical as well as practical implications.

First, the results confirm existing research on the relationship between intensity of Facebook usage and social capital (Valenzuela et al., 2009c) and extend existing research by demonstrating the relative contribution of intensity of Facebook usage on different types of social capital. Facebook not only helps young adults extend their social networks to incorporate diverse connections but also enables them to develop strong ties with selected connections. Intensity of Facebook usage in this research describes the extent to which a user considers Facebook as a part of his/her life and implies the psychological engagement in Facebook interactions. The relative importance of intensity of Facebook usage in different types of social capital development implies that users with high engagement in Facebook interaction are more likely to maintain and develop bonding social capital over Facebook. However, such engagement, although significant for bridging social capital, may not be a major reason for users to develop bridging social capital. It is more likely that bonding and bridging social capital may be subject to different sets of Facebook behaviors, which warrants future research.

Second, this research also provides empirical evidence about the main driving forces for the intensity of Facebook usage (Park et al., 2009) is the UAE context. UAE is a traditional Islamic country with a large population of expatriates from all over the world. The co-existence of youth from multiple cultural backgrounds creates a unique environment for us, from which to examine the universality of motivations to use Facebook. Our findings only support the need for socialization as a driving force for intensity of Facebook usage among the youth in the UAE, implying that the driving forces of Facebook usage are likely vary for different social and cultural contexts. In the future, more cross-cultural studies would be helpful to illuminate the regional distinctiveness in the motivations for Facebook usage.

Finally, adding to the existing evidence about ethnic differences in the relationship between intensity of Facebook usage and social capital building between white and non-white users (Valenzuela et al., 2009b), this research extends this line of research to demonstrate the significant difference between Arab and non-Arab users in intensity of Facebook usage as well as bonding social capital development on Facebook. Moreover, considering strong cultural connotation of gender differences in Arab countries, our results about the gender differences in intensity of Facebook usage and social capital development provide much updated evidence about the users in Arab countries.

As one of the early attempts to examine Facebook users in Arab countries, this research is not without limitations. Particularly, with a cross-sectional design, our empirical results cannot prove causality but correlation. In future research, it would be interesting to examine the evolvement of social capital development associated with Facebook usage. It is likely that more social capital reaped from Facebook interactions will encourage users to get more involved. With a longitudinal design, such dynamics can be well investigated.

In addition, the cross-sectional design also limits our interpretation of the ethnic/gender differences. For example, are such differences inherent and systematic or temporary? Is it possible that the difference between Arab and non-Arab users or gender difference is related to the different stages of Facebook appropriation? In the long run, are Facebook users getting more homogeneous in their usage pattern? Answers to such questions are beyond the scope of this research. But our results imply the necessity to examine the evolvement of Facebook users across different cultural and social contexts. Such research will provide more insight about how Facebook is appropriated over time in diversified contexts.
Finally, it will be interesting to replicate this study in different countries, particularly those Arabic countries with less diversity than UAE. Although we did not consider impact from external environment, such contextual influences indeed shape the social media behaviour and appropriateness. Thus, the future research can extend our model to account for the macro environmental impact.

REFERENCES


AN EMPIRICAL ANALYSIS OF INTRA-TRANSACTIONAL DISCLOSURE IN INTERNET AUCTIONS

Young Jin Lee, University of Denver, USA, email: youngjin.lee@du.edu
Arvind Tripathi, University of Auckland, New Zealand, email: a.tripathi@auckland.ac.nz
Amit Basu, Southern Methodist University, USA, email: abasu@mail.cox.smu.edu
Ananth Srinivasan, University of Auckland, New Zealand, email: a.srinivasan@auckland.ac.nz

Abstract

Despite a substantial literature on authentication mechanisms for online transactions, the impact of intra-transaction information disclosure through question and answer (Q&A) processes has not been explored much. In this paper, we report on the empirical analysis of a large online auction site that supports Q&A processes for buyers and sellers. Our analysis shows that these Q&A processes are significant determinants of auction success. This study provides valuable insights for auction operators in terms of effective online auction mechanism design.

Keywords: Internet auctions; Information asymmetry; Platform design

1 INTRODUCTION

The challenges of online authentication are well known. Transactions in Consumer-Consumer (C2C) online marketplaces are most affected by this, since potential buyers are largely dependent on online mechanisms to estimate the quality of not only the offered products but also the sellers themselves [3]. Further, low barriers to entry for participation in online marketplaces leads to buyers often faced with large numbers of seemingly similar offerings. These problems are exacerbated in online auction marketplaces, where the success of a transaction requires ongoing engagement of participants over substantial periods of time.

These problems have motivated a variety of studies in recent years on the identification and analysis of key facilitators of successful outcomes in online marketplaces. It has been argued that feedback and reviews about products, services, vendors, and other participants might play an important role in enhancing market performance. Examples of such work include personalization of review information (e.g.,[33]), impact of reviews on sales (e.g., [7]); impact of reviews on online auction prices (e.g., [26]); impact of feedback on trust formation (e.g., [28]), and estimating the economic impact of reviews (e.g., [15]). Much of this work emphasizes the combination of quantifiable as well as textual components of user-generated content to examine their impact. In the specific case of internet auctions, it is common to publicize seller and other participant ratings based on their past behavior as authentication signaling to potential buyers. Along with reviews, which are typically textual in nature, such mechanisms typically are built over time and reflect a history of past behavior.

Most of these authentication mechanisms are based on information generated either before or after a transaction. For example, ratings are provided based on user experience of past transactions, product descriptions are provided by a seller as part of information disclosure prior to a transaction, and product reviews are provided by buyers after the transaction has concluded.

In this research, we focus on signaling that is generated while an online auction is in progress. It is our contention that such information enhances participant interest, provides for better disclosure on the part of sellers and other participants, and ultimately increases market efficiency. In order to examine this issue, we examine auction data from a large online auction site. Auctions on this auction site include a Q&A feature that allows potential buyers to pose questions and receive answers from sellers while an auction is in progress. Further, this interaction is publicly viewable by other potential buyers who benefit from the availability of such intra-transaction disclosure. By analyzing auction listing data, as well as data on the intra-transaction disclosure discussed above, we empirically test for the
effects of this signaling mechanism on auction outcomes. Our research goal therefore is to examine the impact of information disclosure while an auction is in progress on auction outcomes.

The rest of the paper is organized as follows. In Section 2, we review the literature that examines various issues pertaining to online marketplaces. This enables us to position our work in context. In section 3, we describe our data collection procedures, the models that were tested and present the results with their analysis and interpretation. Section 4 discusses the implications of these results for the design of auction platforms and raises some important implications of our work while presenting a roadmap of some of the future directions that we intend to pursue.

2 Prior Work

2.1 Influence of Textual Online Word-of-Mouth

Word-of-mouth can be a significant driver of consumer behavior and influence consumer judgment and choice of products by affecting consumer awareness and preferences [18], [16]. While there is a rich tradition of examining word-of-mouth as an area of enquiry, renewed interest in the area is largely due to the influence of current technologies [7]. For example, [6] argues the online version of word-of-mouth has been given new significance since the internet’s capability of bidirectional communication enables individuals to not only receive information from organizations but also make their thoughts and opinions accessible to the global community and therefore create a large-scale network where opinions about a large pool of topics are shared. Dellarocas et al [7] also point out that word-of-mouth now has unprecedented scalability and speed of diffusion, as well as persistence and measurability.

One version of digital word-of-mouth is the online product review, which can influence consumer choices and product sales, as well as revenue forecasting and planning ([33], [5], [7]). Senecal and Nantel [33] suggest that personalized information about product reviews and recommendations have a strong effect on consumers’ online product choices, and Chevalier and Mayzlin [5] found that an increase in the number and length of reviews as well as an improvement in reviews’ star rankings lead to an increase in online book sales. Also, Dellarocas, Awad and Zhang [7] developed a model to forecast motion picture revenue using statistics of online movie reviews posted by users on Yahoo! Movies during the first week of a new movie’s release.

Seller feedback mechanisms in online electronic markets have also drawn much attention from researchers as another form of online word-of-mouth ([6], [30], [26]). This technology has a wider impact on organizations as it allows trust-building, product development, quality assurance and many other activities in online marketplaces, and has emerged as a viable mechanism for fostering cooperation amongst strangers in online marketplaces.

Although the potentially useful role of textual content of user feedback has been suggested [2], many early researchers have focused on numerical information such as star ratings and polarized seller/buyer feedback rating. However, Ghose & Ipeirotis [14] examine the influence of the actual textual content of product reviews from Amazon.com, and show that in terms of consumer-oriented helpfulness, looking at feature-based goods, users prefer mainly objective information with a few subjective sentences. On the other hand, for experience-based goods, users prefer a brief description of objective elements and a personalized, highly sentimental positioning of the goods, not captured by the product description (mixtures of objective sentences and subjective content are highly rated). In terms of manufacturer-oriented sales generation, an increase in average subjectivity of review text leads to increased sales [15] as it incorporates additional measures of text reviews such as informativeness, readability and linguistic correctness.

2.2 Performance of Online Auction Sites

Gefen, Karahanna & Straub [12] state that the physical separation of online auction participants makes it difficult for buyers to use social cues such as physical interaction and body language to assess seller quality. Buyers in the online context can only assess product quality through the online interface rather
than testing goods’ quality by “kicking the tires”, which also falls short in explicitly differentiating complex products from heterogeneous goods ([27], [10]).

Online auction sites frequently employ a feedback system to facilitate trust between seller and buyer. The feedback systems offer the buyer an opportunity to quantify the reputation of the seller, or vice versa, when an auction concludes successfully [22], and allows people to view individuals’ previous auction comments and history [35]. Chen and Wilson [4] state that feedback regarding a particular seller can help buyers to improve their decision making and increase the odds of a successful auction with reduced risk. Various studies including [2], [9], and [28] demonstrate that reputation systems with feedback ratings and text comments provide enriched information on seller quality and therefore reduce uncertainty about the seller.

Ghose [13] demonstrates that the textual feedback generated by buyers and displayed in a user friendly manner can facilitate trust between sellers and buyers. Online auction sites need to utilize this mechanism and implement robust reputation systems that can explicitly display multi-dimensions of a seller’s characteristics, such as customer service, product representation, packaging and shipping. According to seller reputation theory, the information signals about the product in an online auction are dependent upon the reputation of its seller ([24], [34]). One of the major findings of [10] shows that seller uncertainty influences product uncertainty, and seller information signals exhibit complementarity with product information signals.

Recent studies, such as [8], [31] and [10] have applied the concept of product diagnosticity [23] and information signaling to mitigate information uncertainty, mainly by providing the buyers clear and visible signals. Buyers feel confident in assessing product quality if they feel the product information or description is diagnostic, as in [29] and [19].

2.3 Information Disclosure

Information disclosure is described as an activity or a procedure to reveal relevant or essential information to specific audiences with certain degrees of intention [1]. The positive relationship between the clarity of seller provided description and the possibility of a better outcome stimulates seller intention to disclose high quality information to be disseminated among the customer base [21]. Also, consumer beliefs about the historical performance of the seller have a direct effect on outcomes. There is a positive relationship between existing goodwill of sellers and potential profitability, and sellers are encouraged to selectively disclose information, reflecting that goodwill, to achieve a greater profitability [20]. Jin and Leslie [21] argue that in the case of disclosing and propagating information about innovative items, the ice-breakers have greater possibilities of gaining a large portion of market share and obtaining competitive advantages over other competitors, through the timely disclosure of essential information.

Product description has received much attention by researchers, and is deemed to be necessary information disclosure in order to mitigate negative issues of online auction markets. Dimoka and Pavlou [10] argue that high quality combinations of textual, visual and multi-media based product description can competently reveal the details of the item being sold. In a second-hand product auction situation, the product description has even more significance, and Lewis [25] argues that the disclosure of seller private information can help to form a verifiable and enforceable contract. In this case the term ‘private information’ puts more focus on the seller’s own specific experiences with the item, which can be both positive and negative. For example the identification of a dent on a used car by its seller in an auction would improve the quality of information disclosure and thus bring the seller benefit for being honest and thorough, rather than intentionally omitting the defectiveness about the car.

Seller-initiated information disclosure in an online auction also includes the transaction price (start price, reserve price, buy now price) and the third-party certification such as product inspection, historical records and warranties. These types of information serve the same purpose as the product description and the seller private information, namely enriching the auction webpage and mitigating the uncertainty of the specific auction. In online auctions for used goods, buyers mostly receive
information rather than publish it. However, the bi-directionality of the internet enables buyers to be sources of information disclosure [6].

We examine two types of buyer information disclosure in online auctions. The first category of buyer-initiated information disclosure is seller feedback from the reputation systems. The feedback information disclosed can be a numerical star rating, or it can be textual comments that identify the seller’s outstanding or abysmal behavior [28]. The impact of such feedback information, is significant in establishing trust and fostering cooperation, as well as affecting other buyers’ purchasing decisions through the influences from the social word-of-mouth network [11]. Another mechanism through which buyers can disclose information in online auctions is the questions that they pose to the seller. Many online auction sites offer potential buyers the chance to engage the seller with questions about the particular transaction. In their study of the online conversations, Godes & Mayzlin [16] recognized the need to understand such interactions.

However, on the impact of buyer-posed questions and the interactions they trigger in an online auction market has not been examined in the literature. Buyer-posed questions require sellers to respond and thus make a verifiable disclosure on the particular auction webpage. Previous studies have argued that the information disclosed in an auction can mitigate the problem of adverse selection if such information can be verified [17]. Buyer questions during an auction can trigger the disclosure of the seller information, and thus share the ability to define an explicit contract, which can later be enforced.

In contrast to information from feedback mechanisms, buyer-posed questions and seller responses provide a much more timely and relevant input to ongoing transactions. The questions are asked and answered before the auction closes and therefore the information disclosed from such interaction is influential on the current auction. This disclosure can be a substantial performance predictor for transactions, because the questions and seller responses are focused and transaction-oriented, instead of the accumulated seller reputation scores and comments which are directed at the seller rather than the particular auction.

Although voluntarily disclosed product information can be used to form explicit contracts and mitigate product uncertainties, it has some disadvantages in terms of enhancing disclosure compared to (intra-transactional) buyer-posed questions. First, the nature of product description is not ‘bi-directional’ and thus the description can hardly be referred as an interaction, i.e. only the seller can post the product description to the buyers and not the other way around. Second, voluntary information disclosure by the seller can be considered as “passive” information from the perspective of potential buyers in online auction markets. The reason is that the content of this disclosure is solely decided upon by the seller. It is what the seller deems to be important about the product, or what the seller wants to show the potential buyers. Such information is not actively triggered by potential buyers and thus may not be a true representation of what product information a buyer considers necessary to be disclosed. In other words, the potential buyers have no choice but to passively consume the content from the product description and seller private information. On the other hand, the questions asked during an auction are the bidirectional interactions which are actively posed by the buyers. They represent the perfect example of buyer active disclosure and are hypothesized to have the ability to improve the overall information disclosure, since they facilitate the disclosure of information that the buyers really desire.

Thus, an open question is whether intra-transaction information in the form of Q&A interaction between buyers and sellers has a significant impact upon bidders’ participation, which in turn affects the success and outcome quality of online auctions? It is true that a higher number of bids has a significant and positive effect on the likelihood of meeting the auction reserve. Also, for ascending price auctions, prior studies have shown that a higher number of bids is likely to increase the final winning bid Pinker et al. [32]. A higher number of bids indicates higher bidding activity or competition among bidders to win the item. Bidders’ participation is measured as the number of bids in an auction. To address this question, this study aims to investigate specific aspects of the Q&A process that are the key determinants of bidders’ participation in online auctions. The purpose of the current study is to empirically address this specific problem.
3 METHOD, RESULTS AND ANALYSIS

3.1 Data Collection

The data used in this study was collected from a large online auction platform based in New Zealand. Like all other auction houses, bidders’ registration is required before they can participate in any auction. Bidders’ registration allows bidders to be tracked uniquely across multiple auctions. Sellers also need to register on the site, which allows us to track sellers and their activities within and across different product categories. Registration does not cost anything to individuals and hence does not provide any impediment to participants. Authenticated API (application programming interfaces) calls provided by the auction platform were used to collect auction data over a four week period between December 2011 and January 2012. The data that we collected span four categories of products: antiques, books, clothing, and used cars. We wanted the data to span a range of values of price and product complexity to represent the diversity of product offerings. A total of 1112 usable listings were obtained from this effort.

The focus of this study is on the role of intra-auction interactions between seller and potential buyers in an online auction. The Q&A feature of these auctions allows any potential bidder to ask a question pertaining to this auction, to which the seller can respond. This question and answer exchange is displayed on the auction page and provides valuable information to all the auction participants. The questions may relate to product features, type, functionality and concerns related to transaction as a whole. The data that was collected on each auction included various details of the Q&A process, such as the length and time of the questions posed and answers provided, as well as the registered identifiers of the individuals who asked the questions. Our analysis of the data was done in two parts. In the first study, we examined whether seller feedback ratings moderate the effect of intra-auction interactions on bidders’ participation. In the second study we investigated if auction listings that are rich in information moderate effect of intra-auction interactions on bidders’ participation. A listing’s information richness was measured by the number of photos posted of the offered item. A higher number of photos provide more information the product to the potential bidders and hence can enhance the authentication of the product’s features and quality, thereby potentially affecting bidders’ participation. We only focused on listings with at least two bids. The rationale for this design is explained in later sections. Tables 1 and 2 show the characteristics of the data set, and the descriptive statistics of the data collected are provided in Table 2. These variables include traditional variables that are known to affect outcomes such as seller reputation and product descriptors in addition to intra-transaction variables so that we can control for their effects on outcomes.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description/Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>MemberFBCount</td>
<td>Net positive feedback rating for a seller from prior transactions on this auction site.</td>
</tr>
<tr>
<td>SellerCommentCount</td>
<td>Total number of comments made by the seller during an auction. Seller can make comments during an auction, irrespective of questions asked.</td>
</tr>
<tr>
<td>QuestionCount</td>
<td>Total number of question asked during an auction.</td>
</tr>
<tr>
<td>AnswerCount</td>
<td>Total number of answers during an auction (some questions are left unanswered before auction ends).</td>
</tr>
<tr>
<td>PhotoCount</td>
<td>Number of pictures of the item for sale, posted by the seller.</td>
</tr>
<tr>
<td>BidCount</td>
<td>Total number of bids placed in an auction</td>
</tr>
<tr>
<td>SellerQuestionCount</td>
<td>Total number of questions asked by a seller in other auctions</td>
</tr>
</tbody>
</table>

Table 1. Variable Descriptions
Variables | Study 1 (Effect of Feedback Ratings) | Study 2 (Effect of Listing Richness)  
--- | --- | --- 
MemberFBCount | Dataset A (Feedback ratings below median) | Dataset C (Listing Richness below median) | Dataset D (Listing Richness above median) 
SellerCommentCount | 0.24 | 0.11 | 0.36 
QuestionCount | 1.75 | 0.81 | 1.93 
AnswerCount | 1.55 | 0.67 | 1.73 
PhotoCount | 3.18 | 0.99 | 5.85 
SellerQuestionCount_AllCat | 0.44 | 1.13 | 0.91 
BidCount | 19.12 | 15.50 | 25.81 

Table 2. Descriptive Statistics (Means)

3.2 Analysis and Results

This section presents the results of two studies designed to examine the impact of the intra-transaction Q&A process on the number of bids in an auction. Both the studies use BidCount as the dependent variable.

3.2.1 Study 1: Effect of Seller Feedback Ratings

In recent years as the online auction market has gained more attention, the number of listings on auction platforms has dramatically increased. However, even with this huge increase in supply, there has been only a modest increase in demand for items sold via online auctions. With an imbalance in demand and supply, it is not surprising that many auctions in online auction platforms do not get any attention/participation from bidders. Since the auctions we consider are ascending price auctions, a higher number of bids in an auction usually increases the final price as well as the likelihood of a successful auction (which we define as an auction in which the seller’s reserve price is met, and the item sold). Auction platforms, such as the one we consider, facilitate a dialogue between the seller and potential bidders by offering a Q&A feature for each auction. The objective of this feature is to build confidence in the product’s quality and authenticity, which could potentially increase bidders’ participation in these auctions. Hence, we aim to examine if intra-auction interaction does indeed increase the number of bids (Bidcount) in an online auction.

Prior research has shown that seller feedback ratings can be used as a trust building mechanism in online auction markets. Sellers with higher feedback ratings are more trusted and hence, bidders are more likely to bid on auctions from sellers with higher feedback ratings. Consequently, in our first study, we examine if seller feedback ratings moderate the effect of intra-auction interaction on bidders’ participation, as measured by the number of bids in an auction.

The model that we estimate is as follows:

\[
\text{BidCount} = b_0 + b_1 \cdot \text{AnswerCount} + b_2 \cdot \text{SellerCommentCount} + b_3 \cdot \text{PhotoCount} + 
\]

\[
b_4 \cdot \text{MemberFBCount} + b_5 \cdot \text{QuestionCount} + b_6 \cdot \text{SellerQuestionCount_AllCat} \quad (1)
\]

Seller feedback ratings vary by large numbers. Sellers who have been selling on the platform for a long time accumulate higher net positive rating scores compared to newbies or individuals who don’t sell that often. One can argue that the effect of Q&A process on bidders’ participation may vary due to higher and lower feedback ratings. We split our dataset on median seller feedback ratings – creating two datasets, one with feedback ratings below median value (dataset A) and the other above median value (dataset B). We estimate our model in equation (1) for both the datasets. Table 3 and 4 show the results for dataset A and tables 5 and 6 show the results for dataset B.
A stepwise regression model for dataset A yielded an $R^2$ value of 0.183 ($F=30.30$, $p=0.000$). The model characteristics and components are shown below.

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>62901.257</td>
<td>3</td>
<td>20967.086</td>
<td>30.30</td>
<td>.000</td>
</tr>
<tr>
<td>Residual</td>
<td>280943.221</td>
<td>406</td>
<td>691.978</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>343844.478</td>
<td>409</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 3: ANOVA Results, Study 1(Seller Feedback Ratings below Median)**

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
</tr>
<tr>
<td>(Constant)</td>
<td>7.822</td>
<td>2.169</td>
<td>3.606</td>
</tr>
<tr>
<td>AnswerCount</td>
<td>3.211</td>
<td>0.676</td>
<td>0.238</td>
</tr>
<tr>
<td>SellerCommentCount</td>
<td>5.972</td>
<td>1.521</td>
<td>0.188</td>
</tr>
<tr>
<td>PhotoCount</td>
<td>1.234</td>
<td>0.404</td>
<td>0.149</td>
</tr>
</tbody>
</table>

Dependent Variable: BidCount

**Table 4: Model Coefficients**

Results from Tables 3 and 4 show that for sellers with low feedback ratings, the number of answers that sellers provide, sellers’ comments and listing richness (number of photos of the items) all have a significant effect on the number of bids (BidCount) in an auction. These results provide insights and better understanding of the auction dynamics. As expected, we find that higher listing richness can increase the number of bids in an auction. More photos provide better product information, or in other words, reduce information uncertainty regarding the product. One can argue buyers’ trust in a seller is low when feedback rating score is lower [2]. We observe that auction listings from lower feedback rating sellers seem to invite more questions via the Q&A process. We also observe that not all the questions get answered before the close of the auctions. Bidders looking for more information in these listings pay close attention to information (answers) provided by sellers. We find that number of answers provided by the seller has a positive impact on number of bids in these listings. Further, we also find that number of comments posted by the seller has a positive effect on bidders’ participation. Comments allow sellers to clarify their previous answers or provide more information regarding the item for sale. We conjecture that these listings lack information about the product and auction as a whole due to insufficient seller experience. These sellers later use comments to supplement the missing information in listings and in their answers.

To test the effect of the Q&A process on bidders’ participation in listings from higher rated sellers, we estimate the model shown in equation 1 using dataset B. A stepwise regression model for dataset B yielded an $R^2$ value of 0.195 ($F=29.49$, $p=0.000$). The model characteristics and components are shown below.
### Table 5: ANOVA Results, Study 1 (Seller Feedback Ratings above Median)

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>64730.261</td>
<td>3</td>
<td>21576.754</td>
<td>29.49</td>
<td>.000</td>
</tr>
<tr>
<td>Residual</td>
<td>267789.242</td>
<td>366</td>
<td>731.665</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>332519.503</td>
<td>369</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 6: Model Coefficients

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
</tr>
<tr>
<td>(Constant)</td>
<td>4.486</td>
<td>2.813</td>
<td>1.595</td>
</tr>
<tr>
<td>QuestionCount</td>
<td>4.493</td>
<td>0.921</td>
<td>0.238</td>
</tr>
<tr>
<td>MemberFBCount</td>
<td>0.002</td>
<td>0.000</td>
<td>0.234</td>
</tr>
<tr>
<td>PhotoCount</td>
<td>2.122</td>
<td>0.370</td>
<td>0.285</td>
</tr>
</tbody>
</table>

Dependent Variable: BidCount

The results in Tables 5 and 6 indicate that for listings from higher rated sellers, bidder participation is positively affected by listing richness, count of seller feedback ratings and number of questions. The significance of listing richness (count of photos of the item) is consistent with the case of listings from low rated sellers. We find that member feedback count has positive effect on bidders’ participation. An interesting observation is that the number of questions is significant instead of number of answers. From table 2, we observe that there is not much difference in mean number of answers and questions which means that high rated sellers mostly answer all the questions asked. Further, the number of questions is lower compared to lower rated sellers. We conjecture that high rated sellers provide much more detailed information in their listings and hence, bidders don’t need to ask for many clarifications regarding the item or the transaction as whole. Since these sellers are highly trusted, a higher number of questions in these listings signals higher interest about this auction and may attract more potential bidders to bid in this auction. We also find that the number of bids positively impacts the likelihood of an auction meeting its reserve. The variable BidCount, measures the total number of bids placed in an auction. The number of bids may increase due to bidding competition or due to an increase in the number of bidders, and in both cases will result in higher final bid.

#### 3.2.2 Study 2: Effect of Listing Richness

In study 1, we found that sellers’ feedback ratings moderate the effect of Q&A process on bidders’ participation. Our results in study 1 show that listing richness (number of photos of the item for sale) is a significant driver in determining bidders’ participation for new and experienced sellers alike. We find that a higher number of photos (PhotoCount) posted for the item attracts more bids. The photos posted in an auction listing are a key feature of the listing, and contribute significantly to authenticating the offered item. Furthermore, listing richness is critical in providing product related information to prospective bidders. One can argue that listings with lower richness may prompt bidders to ask more questions and require sellers to rely on Q&A process for information dissemination. Hence, Q&A process becomes critical in these listings to induce bidders’ participation. We examine whether listing richness moderates the effect of Q&A process on bidders’ participation.

Using BidCount as the dependent variable, we estimated the following linear regression model to examine the impact of the Q&A process on auction outcome.

We split our dataset on median listing richness - dataset with listing richness below (dataset C) and above (dataset D) median value. We estimate our model in equation 2 for both the datasets. Table 7 and 8 show the results for dataset C and tables 9 and 10 show the results for dataset D. A stepwise regression model of dataset C yielded an R2 value of 0.137 (F=22.68, p=0.000). The model characteristics and components are shown in Table 4.

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>11802.664</td>
<td>2</td>
<td>5901.332</td>
<td>22.675</td>
<td>.000</td>
</tr>
<tr>
<td>Residual</td>
<td>74434.097</td>
<td>286</td>
<td>260.259</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>86236.761</td>
<td>288</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7: ANOVA Results, Study 2(PhotoCount below Median)

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
</tr>
<tr>
<td>(Constant)</td>
<td>8.769</td>
<td>1.776</td>
</tr>
<tr>
<td>AnswerCount</td>
<td>5.100</td>
<td>0.798</td>
</tr>
<tr>
<td>SellerQuestionCount</td>
<td>0.982</td>
<td>0.454</td>
</tr>
</tbody>
</table>

Dependent Variable: BidCount

Table 8: Model Coefficients

The results in table 7 and table 8 show that the number of answers from sellers is a significant driver of bidders’ participation. As we conjectured, listings with lower richness require sellers to provide product related information by responding to questions asked by the bidders. It is interesting to note that the number of questions in these listings does not seem to be high and sellers respond to most of the questions. We also find that number of questions asked by the seller in other auction listings is also significant. Sellers of these listings, on average, have higher feedback ratings. Since higher rated sellers command more trust, questions posted by these sellers in other listings signals their familiarity and knowledge about the market of similar items. In auction market like the one we consider, sellers need to choose appropriate reserve prices based on the market of similar items. Hence, sellers who post more questions on other listings, signal their market knowledge and thus attract higher bidder participation on their listings.

To test the effect of Q&A process on bidders’ participation on auctions with higher listing richness, we estimate the model as shown in equation 2 using the dataset D. A stepwise regression model for dataset D yielded an R2 value of 0.169 (F=52.47, p=0.000). The model characteristics and components are shown in Tables 9 and 10 below.

The results from tables 9 and 10 show that number of answers, members feedback ratings and listing richness (number of photos) are significant drivers for bidders’ participation in auctions where listing richness is high. We have consistently shown that listing richness positively impact bidders’ participation. The significance of listing richness (photo count) shows that higher product-related information helps attract more bidders to an auction. More photos of the
product allow prospective buyers to better evaluate the product and increase their confidence in the transaction as a whole.

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>114619.779</td>
<td>5</td>
<td>38206.593</td>
<td>52.472</td>
<td>.000</td>
</tr>
<tr>
<td>Residual</td>
<td>565027.739</td>
<td>776</td>
<td>728.129</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>679647.518</td>
<td>779</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 9: ANOVA Results, Study 2 (PhotoCount above Median)*

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Constant)</td>
<td>6.529</td>
<td>1.735</td>
<td>3.763</td>
<td>.000</td>
</tr>
<tr>
<td>AnswerCount</td>
<td>3.904</td>
<td>0.548</td>
<td>0.247</td>
<td>7.119</td>
</tr>
<tr>
<td>MemberFBCount</td>
<td>0.002</td>
<td>0.000</td>
<td>0.191</td>
<td>5.770</td>
</tr>
<tr>
<td>PhotoCount</td>
<td>1.858</td>
<td>0.274</td>
<td>0.237</td>
<td>6.780</td>
</tr>
</tbody>
</table>

Dependent Variable: BidCount

*Table 10: Model Coefficients*

Overall, this research presents preliminary findings to show the effect of Q&A process on bidders’ participation. We have also tried to explore the role of sellers’ feedback rating and auction listing richness in bidders’ participation. We are in the process of collecting more data to examine the effect and interaction of sellers’ feedback ratings, richness of auction listings and intra-auction Q&A process on bidders’ participation, which eventually affects success of an auction.

4 Conclusion

Understanding the factors that influence the success and effectiveness of online auctions is an important contemporary problem in electronic commerce, and the literature on this problem has largely focused on authentication mechanisms such as listing richness, online reviews and ratings, and seller reputation. Our study complements this literature by investigating the effect of intra-auction exchange between buyers and sellers in the online auction environment. Intra-auction exchange (Q&A) allows potential buyers to ask any question pertaining to the transaction and helps them decide to commit for the transaction.

Our empirical analysis of online auctions demonstrates that indeed, the Q&A process has a significant impact on the auction outcome in terms of the number of bids, which increase the likelihood of the success of the auction. This is of potential value to online auction operators, since it provides them an additional lever to use for increase both the level of participation and outcomes of auctions. However, our study is only a first step in a relatively rich space. It motivates further analysis of a number of issues, which we outline here as areas for further research.

4.1 Limitations

First, our analysis so far has only considered the aggregate Q&A process. However, it would be useful to drill down into the different types of questions that are posed. As discussed earlier, the problem of online authentication concerns both the items offered for sale, as well as the sellers
themselves. For both these entities, there are a number of existing authentication mechanisms, including tools for rich listings, seller track records, online reviews, ratings and certifications. An interesting question is therefore whether the Q&A process is influential for authentication of either or both entities. For instance, one can ask if sellers’ feedback ratings affect Q&A and listing richness in any way. We observe from table 2 that auctions from high rated sellers tend to get fewer questions. Further, auctions from high rated sellers also have fewer photos (lower listing richness) than from lower rated sellers. These questions certainly require a follow up in-depth analysis to tease out the interactions between seller feedback rating, listing richness and Q&A process.

4.2 Future Research

Extending this line of research, a related question is why people pose questions during auctions. So far, we have implicitly assumed that people pose questions to get a better understanding of the offered item’s features and quality, and possibly to also better understand the quality of the seller. However, questions could also be used strategically, to enhance or even impede the auction process. Anecdotal evidence suggests that people sometimes pose questions that are frivolous, irreverent, and possibly even disruptive. Analysis of the extent of such activities, and who is involved, can help auction operators potentially design mechanisms to mitigate their effects.

Another area for further inquiry is into who poses questions. As suggested by our preliminary results, auction outcomes are influenced by the participation of bidders in the Q&A process. A more comprehensive analysis based on a complete bid history would help shed further light on the significance of bidders actively participating in the Q&A process, along with other passive participants, and even “lurkers”. A related question of interest would be about who answers questions. While it is reasonable to assume that in many cases, the seller is the primary entity providing answers that is not always the case. In fact, the participation of other people in the answering process may play an important role in the authentication process.

A third promising area for further study is the dynamics of the Q&A process. As mentioned earlier, our analysis in this paper focuses on the impact of the level of Q&A on auction outcome. However, the timing, duration and patterns of Q&A may also play significant roles. Analysis based on more detailed data on event timing and frequency may be very enlightening.

Yet another direction for investigation is on whether the type of item offered for auction impacts the role of the Q&A process. In our analysis, we have used data on auctions for a small set of product categories (automobiles, collectibles, books and clothing) . Drilling down into specific product categories may provide useful insights into whether Q&A is more useful for authenticating certain types of products versus others.

References

ANALYSING QUALITATIVE DATA FROM VIRTUAL WORLDS: USING IMAGES AND TEXT MINING

Brad McKenna Operations and Systems Management, University of Portsmouth, United Kingdom 
brad.mckenna@port.ac.uk

Michael Myers, Information Systems and Operations Management, University of Auckland, New Zealand m.myers@auckland.ac.nz

Lesley Gardner, Information Systems and Operations Management, University of Auckland, New Zealand l.gardner@auckland.ac.nz

Abstract

There is an increasing interest within both organisational and social contexts in virtual worlds and virtual reality platforms. Virtual worlds are highly graphical systems in which avatars interact with each other, and almost every event and conversation is logged and recorded. This presents new challenges for qualitative researchers in information systems. This paper addresses the challenges of analyzing the huge amounts of qualitative data that can be obtained from virtual worlds (both images and text). It addresses how images might be used in qualitative studies of virtual worlds, and proposes a new way to analyze textual data using a qualitative software tool called Leximancer. This paper illustrates these methods using a study of a social movement in a virtual world.

Keywords: Virtual worlds, qualitative, images, text, leximancer

1 INTRODUCTION

The recent purchase of Oculus Rift, a virtual reality platform, by Facebook for USD $2 Billion indicates the increasing interest in virtual worlds and virtual reality platforms. Companies such as Toyota, IBM, American Apparel, and Reuters (Wasko et al., 2011) are also investing in virtual worlds for business purposes (Messinger et al., 2009). We suggest this increasing interest in virtual worlds presents a new challenge for qualitative researchers in information systems: if we are going to study these virtual worlds, how can the vast amounts of textual and graphical data that characterize virtual worlds be analyzed? Virtual worlds are highly graphical systems in which avatars interact with each other (Castronova, 2007). Not only are images used extensively, but conversations, events, and almost anything that happens online is logged and recorded.

There is literally a flood of qualitative data pouring into the Internet every day, waiting to be interpreted and analysed by the many companies trying to figure out how it might be possible to mine this huge trove of “big data” for marketing purposes. Whereas many companies are trying to use this data for research purposes, we suggest there is also an opportunity for qualitative researchers to use this data for research purposes. The purpose of this paper is to discuss the challenges of analyzing the huge amounts of qualitative data that can be obtained from virtual worlds (both images and text). Although the collection of qualitative data from virtual worlds is relatively easy (since in many cases the users of virtual worlds are in effect automatically recording any of their own words and actions themselves), the analysis of this data is more problematic.

Based on our prior study of social movements in virtual worlds, found in McKenna et al. (2012), this paper presents two key issues which emerged during the study. The overarching research problem from the study was to understand how social movements are using virtual worlds. The first key issue we discuss in this paper is to propose a new way to analyze the vast amounts of text in virtual worlds. The second key issue, is how images might be analyzed in qualitative studies of virtual worlds. We illustrate
how images and text can be analyzed using a study of a social movement in a virtual world called World of Warcraft (WoW) a virtual world game, made by Blizzard Entertainment.

2 Virtual Worlds

A virtual world can be defined as “a synchronous, persistent network of people, represented by avatars, facilitated by networked computers” (Bell, 2008). Virtual worlds include social virtual worlds such as Second Life, and gaming virtual worlds, such as WoW. Virtual worlds differ from more traditional social networking technologies because they provide a richer, more immersive experience, and can be deeply engaging (Wasko et al., 2011).

Millions of people have invested their time and energy into using these virtual worlds, creating characters, meeting new people, and engaging in new forms of social interaction. Virtual worlds are of interest to IS researchers for both their business and social aspects (Messinger et al., 2009). Games also provide an interesting research stream. For example, WoW offers an alternative world where social functions, learning, and the development of social skills occur in a virtual environment (Davidson and Goldberg, 2009). Virtual worlds have the potential to become laboratories where experiments in social science can test new norms, values, and institutions (Bainbridge, 2010). Virtual worlds also have the potential to be proving grounds for real world social innovations, cultures, and social movements (Bainbridge, 2009) as well as substituting for social institutions in the real world (Williams, 2006).

Virtual worlds create unique challenges for researchers. One unique aspect of virtual worlds is the field site. The main difference between traditional ethnographic field sites and virtual world field sites is that a virtual world researcher does not physically visit a research site (Kozinets, 2010). Instead, the researcher has to log into the virtual world from wherever they happen to be. This means that online ethnographic research is more like deskwork than fieldwork (Rutter and Smith, 2005). Another unique aspect is the ability for researchers to use hidden observations, without members of the online community even knowing the researcher is there (Guimarães, 2005). Data analysis in virtual worlds also introduces some unique aspects, for example, it may not be possible to determine the true age, sex, or ethnicity of virtual world participants (Kozinets and Kedzior, 2009). People in virtual worlds have the ability to be anonymous or use pseudonyms (Christopher, 2009). Virtual worlds also provide users with the ability to express themselves in ways which extend beyond their body, for example, by modifying the environment around them, altering their identity, and creating multiple avatars (Boellstorff, 2008).

3 Images

A key difference between traditional and online ethnographic research is that online studies have new types of digital texts (Urquhart and Vaast 2012). These digital texts include images or photographs (Andrade and Arthanari 2009). Virtual worlds are highly graphical systems in which avatars interact with each other (Castronova 2007). Due to their highly graphical nature, images become an important source of data in virtual world studies. We argue that by analyzing images, a researcher can understand how avatars interact with each other and the virtual objects around them.

Screen captures can be very useful for showing the space and inhabitants of a virtual world, while video captures allow researchers to examine how avatars move through space and how their animations appear (Moore et al. 2009). However, images are not a common data source in information systems research. In marketing, Kozinets (2010) discusses taking screenshots of whatever is on the researcher’s computer screen, but he mostly focuses on screenshots with textual data, not images. Images should be captured at key moments of these interactions, for example when the researcher observes something which is theoretically interesting or related to the research question.

Andrade and Arthanari (2009) advocate the use of images as part of the research process. The primary object of inquiry for many visual studies is a set of materials or activities that a scholar finds visually interesting (Pauwels 2011). Such images could be used to allow readers to understand a particular phenomenon, for example what does a particular area look like, or how did people look a decade ago?
(Pauwels 2011). We believe that presenting images when discussing virtual worlds is absolutely necessary. Virtual worlds differ from the physical world, and the presentation of such images would increase the contextual understanding of the reader.

Another use of images is temporally ordered photographs to show how something has changed over time (Rieger 2011). A repeat photograph, or ‘rephotograph’ is a duplicate photograph used alongside a pre-existing photograph to emphasize certain aspects which may have changed between photographs. The spatial location of the second image is typically repeated, showing the reader the same scene again and inviting comparison (Klett 2011). As virtual worlds are constantly changing (McKenna et al. 2012), this becomes important as images can be presented over multiple timeframes to illustrate those changes.

Cartographic (i.e. maps) have been used in the social sciences (McKinnon 2011) to represent different territories, illustrate the shape and contour of the land, or to indicate landmarks and features of the area (Spencer 2011). Maps have also been used in information systems (see (Walsham and Sahay 1999)). As virtual worlds are also virtual landscapes, maps can be used to represent the different virtual locations which exist within that virtual world, much the same as maps are used to represent the geographical locations of the physical world.

4 TEXTUAL ANALYSIS AND LEXIMANCER

Another challenge involved with the study of virtual worlds if the analysis of textual data. With online studies of a virtual world, the amount of data can be huge. One tool which we have found useful for analyzing large amounts of text is Leximancer, which was developed at the University of Queensland, Australia. Leximancer uses machine learning (content analysis) to analyze large qualitative data sets and to display the results in a visual format. There are examples of research using Leximancer in accounting and management (Crofts and Bismar 2010), conceptual modelling (Davies et al. 2006), human-computer studies (Stockwell et al. 2009), risk management (Martin and Rice 2007), and event management (Scott and Smith 2005). Leximancer has been evaluated for stability and reproducibility and its results so far have been proven to be reliable (Palmer 2013; Rooney 2005; Smith and Humphreys 2006).

Leximancer creates visual output in the form of a conceptual map which presents the main themes contained within the text, and information about how those themes are related. The themes are heat-mapped to indicate their importance. Therefore, the ‘hottest’ (most important) themes appear in red, and the next most important theme in orange, and so on. Leximancer also allows the researcher to extract the actual pieces of text which were used to create the themes.

Content Analysis in Leximancer can be supervised or unsupervised. If using the supervised approach, the researcher will construct a set of key terms (known as concepts) usually with some background knowledge within the domain, or with some theoretical sensitivity. Alternatively, in the unsupervised approach, the algorithm will discover the concepts on its own via reading and re-reading the data. It is the unsupervised approach which is the greatest strength of Leximancer as it is very useful when there is no prior model or set of factors by which to analyze the data (Davies et al. 2006; Palmer 2013). This approach relies on the algorithms in Leximancer to detect the main themes and concepts arising from the data (Palmer 2013).

The remainder of this paper will introduce the virtual world that we studied, and explain how we analyzed our graphical and textual data using Leximancer.

5 RESEARCH APPROACH

In the original study we decided to focus our exploration on one of the largest social movements within WoW, a Lesbian, Gay, Bisexual, and Transgender movement, hereafter referred to as the LGBT movement. LGBT aims to create awareness for LGBT issues, both in game and out. By early 2013 LGBT has over 7,800 members (players) in WoW and has over 15,000 characters (it is possible for one player to add multiple characters). LGBT was established on a WoW server in October 2006 to "better
service the LGBT community and offer a safe, inclusive place to game for members of any sexual orientation or gender identity” (LGBT movement website, 2010).

LGBT has been profiled in a number of gay and lesbian magazines and in a prominent WoW blog website. LGBT is a global social movement with members from many countries. LGBT also maintains a website with discussion forums. LGBT holds many regular activities inside WoW such as an annual pride parade with floats, model competitions, dance parties, group photographs, and events for Valentine’s Day. These events are generally organized by the leaders of LGBT, often with input from members via the discussion forum. They also organize member meetings in the real world, and have had meetings in Australia, Canada, and the United States. In McKenna et al. (2012) we realized that these activities were influenced by patches, which are changes to the game released by the game designers. We followed how these changes to the technical ecosystem influenced the social activities of LGBT. The research question addressed in this study is: how does the technological artifact (the virtual world) and the social world (the social movement) co-evolve?

The lead researcher joined the LGBT movement in WoW and participated in a number of movement activities such as virtual pride parades, dance parties, and group photographs. During the fieldwork, field notes were taken. These notes included digital texts, suggested by Urquhart and Vaast (2012) to include images. Field notes based on participant observation online took a different form than in traditional ethnographic studies. The lead researcher did not actively take field notes during interactions with LGBT, but rather chose to record details with screen captures. On occasions, LGBT members would meet together and perform social movement activities, such as the pride parade. The lead researcher participated in these activities, for example, by marching in the parade. However, marching in the parade involves controlling an avatar’s movements with a keyboard and mouse, which means one’s hands are not free for active note taking.

In our study we found that the discussion forum data was also an extremely valuable data source. In some ways it can be argued that the discussion forum data is more “authentic” than interview data. As Myers and Newman point out, the interview is an artificial situation (Myers and Newman 2007). The interviewee is prompted by a stranger (the interviewer) to answer some questions at a specific point in time, whereas discussion forum data records what people actually thought and posted at the time and is continually updated.

The multiple sources of data collected throughout the research project are listed in Table 1. In total, the lead researcher spent over 1,600 hours engaging with LGBT. The next section of this paper will discuss more specific methodological issues from the original study.
6 DATA REDUCTION AND ANALYSIS

The discussion forum data proved problematic due to the large amount of data collected. It was necessary for the researchers to find a way to filter out the data which was not considered important for answering our research questions.

For the analysis of the dataset we first loaded the entire dataset into Leximancer. We soon realized that this was not a good approach as the dataset contained many posts which were irrelevant to the research questions proposed. Therefore we had to find a way to reduce the dataset. To reduce the dataset we read the patch notes of 114 patches in an attempt to discover which patch had an impact on LGBT. After analysis of the patches, we discovered 3 patches which had a strong influence. Second, we had to discover the impact that these patches had on LGBT. Often the patch notes are released before the patch is implemented into the game. Therefore we were able to filter the data from the discussion forum by extracting only those posts made about a certain patch before and after the patch was released. We also performed keyword searches using keywords from our theoretical approaches along with keywords based on our knowledge of the game and the impacts from analyzing the patch notes. Therefore we were able to disregard most of the posts, which gave us a final count of 405 posts. The chat logs also proved to be difficult to extract useful data. Having collected 1.5 years of chat logs, the data set was massive. Therefore we used keyword searches to extract useful text. In total we extracted 10 useful text excerpts from the chat logs.

6.1 Leximancer

Data analysis subsequently involved both the use of NVivo and Leximancer. NVivo is a well-known qualitative software package so a discussion of its use of beyond the scope of this paper. Because the nature of this study was exploratory, the unsupervised approach was used as we were interested to see what emerged from the data. Only the discussion forum data was analyzed in Leximancer due to the large number of posts downloaded.

The entire dataset was first loaded into Leximancer, and the unsupervised approach was executed. This created a set of Leximancer themes which were then filtered out through the researcher’s knowledge of the patches which were implemented in WoW. This reduced the dataset to a total of 405 posts which were useful for answering the research question. The reduced dataset was loaded into NVivo for coding manual coding with theoretical sensitivity to actor-network theory (ANT), which is an appropriate tool.

<table>
<thead>
<tr>
<th>Source of Data</th>
<th>Nature of Data Collected</th>
<th>Quantity Collected</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant Observation</td>
<td>Screen captures from movement activities.</td>
<td>At least 50 screen captures.</td>
<td>Images</td>
</tr>
<tr>
<td>Discussion Forum Posts</td>
<td>Discussion posts from movement website.</td>
<td>128,773 posts dating back to 2006.</td>
<td>Text</td>
</tr>
<tr>
<td>Chat Logs</td>
<td>Chat logs from movement in-game chat channels.</td>
<td>Approximately 1.5 years worth of chat logs.</td>
<td>Text</td>
</tr>
<tr>
<td>Social movement’s website</td>
<td>Textual information relating to background information about the movement and rules of membership.</td>
<td>Approximately 20 pages of information.</td>
<td>Text</td>
</tr>
<tr>
<td>WoW Patch Notes</td>
<td>Documents the changes implemented by a patch.</td>
<td>Patches dating back to 2006. 114 patches.</td>
<td>Text</td>
</tr>
<tr>
<td>Other WoW websites</td>
<td>Textual information relating to aspects of WoW gameplay.</td>
<td>Not documented but estimated to be over 100 pages of information</td>
<td>Text</td>
</tr>
</tbody>
</table>

Table 1. Data sources used in this study.
for socio-technical research (Callon 1986), and has been advocated for use in information systems research by Walsham (1997). ANT does not distinguish between human and non-human elements, including people, software, hardware, organizations, processes, and treats the social and the technical as inseparable (Walsham 1997). ANT is a useful means for data analysis in this study for its ability to examine the co-evolution of society and technology (Callon 1986). A more detailed discussion of theory choice is beyond the scope of this paper, but can be found in McKenna et al. (2012). The reduced dataset was also loaded back into Leximancer, and the unsupervised approach was rerun. This created a new set of themes (Figure 1), and knowledge pathways. The text which created those knowledge pathways was then loaded into NVivo and further coded manually with theoretical sensitivity to ANT. Next the codes from both rounds of NVivo coding were compared with each other. This process is illustrated in Figure 1.

![Coding Process](image)

**Figure 1. Coding Process**

An example of the Leximancer output is illustrated in Figure 2. In the example illustrated, the sub-set of data set was used to extract the common themes evident in the discussion forums when a new patch was released which placed a cap on the size of guilds. On the left hand side is the conceptual map. Each circle within the map represents a theme. We can see from this image that the most important theme was “guild”, followed by “members”. Each theme contains multiple concepts (nodes) which make up that theme. The solid line indicates the knowledge pathway, which shows the connections between concepts. These pathways were used to empirically link concepts together which provided a very useful way of understanding the data, and analyzing relationships between concepts. The right hand side indicates the actual text extracted from Leximancer which creates the knowledge pathway, i.e. the text which supports the relationship between the concepts guild and Blizzard. Note that only one knowledge pathway is shown, but can be created between any concepts in the map.
6.2 Images

The other main data source was the screenshots (images). We coded the screenshots using NVivo. In the following section, we will briefly review some key points from Andrade and Arthanari (2009) who present three stages for using images in research: image maker, image analyzer, and image presenter.

The first stage of using images in a research report is the information systems researcher playing the role of an image maker. Andrade and Arthanari (2009) discuss this role in relation to that of a photographer. However, in our case no camera was needed. The image was taken by the researcher typing a keyboard command on his computer which took a screenshot of the researcher’s computer screen (WoW has a built in command for taking screenshots). None of the other players would be aware that the screenshot is taken. These screenshots are saved automatically in a WoW folder within the computer where the game is installed.

The second stage of using images requires the researcher to become an image analyzer. In order to interpret an image, the researcher must understand the subject matter and understand the context of the image (Andrade and Arthanari 2009). In our study images were used to aid the analysis of social movements. For example, the image presented in Figure 3 illustrates players getting their avatars ready to begin the virtual parade. Avatars group together into pre-organized floats, and then begin to form into lines in order to march through the virtual landscape.

Another example of images taken which fits within the context of social movements is illustrated in Figure 4. This image illustrates the actual parade. Here we can see several items of interest to our study. Firstly we can see the parade itself, in which both friendly and enemy players are participating. In WoW, Horde and Alliance players are meant to be fighting each other, but during the parade they come together and fighting is discouraged. In fact the lead researcher had installed an add-on to the game which notifies him when an enemy player is nearby. This add-on is displayed at the bottom right of the image, and is creating many warnings due to many of the parade participants being enemy players.
Players form into groups (floats) before the parade begins.

Figure 3. Avatars getting ready to start the parade. Image: ©2004 Blizzard Entertainment, Inc. All rights reserved.

Some parade participants decide to fly above the parade rather than march.

Non-members of LGBT observing or participating in the parade.

Both friendly and enemy players marching together in the parade.

Add-on warning that enemy players are nearby.

Horde and Alliance marching together.

Figure 4. Virtual Pride Parade. Image: ©2004 Blizzard Entertainment, Inc. All rights reserved.

The image also shows non-members of LGBT who come out to watch the parade. In this image they are standing to the side of the parade, however it is not uncommon for non-members to also participate in
the parade. The image also illustrates that not all parade participants choose to walk with the parade, as in the top left of the image some participants choose to fly above it.

The final stage of using images is where the author becomes an image presenter, where researchers produce a written report of their findings. Alongside the written text, images can be presented to provide an understanding of a phenomenon which might be difficult to present in textual form.

7 DISCUSSION

According to van Leeuwen (2011) “text and image depend on each other to convey the whole of the content” (p. 551). The following example illustrates how multiple data sources were used to triangulate our analysis, in order to tell a convincing story.

7.1 Pride Parades

The pride parades were directly impacted by a patch. This example illustrates the use of patch notes, images, chat, and discussion forum data. This patch was implemented in December 2010 and changed the virtual landscape of WoW. Some virtual areas within the game become flooded, while in others giant canyons were formed. The patch notes are illustrated in Table 2. If a reader was unfamiliar with WoW, this would mean absolutely nothing to them. However, if an image was to accompany the text, it would add more value to the full understanding of the context.

Table 2. Patch 4.0.3a patch notes

| Azeroth Shattered: Deathwing’s return has had an immeasurable impact throughout the Eastern Kingdoms and Kalimdor. Players will notice drastically altered terrain. |

Therefore in this example we have used an image to illustrate the changes to the virtual landscape (see Figure 5), or rephotograph as advocated by Klett (2011). In this example we can see the virtual landscape before and after patch implementation. The images at the top are before patch, the images at the bottom are after patch. The image at the top-left shows a peaceful village called Camp Taurajo in the region known as The Barrens which was used as the starting point of the pride parade. The image at the top-right shows a flat plain land where the pride parade used to march through. On the bottom-left, the image shows how Camp Taurajo was destroyed after the patch, and on the bottom-right how the plain land has now become a deep canyon which has split The Barrens into North Barrens and South Barrens. In Figure 6, we illustrate how we used a map (McKinnon 2011; Spencer 2011) to aid our analysis of how the parade route had to change after the patch was implemented.

We also captured some chat from LGBT members around this event. Note that names have been disguised.

[Pradush]: aww poor camp taurajo
[...]
[Hasan]: I'm loving how southern barrens turned out
[Adwardomos]: How does one get from northern barrens to southern barrens?
[Sramak]: Go south?
[Zeb]: I had to go all the way around it
[Adwardomos]: To the west or east?
This chat was especially important because it illustrates that players were having difficulty crossing the canyon. This was the same route that LGBT had previously taken for their pride parade. Therefore LGBT had to move the parade to a new virtual location for the parade.

![Figure 5. Parade starting location before patch (top-left), parade route before patch (top-right), parade starting position after patch (bottom-left), parade route after patch (bottom-right). Image: ©2004 Blizzard Entertainment, Inc. All rights reserved.](image)

![Figure 6. Old parade route (triangles), new parade route (circles). Image: ©2004 Blizzard Entertainment, Inc. All rights reserved.](image)

Together these multiple data sources are able to tell the story of how changes in configurations to a virtual world affect the inhabitants of that virtual world. The next example illustrates a further impact, this time on the size of guilds.
7.2 Guild Size Caps

During beta testing of the expansion pack called Cataclysm, Blizzard became aware that the size of the guilds must be capped at 600 members (later increased to 1,000). Blizzard was implementing new systems to manage guilds, which contained complex systems which track the (gaming) contributions that an individual guild member makes to the guild. Larger guilds would therefore place more strains on these systems. Therefore in order to allow the system to run smoothly, they found that capping the size of guilds was necessary. At the time this patch was implemented, LGBT had approximately 6,200 members. In this example mostly the discussion forum data was used, and NVivo and Leximancer were used for the analysis.

On the left side of Table 3 the Leximancer conceptual map is illustrated which indicates the discovered themes and concepts from the data relating to the guild size caps. This example also illustrates the knowledge pathway (i.e. actual text) which represents the linkage between guild and size. This pathway connects these two end concepts with the intermediate concepts of members and chat. On the right hand side the text which creates the knowledge pathway is presented.

<table>
<thead>
<tr>
<th>Leximancer Conceptual Map</th>
<th>Knowledge Pathway</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Leximancer Conceptual Map" /></td>
<td>1. Hello lovely members of LGBT! As you all know by now, Blizzard is instituting a 600-person hard cap on guild size starting in patch 4.0.1, which we expect to be released next week [discussion forum].</td>
</tr>
<tr>
<td></td>
<td>2. Previously, adding players to a guild was as simple as adding people to a chat channel. While approximately 500 members were visible in the UI, there was no real need to limit guild size [discussion forum].</td>
</tr>
<tr>
<td></td>
<td>3. While some players have used options including mods and custom chat channels to support large player and guild alliances that number multiple thousands, groups of that size aren't ideally suited to our design philosophy. As always, we're continually looking into adding new features to help facilitate guild management, scheduling, and player communication [Blizzard official announcement].</td>
</tr>
</tbody>
</table>

Table 3. Leximancer Output

We also created manual codes in NVivo based on the discovered Leximancer concepts. The Leximancer output was loaded into NVivo and further coded. This gave us our second round of coding. Table 4 illustrates some of the Leximancer discovered concepts, and the manual coding performed in NVivo from the same text excerpt.
Supporting Text | Leximancer Concepts Discovered | NVivo Codes
--- | --- | ---
With the advent of the new guild system in Cataclysm we are tracking many more things on each individual player in a guild and in order to support that, we need to limit the amount of members to a reasonable level. The new cap of 600 members is fully supported in the new guild system and that means that everyone will be visible in the UI and able to contribute to all guild functions like experience and reputation gain. | System Support Members Experience Level | Blizzard’s reasoning Potential Impact on LGBT Limitations over code

When patch 4.0.1 goes live (most likely next week), we will reactivate the LGBT guild for all future invites, since we won’t be able to add any toons to LGBT. We will not be making any other changes or adding any other sub-guilds until we see exactly how the new system works and the results of our advocacy efforts with Blizzard. | Patch Future System Toons (Avatar/Character) Guild | Potential Impact on LGBT Game systems Potential solutions Discussions with Blizzard New sub-guilds

Table 4. Leximancer concepts and NVivo codes

8 Challenges

Our study has presented a number of unique challenges not evident in traditional ethnographic studies. First, huge amounts of data potentially creates a problem of qualitative data analysis. With such a large data set, it becomes necessary for the researcher to weed out data which is not relevant to the study, and to hunt through the data set to find the correct data points relevant to the research question. In some instances this is like finding a needle in a haystack. Our advice to researchers is to have some method to reduce the dataset. For example, we used patches, but other methods could be used such as filtering for research question specific concepts, or theoretical sensitivity.

The second challenge is how to analyze such a large data set. One obvious solution is to use automatic data analysis software. In this study, we realized that coding 128,773 discussion forum posts would be time consuming if using a qualitative data analysis software package like NVivo. Therefore, we decided to use Leximancer, which provides automated qualitative data analysis. Because Leximancer is an automatic text analysis tool, we suggest that researchers also compare the Leximancer outputs with codes created in NVivo. Not only does this give multiple rounds of coding (automatic and manual), but it also ensures that every possible theme of concept is extracted from the dataset.

However researchers need to be aware of the ‘garbage in garbage out’ problem, which is especially problematic when using live data such as discussion forums. Discussion forum data contains many threads and posts about an unlimited number of topics, and conversations which may start about one particular topic may suddenly go off on a tangent. Therefore we believe that even in the Leximancer approach some filtering is required by the researcher. We found that without this prior filtering, the Leximancer analysis was producing many inappropriate results. It was also necessary to further filter the data after the Leximancer analysis. For example the data from the discussion forums often had HTML tags embedded within it. As these tags are text, they were included in the results. Leximancer contains a pre-defined set of stop words, which are common English language words which will be
skipped over by the algorithm. It is possible to edit the stop word list so the HTML tags were added to the list and the algorithm was run again. This iterative process was completed many times to remove words unnecessary for the analysis. Other words were also added to the stop word list based on the prior knowledge of the researcher. It is claimed that the strength of Leximancer is its unbiased analysis of the data, however this iterative process by definition adds some amount of subjectivity and bias to the results based on the words the researcher decides to exclude from the analysis. Therefore, researchers must still be very familiar with the dataset prior to processing in Leximancer. Researchers should also have background contextual knowledge to ensure that words or not included or excluded from the analysis.

The third challenge is how to take field notes. We decided to capture what was happening in the field by taking screenshots. This requires data analysis similar to that presented in Andrade and Arthanari (2009). Because virtual worlds are highly graphical, with avatars interacting with other avatars and virtual objects, we found that that taking screenshots of these interactions with useful. But the researcher must know the appropriate time to take a screenshot. Therefore we recommend that the researcher spend time getting to know the virtual environment and the people and avatars within it. In our study the lead researcher spent over 1,600 hours interacting with members of LGBT. The researcher built up an intimate knowledge of their activities, which enabled the selection of images to ensure they were related to the research question.

The final challenge is combining all of the many different digital texts and analysis (van Leeuwen 2011) into the written report to tell a convincing story. We have presented examples of how this can be achieved. It is our hope that our experience in trying to use these alternative data sources might prove valuable for other IS researchers.

9 CONCLUSION

This paper has presented how we analyzed our qualitative data based on a study of a virtual world social. We have discussed how we analyzed images and the huge amount of textual data. As the quantity of textual data in online forms is increasing, we need to find new ways to manage and analyze this data. We have presented our approach to this problem by using a machine learning software tool (called Leximancer) to automatically discover key themes in vast quantities of text.

This study has its limitations. One limitation is that we used only a small subset of the dataset (405 posts) for analysis. Future studies could report on findings from Leximancer with much larger datasets. Because our data was downloaded directly from a discussion forum, it also contained much noise. Other textual based data sources, such as interviews or questionnaires, may contain less noise and would be easier to analyze without the need for a filtering process such as the one we used. Future studies could also look into using object recognition tools to analyze the content found within images.

References


A STUDY ON USERS’ ATTITUDE TO PERVERSIVE GAMES IN CHINA

Shang Gao
Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway

Zhe Zang
School of Business Administration, Zhongnan University of Economics and Law, Wuhan, China

Hong Guo
Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway

Abstract

Today, pervasive game has been becoming more and more popular in peoples’ daily life. It has been attracting increasingly attention from both researchers and practitioners. However, there is not much research about users’ attitude to pervasive games in China. In this research, we investigate users’ attitude to pervasive games in China with 373 valid questionnaires from the biggest city in central China. The results indicate that 34.85% of the participants would like to play video games in their spare time. A majority of the participants (60.32%) choose mobile/portable devices to play games, while the others play games on PC or game consoles. Further, most participants are likely to play games on mobile devices while they are waiting for something. In addition, most participants do not want to disclose their geographical location publicly when they are playing games. Last but not least, majority of the participants would like to play games with their acquaintances, rather than the strangers.

Keywords: Pervasive Games, Survey, Attitude

1. INTRODUCTION

Pervasive game is an emerging game genre which brings more physical movements and social interactions into game (Guo et al., 2010). There is no doubt that the use of pervasive games is growing. There are more and more pervasive games available on various devices. While there has been an increasing availability of pervasive games, limited attention has been given to users’ attitude to pervasive games.

The objective of this study is to investigate users’ attitude to pervasive games in China. In order to address this objective, we defined two research questions as outlined in Section 3. Consequently, a survey based on these two research questions was carried with 373 respondents in the biggest city in central China. The answers to these questions could further improve our understanding of the status of pervasive games in China by taking users’ attitude to this concept into account.

The remainder of the paper is structured as follows: Section 2 reviews relevant literature. We present the research methods in Section 3. Section 4 presents the survey results. In Section 5, we conclude this research and point out some future research directions.
2. LITERATURE REVIEW

The term pervasive game has been defined differently in literature. Montola et al. (2009) defined Pervasive Game as having "one or more salient features that expand the contractual magic circle of play spatially, temporally, or socially. A pervasive game is one where the gaming experience is extended out in the real world (Benford et al., 2005) or where the fictive world in which the game takes place blends with the physical world (Nieuwdorp, 2007a). Further, Nieuwdorp (2007b) tried to summarize different definitions of the term and ended up with a conclusion that “we need to let go of the notion of coining the ultimate definition of pervasive games, and instead ask ourselves what makes a game pervasive.” His argument brings a research question to the field that what the essential characteristics of pervasive games are.

There are some empirical studies on physical games and pervasive games. We list some existing literature as follows. For instance, Sweetser et al. (2005) proposed the GameFlow model where eight overall goals were defined: concentration, challenge, skill, control, clear goals, feedback, immersion, and social engagement. Jegers (2007) extends criteria of elements in general GameFlow Model according to attributes of pervasive gaming. This kind of work in literature does help improve understanding of pervasiveness.

Guo et al. (2010) investigated the essential characteristics of pervasive games by proposing TeMPS. TeMPS is a conceptual framework to characterize pervasive games through the four dimensions: Temporality, Mobility, Perceptibility, and Sociality. Moreover, based on this framework, Wang et al. (2010) studied consumers’ attitude towards pervasive games in Norway.

The four characteristics of TeMPS are summarized as follows:

- Temporality: addressing the temporal property about the game, i.e. the game is played in fixed time/round or not (open beginning and/or open ended);
- Mobility: addressing the spatial property whether the game could be played anywhere or fixed in one place;
- Perceptibility: addressing how the game is mixed with the reality, e.g. does the game construct the appearance of the player proxy in game by sensing the player’s real world appearance?
- Sociality: addressing the player’s relationship and social influence of the game.

Faria and Wellington (2005) described a large survey on business simulation games where they focused on how games were first adopted, objectives for game use, and usage of such games. Tychsen et al. (2008) presented a survey on the motivations for playing single- and multi-player role-playing games. In this survey, they found that the motivations for playing pervasive games were not simple constructs, but rather multiple motivational drivers that are interrelated.

As to empirical studies on pervasive games, HeartBeat, was designed to demonstrate the design vision of Head-Up games that put outdoors play center stage combining the benefits of traditional outdoor games with the opportunities for richer experiences and innovation offered by new media. The evaluation involving 32 children playing outdoors showed how the game promotes physical activity and social interaction between children in ways one would expect from traditional outdoor games (Magielse and Markopoulos, 2009).
Furthermore, TAM (Davis, 1989) is widely acknowledged as one of the most robust and influential models for explaining user acceptance of technology. TAM has been revised to incorporate additional variables in specific contexts, such as WWW (Johnson and Hignite, 2000), and mobile information services (Gao et al., 2014). Gao et al. (2015) extended TAM with additional factors (e.g., social factors and flow experience) to study the adoption of mobile games in China. The research findings demonstrated that flow experience played an important role in the adoption of mobile games.

To our knowledge, we did not find any studies aiming at studying users’ attitude to pervasive games in China. Therefore, we wanted to evaluate users’ attitude towards pervasive games in general in China in terms of four essential characteristics of pervasive games in TeMPS.

3. **RESEARCH DESIGN**

To study users’ attitude to pervasive games in China, we defined two research questions in this section. These research questions were examined in terms of a survey, which was carried out with college students in the biggest city in Central China. The survey instrument was developed based on extensive review on the existing literature and discussions with some experts in the field of pervasive games. In this survey, special emphasis was laid on the usage and users’ perspectives of pervasive games in China.

3.1 **Research questions**

We developed the following research questions (RQ) in this research.

RQ1. What is the general usage of pervasive games in China?

RQ2. What are users’ expectations on major characteristics of pervasive games in China?

3.2 **Scale design**

The survey consists three parts. The first part was intended to study users’ personal background (e.g., age, gender, educational background, etc). The second part was designed to measure the current usage of pervasive games in general in China. To better understand what types of pervasive games Chinese users would be interested in, we designed some items based on the major characteristics of pervasive games and asked the participants for their opinions on these in the third part of the survey. The development of the instrument for the third part of the survey is described as follows.

The third part of the survey was designed based on the TeMPS framework (Guo et al., 2010). The framework classifies pervasive games according to four main characteristics: temporality, mobility, perceptibility and sociality. We also made necessary adjustments to fit the scenario of this study. Further, we developed some items through some discussions with some senior researchers in information systems and some postgraduate students majoring in electronic business in China.

As a result, a total of 32 questions were included in this survey. A Five-point Likert scale, with 1 being the negative end of the scale (strongly disagree) and 5 being the positive end of the scale (strongly agree), was used to examine participants’ responses to some of the items in the survey.

3.3 **Sample**
In sampling, college students in the biggest city in central China were chosen to finish the survey. The survey was distributed in terms of online questionnaires from October 15, 2013 to December 15, 2013. As a result, 394 completed questionnaires were collected, among which 373 were valid questionnaires (i.e., valid respondent rate 94.7%).

4. DATA ANALYSIS AND DISCUSSION

4.1 Descriptive analysis

A total of 373 valid questionnaires was collected in this survey. As the test persons are all college students, the age structure was relatively young, with 97.86% of the participants between age 18 and 25. The sample consisting of 41.02% male and 58.98% female participants can reflect a relatively balanced ratio. Fig. 1 shows the major shares of the participants. According to these results, most respondents majored in financial management.

![Figure 1. Major shares of the participants](image1)

Besides, every participant had a mobile device and 76.41% of them owned a smartphone. The shares of operating systems used by the test persons are depicted in Fig. 2. According to these results, Android was the most used operating systems. 23.59% of the participants were non-smartphones users. And the second and third most used Operation Systems was Symbian and Apple IOS system.

![Figure 2. Operating systems of mobile end devices](image2)
4.2 Results on RQ1

Fig. 3 presented the main spare time activities of the participants. We found that 34.85% of the participants would like to play video games in their spare time.

![Figure 3. Shares of the main spare time activities](image)

47.72% of the participants spent less than 1 hour on playing video games. For more details, see Fig. 4.

![Figure 4. The time spent on game activities of the participants](image)

A majority of the participants (60.32%) chose mobile/portable devices to play games, while the others played games on PC or game consoles. Fig. 5 showed the frequently used mobile device types. When playing games on mobile/portable devices, most participants wanted to use smartphones, and another 28.69% wanted to use tablet PC. The other kinds of mobile devices shown by Fig. 5 were seldom used.

![Figure 5. Frequently used mobile device types](image)
According to Fig. 6, the results indicated most participants wanted to play games on mobile devices (e.g., smartphones, iPad, iTouch and similar) while they were waiting for something. At the same time, they were likely to play games at home or in class as well.

![Figure 6. The situations chosen to play games on mobile devices](image)

Most participants agreed that playing games in the spare time was a way to relax and relieve stress, rather than a waste of time (69.97%). Another interesting finding was that most participants had the experience of playing games on the mobile devices while in parallel with another activity (e.g., watching TV, waiting for the bus, cooking, working and similar). Fig. 7 showed the reasons for gaming relevant to the participants. The results indicated that time consuming, entertainment and relaxing were the main reasons for gaming.

![Figure 7. The reasons for gaming relevant to the participants](image)

A total of 78.02% of the participants rarely played games with others by using mobile devices. Besides, 61.12% of the participants would be more likely to stay at the same place when playing games on the mobile devices. Most participants did not like to switch their devices when playing games.
4.3 Results on RQ2

**Temporality**

In order to get the information of temporality in games, four questions were tested in this part. Fig. 8 showed the results of the four questions. In general, respondents did not like games with a fixed amount of time for a session which should have a definitive start and end, and quitting during the session would mean the users do not gain points for the session. Meanwhile, we found that most participants preferred games which were possible to start and end at any time he/she wants, and he/she should be able to play for any amount of time before quitting. The number of participants who prefer games without timing function were a little more than the ones who prefer games with timing function.

![Figure 8. The results of the questions about the temporality in games](image)

**Mobility**

Three questions were tested in this part to examine the mobility characteristic in pervasive games. The results were depicted in Fig. 9. For most participants, the value of a game would increase if the game contains the information about the real world. But most of them did not want their geographical location displayed publicly in the game process.

![Figure 9. The results of the questions about the mobility in games](image)
**Perceptibility**

In this part, perceptibility in games was examined in terms of three questions. Fig. 10 presented the results in detail. Some participants agreed that the value of a game would increase if he/she could control the game by moving real world objects (for example by opening a door in real world would open a door in game). The man-machine interactive function of the game was preferred by some participants. Thus, the value of a game would increase if the game could record the users’ feelings and act based on them (for example that the game measures heart rhythm and knows when the player is scared, relax and so on). Furthermore, some participants preferred to play games that could be controlled with voice commands.

![Figure 10. The results of the questions about the perceptibility in games](image)

**Social**

The social characteristic of pervasive games was examined in terms of four questions. The results were depicted in Fig. 11. Some participants indicated that the value of a game would increase if the user can compete with others to reach the goals. Meanwhile, cooperating with others to reach the game goals was also accepted by most participants. When playing the game, more users wanted to be able to communicate with others. However, the majority of the participants would like to play games with their acquaintances, rather than the strangers.

![Figure 11. The results of the questions about the social element in games](image)
5. CONCLUSION AND FUTURE RESEARCH

This research investigated users’ attitude to pervasive games in China with the responded valid 373 questionnaires from the biggest city in central China. The results of survey provided some insights into many new findings, especially with respect to users’ perspectives on four essential characteristics of pervasive games in TeMPS framework. The results indicated that 34.85% of the participants would like to play video games in their spare time. A majority of the participants (60.32%) chose mobile/portable devices to play games, while the others play games on PC or game consoles.

Other interesting findings are as follows: 1). Most participants were likely to play games on mobile devices while they were waiting for something. 2). Users would like to see a session length that fit their playing style. There should not be a limit on how long a player has to play in order to gain something from the session. 3). Most participants did not want their geographical location displayed publicly in the game process. 4). some participants preferred to play games that could be controlled with voice commands. 5). the majority of the participants would like to play games with their acquaintances, rather than the strangers. We believe that the findings of this research can help developers create a better game experience with the use of pervasive elements and mobile technology in games.

We were also aware of some limitations in this research. Firstly, the sample size of the study was relatively small. This might reduce the power of our findings. Moreover, the sample used for analysis was drawn from a city in the central part of China, so the generalizability of the results remains to be tested. Last but not least, all the data were collected using self-reported scales in terms of online questionnaires. The results might exist the bias for online questionnaire (Dollman et al., 2004).

As for future research, we plan to make a deeper exploration on the usage and acceptance of pervasive games in China by attempting a nationwide sample. Another possible direction is to update the present study with a longitudinal design.

REFERENCES


THE INTERPLAY OF OMNIRETAILING & STORE ATMOSPHERE ON CONSUMER’S PURCHASE INTENTION TOWARDS THE PHYSICAL RETAIL STORE

Chris Lazaris, ELTRUN - The E-Business Research Center, Athens University of Economics & Business, School of Business, Department of Management Science & Technology, lazaris@aueb.gr
Adam Vrechopoulos, ELTRUN - The E-Business Research Center, Athens University of Economics & Business, School of Business, Department of Management Science & Technology, avrehop@aueb.gr
Georgios Doukidis, ELTRUN - The E-Business Research Center, Athens University of Economics & Business, School of Business, Department of Management Science & Technology, gjd@aueb.gr
Katerina Fraidaki, ELTRUN - The E-Business Research Center, Athens University of Economics & Business, School of Business, Department of Management Science & Technology, fraudaki@aeub.gr

Abstract

The increasing use of consumer supportive technologies and applications within the physical retail store in the context of the omnichannel retailing era has enhanced shopping experience & store atmosphere. In parallel, emerging consumer behavioural patterns indicate that physical retail store visitors interact with the offline and the online channel simultaneously, mainly through their smart mobile devices. Relevant research, however, has not thoroughly investigated that issue from a consumer perspective. The present study investigates the importance consumers attach to a series of online practices offered in the physical retail stores and explores consumers' preferences regarding the combined use of online & offline store atmosphere and omnichannel criteria that affect their purchasing intentions towards the physical store. The findings show that both in-store Internet users and non-users attach more or less the same importance to some specific online practices offered within the physical store. On the other hand, consumers with different levels of in-store internet use evaluate online & offline store atmosphere and omnichannel criteria in a differentiated manner. The study ends by providing some implications to practitioners and researchers regarding the omniretailing technologies that should be applied in the physical stores, as well as the importance of online & offline store atmosphere in order to purchase from them.

Keywords: Store Atmosphere, Omnichannel Retailing, Consumer Behaviour, Retail Technologies.

1 INTRODUCTION

Since the introduction of the online channel, mainly in the form of the world wide web, retailers have attempted to establish it as a retail medium, complementing or replacing existing offline ones. In order to make ICT retailing technologies more consumer friendly, retailers tried to adapt offline practices to online ones. Online store features, such as the electronic shopping cart, electronic product catalogs, advertising banners, e-mail campaigns & store layout tried to resemble their offline counterparts, in order to assist consumers' buying behaviour and enhance their shopping experience.

Part of that approach was to replicate a store’s conventional atmosphere to an online one. Store atmospherics consist of “all of the physical and non-physical elements of a store that can be controlled in order to enhance (or restrain) the behaviors of its occupants, both customers and employees”
(Eroglu & Machleit, 1993, p.34). In accordance, a web atmospheric cue is “comparable to a brick-and-mortar atmospheric cue and can be defined as any web interface component within an individual’s perceptual field that stimulates one’s senses” (Dailey, 2004, p.796).

As the internet started to establish itself as a part of everyday life, e-commerce technologies and practices began to mature and consumers started to familiarize with them at high levels. Both retailers and consumers became multichannel, in the sense that they choose either channel in order to satisfy their purchase intentions. The introduction of the mobile and social media as new online channels increased consumer engagement towards online, as well as internet use. Consumers were now able to interact with all channels at all times, exploiting unique advantages that emerged. The boundaries of online and offline channels became blurred and a disruption in retailing industry took place: the emergence of omnichannel retailing, as an evolution of multichannel retailing.

Online best practices and technologies were brought to the physical store, as retailers’ strategies turned to the opposite direction: to replicate the online environment to the offline one. Nevertheless, the majority of retailers haven’t established a clear omniretailing strategy regarding the optimum application of related ICT technologies or how to seamlessly integrate them (e.g. consider the potential role of the physical store’s salespeople that use innovative ICT technologies to support the consumer buying processes). In fact, most of them try to empower consumers with in-store self-service technologies in order to create an enhanced store atmosphere and shopping experience (Wurmser, 2014).

The purpose of this study is to explore the consumer’s perception about this new hybrid retailing environment, in relation to online & conventional atmospherics, that are integrated through omnichannel retailing practices and technologies. To this end, the study attempts to discover what omniretailing technologies & practices add value to the in-store shopping experience, from the consumer’s point of view.

2 STORE ATMOSPHERE & OMNICHANNEL RETAILING THEORETICAL FOUNDATIONS

It was not until Kotler (1973) introduced the store atmosphere notion based on the four human senses. Store atmosphere initially consisted of four atmospheric dimensions: visual, aural, olfactory and tactile.

Later on, these dimensions expanded with the addition of human & social factors. Baker (1986) added the social dimension (store employees and customers) and Turley & Milliman (2000) as well, included the human variable (employee & customer characteristics). At that point, store atmosphere components also included external, general interior, layout & design, point of purchase & decoration variables (Turley & Milliman, 2000).

In parallel, with the introduction of the internet medium, store atmosphere was translated to the virtual world by various researchers. Dailey (1999) and Vrechopoulos et al. (2000) were among the first ones, adapting conventional atmospherics principles. According to Vrechopoulos et al. (2000), virtual store atmosphere consists of virtual layout & design (tree/hub, pipeline & guiding pathway structure), virtual atmospherics (site view, sound, scent) and virtual theatrics (animation techniques).

Afterward, Manganari et al. (2007) presented mobile commerce (m-atmospherics) and also added the social factor (virtual social presence), in accordance to conventional atmospherics (Manganari et al., 2009). Newly, Park et al. (2014) following the paradigm of social networks, introduced the social networking atmosphere. In this way, online store atmosphere included all main online channels, but there was no framework that brought them altogether.

That was until Poncin & Mimoun (2014) united Baker’s and Turley & Milliman’s framework in a single store atmosphere study. Findings from that study included that in-store technologies & multichannel integration affect consumer behaviour in the physical store.

However, that study did not take into consideration the omnichannel phenomenon. Originating from the latin word omnis (meaning: all, everything) it made its first appearance in the practitioners’ domain. It was first introduced in order to separate itself from the multichannel concept through the
simultaneous and not just parallel use of channels by consumers (Parker & Hand, 2009 and Ortis & Casoli, 2009).

In academia, Rigby (2011, p.4) was the first to define omnichannel retailing as “an integrated sales experience that melds the advantages of physical stores with the information-rich experience of online shopping”. Recently, Omniretailing, synonymous to this concept, appears by Levy, et al. (2013, p.67) as “a coordinated multichannel offering that provides a seamless experience when using all of the retailer’s shopping channels”.

To that end, although omniretailing extends from the physical retail store to the whole supply chain, the simultaneous presence of all channels take place within the offline store, altering the shopping environment. We could postulate that this hybrid store atmosphere, consisting of both online & offline channels atmospheric cues is the result of the omniretailing interplay. In a sense, it could be described as the Omnichannel Retailing Store Atmosphere (ORSA).

3 **RELEVANT EMPIRICAL STUDIES & RESEARCH HYPOTHESES**

Omniretailing practices in the physical store take place when consumers use all available retailer’s channels (Levy, et al. (2013, p.67). Therefore, consumers should utilize the online channel within the physical store, in order to engage in omnichannel while present in the conventional store. The prevalent way to achieve this, is by using their mobile phones’ internet capabilities. In fact, the degree that shoppers use omnichannel retailing practices, encountered as omnichannel retailing intensity by Lazaris et al. (2014), affects the frequency of their mobile Internet use (Lazaris et al. 2014).

On the other hand, the frequency of online shopping imposes several implications to shopping orientation across several criteria (Schramm-Klein et al., 2007). Schramm-Klein et al. (2007) have compared offline vs. online stores' frequency of use among internet users, in terms of the orientation towards store atmosphere, service, personal communication and contact. They found out that there is no statistical significant relationship between frequent online & offline shoppers in terms of shopping orientation towards store atmosphere (both online & traditional) & service. In contrast, there was found to be significant statistical relationship between frequent online & offline shoppers in terms of shopping orientation towards personal communication and contact: frequent offline shoppers were more oriented towards these criteria than frequent online shoppers.

In parallel, Pauwels et al. (2011) discover that, in general, frequent online visitors spend more offline. Nevertheless, there was also a consumer group (entertainment oriented) that reduced their offline visits and revenues, since they replaced them online.

Later on, Schramm-Klein et al. (2011) have shown that positive evaluation of perceived multichannel integration favourably affects a retailer’s image, which in turn affects the frequency of multichannel use. Similarly, Bendoli (2005) discovered that high levels of multichannel integration perception lead to loyalty and purchase intention. Last but not least, Chen (2007) provides empirical evidence which indicates that enhanced multichannel store image increases purchase intention and, therefore, proposes integration & synergies across channels.

Drawing from the aforementioned literature findings, four store atmosphere & omnichannel criteria were selected (analytically presented in Table 2) in order to test whether in-store internet use by consumers affects the importance they attach to them in order to purchase from a physical store. Apart from the conventional & online store atmosphere, we included the human variable atmospheric component, in the form of the salesperson. We chose to empower this human variable with technologies, in order to test the impact of online cues on conventional ones. In addition to this, we also added the criterion of multichannel integration that creates a seamless shopping experience, in order to test the effect of omnichannel. Therefore, the following research hypotheses are formulated:

**H1.1:** There are statistically significant differences among consumers with different frequency levels of in-store internet use, in terms of the importance they attach to a physical store atmosphere, in order to purchase from that store.
**H1.2:** There are statistically significant differences among consumers with different frequency levels of in-store internet use, in terms of the importance they attach to a physical store’s service support offered by salespeople that utilize sales supporting technologies, in order to purchase from that store.

**H1.3:** There are statistically significant differences among consumers with different frequency levels of in-store internet use, in terms of the importance they attach to the online store’s atmosphere, in order to purchase from its physical store counterpart.

**H1.4:** There are statistically significant differences among consumers with different frequency levels of in-store internet use, in terms of the importance they attach to a store’s multichannel integration that creates a seamless shopping experience, in order to purchase from that physical store.

On the other hand, apart from the frequency of in-store internet use effects on the previous factors, subsequently, another research question emerges: which store atmosphere & omnichannel criteria are more important for consumers in order to purchase from a physical store?

Past literature findings reveal that conventional store atmospherics influence store patronage intentions (Grewal et al., 2003). In parallel, online store atmosphere influences online store patronage intentions (Kim et al., 2007).

Moreover, investigating the role of technology at the interface between salespeople and consumers, Ahearne & Rapp (2010) propose that salesperson technologies increase effectiveness & efficiency, whereas consumer-centric technologies reduce the need of a salesperson.

To that end, Mitchell (2010) discovers that store atmospherics amplify expectations of salespeople orientation and patronage intentions. Nonetheless, store atmospherics effects lessen at the introduction of salespeople.

Furthermore, Kwon & Lennon (2009) find out that online & offline store image are enriched via multichannel integration and affect customer loyalty and store patronage intention. Actually, offline store image affects online, which in turn affects online loyalty.

Next, Verhagen & van Dolen (2009) discover that offline store image (consisting of store atmosphere & service among other components) affects online purchase intentions. Moreover, online & offline store constitute the same multichannel experience and therefore retailers could benefit from a seamless multichannel integration.

Finally, Badrinarayanan et al. (2010) reveal that there are transference and congruence effects on purchase intentions in online stores from their offline counterparts.

Elaborating on these research insights, the following research hypotheses are formulated:

**H2.1:** There are statistically significant differences among different store atmosphere & omnichannel criteria, in terms of the importance that consumers attach to them in order to purchase from a physical store.

**H2.2:** There are statistically significant differences among different store atmosphere & omnichannel criteria, in terms of the importance that in-store Internet users attach to them in order to purchase from a physical store.

Beyond the previous research studies, it is obvious that omnichannel behaviour inside a physical store involves the application of online practices & technologies within the physical context. After all, according to Rigby (2011, p.4) that is the very essence of omnichannel retailing: blending the advantages of both physical stores & online shopping.

But, which online practices & technologies are more important for consumers inside physical stores? Business reports that have emerged recently, attempt to answer to this question (Wurmser, 2014). These reports either provide statistics about how retailers integrate channels inside the physical store, or depict what ICT retailing technologies consumers actually use in-store, giving emphasis on mobile technologies.

Academia has approached this question long ago in several studies. Burke (2002) investigated what technologies consumers want in the physical and the virtual store. He investigated 128 different
aspects of the shopping experience both online & in-store by conducting a national survey with 2,120 online users. He concluded that retailers should assist consumers to make the best use of every channel by seamlessly integrating them. Features & technologies of the shopping experience included sales assistance & customer service, digital signage & info kiosk, in order to assist product information, availability & pricing provision, support, payments and checkouts among others.

At the same period, Dahbolkar & Bagozzi (2002) explored technology-based self-service importance for consumers and find out that it contributes to enhancing user experience, if accompanied with ease of use and performance characteristics.

In addition to this, Ahn & Han (2004) discovered that a combination of online and offline features amplify the impact on usefulness, attitude, and intention to use, in comparison to each one on its own.

At another study, Andersson & Nordmark (2008) elaborated on the notion of the info kiosk as a means of “bringing the web to the shop floor”. They conclude that web features & multichannel integration are crucial for the in-store environment.

Recently, Magrath & McCormick, (2013) emphasized the importance of mobile apps as a marketing and sales channel.

The previous studies as well as the business insights reported above, led us to select specific online practices and technologies (analytically presented in Table 1), suitable for an omnichannel retailing environment, in order to explore the importance that consumers (both in total and in-store internet users only) attach to them within the offline channel. Thus, based on the previous literature findings, the following research hypotheses are formulated:

**H3.1:** There are statistically significant differences across a series of online practices & technologies applied inside physical stores, in terms of the importance that consumers attach to them.

**H3.2:** There are statistically significant differences across a series of online practices & technologies applied inside physical stores, in terms of the importance that in-store Internet users attach to them.

4 RESEARCH METHODOLOGY

The present study employed an exploratory quantitative empirical research design that took place in Greece in November 2014, in the context of the ELTRUN - The E-Business Research Center annual E-Commerce survey. The data collection instrument of the national survey was an online questionnaire which received 1022 answers from Internet users. The questionnaire was created in google forms and hosted on the research center web site. Internet users were informed of its existence via e-mail campaigns, display banners on popular Greek news sites & e-shops, social media, as well as from the hosting site directly.

Participants where questioned about their frequency of internet use across multiple channels, with the physical retail store included. Consumers were separated into different groups (quasi experimental design) regarding the frequency levels of their in-store internet use. Survey questions included various store atmosphere criteria & components, along with omniretailing technologies & practices (see Tables 1, 2). Specifically, store atmosphere criteria included both offline & online ones, with the inclusion of omniretailing characteristics, such as the technology-empowered salespeople and the multichannel integration in order to create a seamless shopping experience. In this way, we attempt to explore the link between store atmosphere and omnichannel retailing, regarding purchase intention towards the physical store.

In the context of this exploratory study, specific omniretailing technologies & practices were selected based on previous academic literature (Burke, 2002), as well as recent relevant business reports (Wurmser, 2014). It should be clarified that participants were asked about their preferences (importance attached) and not actual use of these technologies since most of them, though popular online, have not high penetration offline (because of either store availability or consumer mobile phone compatibility).
It should be noted that a similar survey took place in 2013 (also in the context of the ELTRUN’s annual E-Commerce survey), aiming to investigate the existence of omnishoppers & omniretailing practices. Based on those results (Lazaris et al. 2014), the present study aims to further clarify their preferences, purchase intentions & the importance they attach to online & offline environmental factors along with supporting e-commerce technologies.

5 FINDINGS AND DISCUSSION

5.1 Descriptive Statistics

Descriptive statistics reveal that mobile phones are the second device of choice for internet browsing, right after laptops, surpassing even desktops. In fact, 88% choose laptops for frequent internet use and 86% choose mobile phones. What’s more, 56% frequently use the internet on the move, whereas 41% use it frequently in the physical retail store. Only 21% of them have never used it in-store.

Regarding the four store atmosphere – omnichannel criteria, 89% of consumers regarded service support by salespeople utilizing sales supporting electronic technologies as very important in order to purchase from the physical store (Figure 1). In second place, the online store’s atmosphere collects 77% of consumers that attach high importance (Figure 2), which is marginally higher than the conventional store atmosphere (76%, Figure 3). Last place, though not low in percentage, is multichannel integration in order to create a seamless shopping experience, collecting 69% of our sample (Figure 4).

These results alone, show that the combination of the human-technology factor is regarded by the majority of internet users as the most important criterion in order to purchase from the physical retail store. Another striking result is the equal attention to online & offline store atmosphere. This could be attributed to the fact that they are both internet users and also possess high omnichannel retailing intensity. This intensity is also evident by the high percentage of importance they attach to multichannel integration.

Figure 1. The importance that consumers attach to service support by salespeople utilizing sales supporting technologies, in order to purchase from the physical retail store.

Figure 2. The importance that consumers attach to online store’s atmosphere, in order to purchase from the physical retail store.
As far as the online practices & technologies applied inside physical stores are concerned, it was discovered that the ones that gathered the greatest importance were #6, #16 and #3 (Table 1). It should be noted that although in-store price comparison, which could lead to showrooming (#6) scores lower at the mean ranks, it takes one of the first places among the number of consumers that regard it as very important. Showrooming, also encountered as free riding behaviour in multichannel literature (Heitz-Spahn, 2013), is the behaviour where a consumer researches a product or service at a retailer, only to purchase it from another one, at the same or different channel.
Online practices & technologies applied inside physical stores | % of consumers that attached great importance
--- | ---
6. In-store price comparison, which could lead to showroming | 77%
16. Fast electronic checkouts without queues | 77%
3. Ability to buy in-store with internet prices, as a result of an electronic check-in in the physical store | 74%
11. Product stock electronic availability | 69%
4. Special prices, coupons, offer alerts at the store’s entrance | 69%
12. Loyalty points electronic access | 61%
15. Product electronic search & map navigation to them | 57%
1. Free in-store wifi | 55%
18. In-store retail-assisting mobile site (accessible via wifi) or mobile app | 55%
8. Access to user opinions, product presentations & reviews | 53%
14. In-store location-based offers | 50%
13. Self-service assisting technologies | 47%
17. Mobile payments | 46%
5. Access to electronic profile & purchase history | 43%
7. Electronic recommender systems | 34%
2. Electronic check-in in the physical store (e.g. via wifi, foursquare, swarm, facebook, etc) | 32%
9. Product/service posts and comments on social networks | 29%
10. Email send & receive | 28%

Table 1. Online practices & technologies applied inside physical stores ranking, in terms of the percentage of consumers that attach great importance

5.2 Hypotheses Testing Results and Discussion

Non parametric statistical tests were employed due to the violation of the normality distribution assumption. Kruskal-Wallis H tests showed statistically significant differences among consumers with different frequency levels of in-store internet use in terms of the importance they attach among different store atmosphere & omnichannel criteria, in order to purchase from a physical store (Table 2). Specifically, significant differences are observed among consumer groups with different frequency levels of in-store Internet use regarding all the store atmosphere & omnichannel criteria. Thus, hypotheses 1.1 to 1.4 are accepted.

<table>
<thead>
<tr>
<th>Store atmosphere &amp; omnichannel criteria</th>
<th>Consumer groups with different frequency levels of in-store Internet use</th>
<th>N</th>
<th>Mean Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Criterion 1: Store’s conventional atmosphere</strong></td>
<td>Non In-Store Internet Consumers</td>
<td>186</td>
<td>369,97</td>
</tr>
<tr>
<td></td>
<td>Low In-Store Internet Consumers</td>
<td>287</td>
<td>419,76</td>
</tr>
<tr>
<td></td>
<td>Frequent In-Store Internet Consumers</td>
<td>215</td>
<td>414,28</td>
</tr>
<tr>
<td></td>
<td>Heavy In-Store Internet Consumers</td>
<td>108</td>
<td>359,73</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>796</td>
<td></td>
</tr>
</tbody>
</table>
**Criterion 2:** Service support by salespeople utilizing sales supporting electronic technologies

<table>
<thead>
<tr>
<th>Category</th>
<th>Non In-Store Internet Consumers</th>
<th>Low In-Store Internet Consumers</th>
<th>Frequent In-Store Internet Consumers</th>
<th>Heavy In-Store Internet Consumers</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>186</td>
<td>306</td>
<td>206</td>
<td>108</td>
<td>806</td>
</tr>
<tr>
<td></td>
<td>457.82</td>
<td>397.33</td>
<td>393.72</td>
<td>346.08</td>
<td></td>
</tr>
</tbody>
</table>

**Criterion 3:** Online store’s atmosphere

<table>
<thead>
<tr>
<th>Category</th>
<th>Non In-Store Internet Consumers</th>
<th>Low In-Store Internet Consumers</th>
<th>Frequent In-Store Internet Consumers</th>
<th>Heavy In-Store Internet Consumers</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>186</td>
<td>297</td>
<td>206</td>
<td>108</td>
<td>797</td>
</tr>
<tr>
<td></td>
<td>372.73</td>
<td>390.86</td>
<td>377.68</td>
<td>507.27</td>
<td></td>
</tr>
</tbody>
</table>

**Criterion 4:** Multichannel integration in order to create a seamless shopping experience

<table>
<thead>
<tr>
<th>Category</th>
<th>Non In-Store Internet Consumers</th>
<th>Low In-Store Internet Consumers</th>
<th>Frequent In-Store Internet Consumers</th>
<th>Heavy In-Store Internet Consumers</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>186</td>
<td>296</td>
<td>206</td>
<td>118</td>
<td>806</td>
</tr>
<tr>
<td></td>
<td>353.79</td>
<td>433.84</td>
<td>399.01</td>
<td>413.58</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Chi-Square</strong></th>
<th><strong>df</strong></th>
<th><strong>Asymp. Sig.</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Criterion 1:</td>
<td>11,409</td>
<td>.010</td>
</tr>
<tr>
<td>Criterion 2:</td>
<td>21,546</td>
<td>.000</td>
</tr>
<tr>
<td>Criterion 3:</td>
<td>32,641</td>
<td>.000</td>
</tr>
<tr>
<td>Criterion 4:</td>
<td>15,217</td>
<td>.002</td>
</tr>
</tbody>
</table>

| **Table 2.** | **Kruskal-Wallis test statistics for Hypotheses H1.1 & H1.2** |

Regarding Criterion 1, post-hoc pairwise comparisons show that the null hypothesis is rejected, since the differences observed among groups are significantly important. Nevertheless, there are no statistically significant differences observed between Non In-Store Internet Consumers and Heavy In-Store Internet Consumers. These results could be explained through the notion of atmospheric responsiveness, which is the tendency to behave according to environmental factors. Heavy In-Store Internet Consumers seem to be more online-oriented and therefore score lower at offline cues. On the other hand, Non In-Store Internet Consumers could mean that this group is more utilitarian in-store and don’t care about the shopping environment.

As far as Criterion 2 is concerned, it is evident that significant statistical differences exist only between Non In-Store Internet Consumers and all the others. Also, this consumer group scored the highest mean rank, which implies that the need for service support by salespeople utilizing sales supporting electronic technologies decreases, as the frequency of in-store internet use by consumers increases (they are probably served by themselves).

In contrast, Criterion 3 test shows that significant statistical differences exist only between Heavy In-Store Internet Consumers and all the others. Also, this consumer group scored the highest mean rank, which implies that as the frequency of in-store internet use by consumers increases, they start to pay more attention to online atmospheric components, even in order to purchase from a physical store.

Finally, Criterion 4 post-hoc pairwise comparisons show that significant statistical differences only exist between Non In-Store Internet Consumers and Low In-Store Internet Consumers, which could be attributed to the fact that Low In-Store Internet Consumers have a more balanced approach towards multichannel perception, whereas the other consumer groups use either online or offline channels at most. Therefore, we expect that significant statistical differences are more probable to be found.
between this group and the offline channel one (since the other online groups are also multichannel, too).

Regarding hypotheses 2.1, a Kruskal-Wallis H test showed that there is a statistically significant difference among these criteria ($\chi^2(3) = 194,758, p = 0.000$), with a mean rank importance score of 2312.99 for service support by salespeople utilizing sales supporting electronic technologies (criterion 2), 1874.82 for multichannel integration in order to create a seamless shopping experience (criterion 4), 1871.26 for online store’s atmosphere (criterion 3) and 1673.13 for store’s conventional atmosphere (criterion 1). Thus, hypothesis 2.1 is accepted.

Post-hoc pairwise comparisons revealed that there are significant statistical differences between all criteria except #3 & #4, which could be an indication that online atmosphere are almost equally important to multichannel integration in order to provide a seamless shopping experience in the physical store.

Regarding hypothesis 2.2 (i.e. in-store Internet users), the results are more or less the same ($\chi^2(3) = 105,963, p = 0.000$), except for the lower mean rank importance scores. Specifically, there was a mean rank importance score of 1799.13 for service support by salespeople utilizing sales supporting electronic technologies (criterion 2), 1559.00 for multichannel integration in order to create a seamless shopping experience (criterion 4), 1521.92 for online store’s atmosphere (criterion 3) and 1365.92 for store’s conventional atmosphere (criterion 1). Thus, hypothesis 2.2 is accepted.

Nevertheless, for this target group it is observed that pairwise statistical differences between criteria #3 & #4 are roughly the same, whereas pairwise statistical differences between criteria #2 & all the others are decreased. This could also be an indication that in-store internet consumers attach less importance to salespeople utilizing sales supporting electronic technologies in comparison to the other criteria, than the hypothesis 2.1 group (i.e. all consumers).

Furthermore, a Kruskal-Wallis H test for testing hypothesis 3.1 showed that there is a statistically significant difference on the importance score that consumers attach across a series of online practices & technologies applied inside physical stores ($\chi^2(17) = 1890.866, p = 0.000$), with the mean rank importance scores reported in Table 3. Thus, hypothesis 3.1 is accepted.

<table>
<thead>
<tr>
<th>Online practices &amp; technologies applied inside physical stores</th>
<th>N</th>
<th>Mean Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Importance attached by consumers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Fast electronic checkouts without queues</td>
<td>914</td>
<td>10882.56</td>
</tr>
<tr>
<td>3. Ability to buy in-store with internet prices, as a result of an electronic check-in in the physical store</td>
<td>972</td>
<td>10632.88</td>
</tr>
<tr>
<td>6. In-store price comparison, which could lead to showrooming</td>
<td>924</td>
<td>10108.59</td>
</tr>
<tr>
<td>11. Product stock electronic availability</td>
<td>894</td>
<td>10034.16</td>
</tr>
<tr>
<td>4. Special prices, coupons, offer alerts at the store’s entrance</td>
<td>914</td>
<td>9426.63</td>
</tr>
<tr>
<td>12. Loyalty points electronic access</td>
<td>913</td>
<td>9157.78</td>
</tr>
<tr>
<td>1. Free in-store wifi</td>
<td>952</td>
<td>8791.43</td>
</tr>
<tr>
<td>15. Product electronic search &amp; map navigation to them</td>
<td>894</td>
<td>8560.32</td>
</tr>
<tr>
<td>18. In-store retail-assisting mobile site (accessible via wifi) or mobile app</td>
<td>913</td>
<td>8284.34</td>
</tr>
<tr>
<td>14. In-store location-based offers</td>
<td>904</td>
<td>8061.67</td>
</tr>
<tr>
<td>17. Mobile payments</td>
<td>904</td>
<td>8015.85</td>
</tr>
<tr>
<td>8. Access to user opinions, product presentations &amp; reviews</td>
<td>924</td>
<td>7960.35</td>
</tr>
<tr>
<td>13. Self-service assisting technologies</td>
<td>924</td>
<td>7577.87</td>
</tr>
<tr>
<td>5. Access to electronic profile &amp; purchase history</td>
<td>904</td>
<td>7159.09</td>
</tr>
<tr>
<td>7. Electronic recommender systems</td>
<td>914</td>
<td>6369.36</td>
</tr>
<tr>
<td>9. Product/service posts and comments on social networks</td>
<td>894</td>
<td>5920.61</td>
</tr>
<tr>
<td>10. Email send &amp; receive</td>
<td>914</td>
<td>5916.97</td>
</tr>
</tbody>
</table>
2. Electronic check-in in the physical store (e.g. via wifi, foursquare, swarm, facebook, etc)

<table>
<thead>
<tr>
<th></th>
<th>All Consumers</th>
<th>Internet In-Store Consumers</th>
<th>Non-Internet In-Store Consumers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>N</td>
</tr>
<tr>
<td>Online practices &amp; technologies</td>
<td>16</td>
<td>914</td>
<td>4.07</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>972</td>
<td>3.99</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>924</td>
<td>3.88</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>894</td>
<td>3.84</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>914</td>
<td>3.68</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>913</td>
<td>3.62</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>952</td>
<td>3.53</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>994</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>913</td>
<td>3.37</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>904</td>
<td>3.35</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>924</td>
<td>3.31</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>904</td>
<td>3.29</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>924</td>
<td>3.21</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>904</td>
<td>3.06</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>914</td>
<td>2.85</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>914</td>
<td>2.76</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>894</td>
<td>2.73</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>954</td>
<td>2.61</td>
</tr>
</tbody>
</table>

Table 3. *Kruskal-Wallis test statistics for Hypotheses H3.1*

Post-hoc pairwise comparisons revealed that there are no significant statistical differences between online practices & technologies #7, #9, #10 & #2, which have the lowest mean ranks. Therefore, these are equally considered by consumers as less important in comparison to the others. Next, online practices & technologies #15, #18, #14, #17, #8, #13, #5 are positioned in the middle of importance by rank, and #17, #8, #13, #5 have no significant statistical differences between the other three (#15, #18, #14). All the other ones have significant statistical differences between the rest, with #16, #3, #6 & #11 being considered as the most important ones by consumers. In fact, #6 (In-store price comparison, which could lead to showroming) scores the highest percentage among the rest, according to the number of consumers that consider it of primary importance in individual scores.

Finally, regarding hypothesis 3.2 (i.e. in-store Internet users), the results are more or less the same ($\chi^2(17) = 1627.702, p = 0.000$), except for the generally higher mean scores (Table 4). Nevertheless, for this group it can be observed that online practices & technologies #1 & #17 have each risen by a single rank in the final classification. This could be attributed to the fact that this target group attaches greater importance to in-store internet facilities (#1, free in-store wifi) & therefore wants to use mobile devices in order to satisfy the primary importance of fast electronic checkouts without queues (#16) via mobile payments (#17). In contrast, consumers that don’t use internet in-store have scored significantly lower for free in-store wifi, something which is self-justified (Table 4). Thus, hypothesis 3.2 is also accepted.
6 CONCLUSIONS AND IMPLICATIONS

6.1 Conclusions

Several interesting conclusions and corresponding implications are derived through the present study. These mainly refer to the promising role that in-store online practices, retail technologies and employees could play in the in-store consumers' shopping process.

Specifically, it seems that some online practices-services offered/applicable within the physical retail store (e.g. fast electronic checkouts without queues, ability to buy in-store with internet prices as a result of an electronic check-in in the physical store and in-store price comparison, which could lead to showroombing, etc.) are more attractive and promising than others, since consumers (both in-store Internet users and non-users) attach greater importance to them. Also, electronic check-in in the physical store is not desirable, unless it offers important benefits (e.g. price matching). Moreover, it seems that consumers are more eager to adopt cost-cutting technologies, than shopping-assisting or experience-enhancing technologies. In sum, it is clear that both in-store Internet users and non-users do not possess significant differences as far as the importance they attach to these practices (with some slightly differences observed in some online practices, as discussed in the previous section).

In parallel, it seems that for both in-store Internet users and non-users, the service support offered by salespeople that utilize sales supporting technologies is the most important store atmosphere relevant feature. While it was expected that especially for in-store Internet users this feature would not be so important (in the sense that they could potentially execute these tasks by themselves through their mobile phones), the fact that also this group attach the greatest importance to this particular attribute/feature highlights the promising role that salespeople (equipped with technology) could play in the physical retail store. Nevertheless, it is evident that heavy in-store internet users seem to care less for technology-empowered employees, probably because they serve their own needs appropriately utilizing their own technology. Also, this group attaches the greatest importance to the store’s online atmosphere, which comes at no surprise since they probably admire the store’s online characteristics highly due to in-store internet use. In fact, the online & the offline store atmosphere are generally perceived of equal importance by all participants. Specifically, the online one gathers 77% of consumers that attach high importance to it (Figure 2), in comparison to 76% for the conventional store atmosphere (Figure 3).

6.2 Managerial Implications

The findings strengthens the implications to practitioners regarding the selection of the most attractive and promising in-store technologies that they could first adopt in their physical retail stores in order to satisfy their customers. Besides, comparing the study results with current business practice, it seems that several innovative retailers have, more or less, already adopted several of the aforementioned online practices/services.

Nevertheless, since in some cases ICT in-store technologies may empower consumers and not employees, a retailing “paradox” could be potentially created: consumers utilizing smartphones and integrating channels could outperform employees who may feel helpless inside their own shopping environment (i.e. physical retail stores) and, thus, cannot support consumers effectively. In other words, in these cases, salespeople can only depend on their existing experience (e.g. personal selling techniques). However, ICT (e.g. mobile technologies) could prove to be important assets also for the salespeople, in order for them to provide enhanced and innovative services and effective support to consumers during their in-store visit (as also derived by the present study results). Thus, it could be suggested to retailers to invest on educating and training their salespeople along with offering them access to Information and Communication Technologies infrastructure (e.g. CRM and ERP systems, smart devices). In addition to this, multichannel retailers should pay increased attention to their online stores’ atmosphere along with their conventional counterpart, since it is considered to be equally
influential among consumers. In sum, since retailers are trying to find ways to adjust their presence in the context of the omnichannel retailing era (by exploring which innovative technologies to adopt), it seems that they should also place emphasis on the role of their employees (e.g. in-store salespeople equipped with technology) in order to offer value added services to their physical store visitors (i.e. services that could not be offered exclusively through technology).

6.3 Limitations

It should be underlined that, since the study was conducted online, the results cannot be generalized to non-Internet users. Apparently, several offline retail store visitors do not use the Internet in general (i.e. either within a physical retail store or not), and, thus, they may not intend to adopt omnichannel retail practices, in-store retail technologies, etc. (at least for the near future). However, also for this group, the role of salespeople may be important (e.g. provision of detailed product information, customer service, etc.). Last but not least, we should not underestimate the potential effects of the economic crisis in Greece. To this end, a multinational study could reveal different preferences among omniretailing technologies, which could possibly outscore cost-cutting ones.

6.4 Future Research Directions

Future research could further investigate this topic through several perspectives and approaches. Indicatively, some of the online practices and the store atmosphere relevant criteria reported in the paper could serve as the treatments of an experimental design aiming to investigate causal relationships between retailers' omnichannel retail practices and consumer behavioural patterns and evaluations. Also, an interesting research direction could be to investigate showrooming behaviour at the presence of omniretailing technologies at the hand of both consumers and salespeople, in order to test which scenario is more effective for coping with such behaviour. Similarly, qualitative research approaches (e.g. in depth personal interviews) could also contribute towards thoroughly exploring and understanding consumers' concerns, needs and intentions in the context of the evolving landscape of omnichannel retailing.

References


IMPROVING THE VALUE OF IS – SELF PARTICIPATION AND HEALTHY LIFESTYLE

Anton Manfreda, University of Ljubljana, Faculty of Economics, Academic Unit for Business Informatics and Logistics, Kardeljeva ploscad 17, Ljubljana, Slovenia, anton.manfreda@ef.uni-lj.si

Abstract

Numerous studies have been done in the last decades to promote IS and to present how IS can increase the business value. However, IS departments are still often considered merely as a secondary activity in many organisations. Consequently, a perception of IS as a cost instead of an enabler of business value is still frequently present in many companies. Nevertheless, by the advent of social networks new opportunities have emerged for researchers and organisations to recognize the new value of IS. Thus, there is a strong need to improve the importance of IS by presenting the value of IS from a different perspective. The purpose of this research in progress is therefore to present IS as an enabler and important element of a better future and consequently to emphasize a different perspective that may also increase the business value of IS. More than 300 students were invited to participate in the research allowing to present the value of IS in a way that improves the healthy style of living.

Keywords: IS value, healthy lifestyle, motivation, digital services

1 INTRODUCTION

Studying the influence of IS on the business value has been a main challenge for researchers in the last few decades (Luo et al., 2012) with the focus on promoting IS and presenting how it can increase business value. Therefore, several attempts have been made to improve the relationship between IS departments and business departments by presenting the importance of IS (Milis et al., 2008). Researchers were devoting considerable effort to this issue due to the extension and consequences of inefficient business-IS relationship on the successful IS implementation projects. Presenting IS as an enabler of business process improvement was often seen as a solution for the problematic business-IS relationship.

However, despite several efforts to overcome the problem of treating IS merely as a cost in organizations, it seems that this problems still remains an important issue not being properly solved. An important and often neglected issue is the perceived value of IS which still remains vaguely answered.

It has been suggested that due to the important role of IS presenting the value of investing in IS should be an important contribution of the IS discipline (Agarwal and Lucas Jr, 2005). Therefore, several researchers have been motivated to promote the value of IS. However, they were focusing mostly on the business value of IS and regardless their effort the problems in the business-IS relationship that were existing a decade ago are still present. Therefore, there is a strong need to present a value of IS from a different perspective, namely to present IS as an enabler and an important element of a better future.

Nevertheless, by the popularity and increased usage of different digital services and social networks also new opportunities are emerging for organizations and researchers. These opportunities are enabling to identify the new value of IS, namely presenting IS as a tool that helps people live healthier. The purpose of this research in progress is thus to focus on how IS solutions can help people to improve their healthy life. Several students participated in the research which will help us to recognize explanations and factors for considering or not considering healthier way of living and finally to present how IS can be used as a tool for increasing the importance of a healthier life.
The paper is divided into four main parts. It begins by a short theoretical background presenting the value of IS and the gamification concept. Second, the research methodology is briefly described, followed by the data analysis. At the end, concluding remarks are presented along with research limitations and future research opportunities.

2 THEORETICAL BACKGROUND

2.1 The value of IS

Presenting the business value of IS has been a main challenge for researchers in the last few decades (Pang, 2014, Luo et al., 2012). It has been claimed that presenting the value and importance of investing in information systems is a particularly important contribution of the IS theoretical field (Agarwal and Lucas Jr, 2005). Therefore, several researchers have been motivated to understand the influence of IS on the improved organizational performance (Melville et al., 2004).

It has been argued that IS enable business process reengineering, strategic alliances and competitive advantages (Avison et al., 1999), and can consequently present the value to the organization (McKeen and Smith, 1996). Moreover, they also help organizations to be innovative by providing appropriate infrastructures and consequently sustaining competitiveness (Hewitt, 1995). Nevertheless, IS may create business value by enabling efficient business processes and by enabling organizations to perform their activities more efficiently compared to their competitors (Luo et al., 2012).

However, opportunities for obtaining strategic advantages exclusively based on IS are slowly disappearing since merely technology itself is not a sufficient factor for successful IS implementation (Dhillon, 2008). Organizations should focus on organising IS instead of focusing just on technology in order to obtain a sustainable competitive advantage (Bharadwaj, 2000). Nevertheless, companies with the largest IS investment were not performing the best financial results (Carr, 2003) signifying that IS has become a commodity for organizations and therefore does not create a competitive advantage by itself.

2.2 Gamification as a new opportunity

Gamification emerged in 2010 and can be defined as using game elements and game design techniques in a non-game contexts (Deterding et al., 2011) with the purpose to increase individuals’ commitment. The concept of gamification has been mainly used in the marketing; however, its application has been extended also to other areas like environment, energy, health or education (Deterding et al., 2011).

Gamification is closely connected with social elements, and therefore it has the greatest impact on the generation of people who use and share their everyday life in social networks (Simões et al., 2013). Several features are namely common to social games like user loyalty or different achievements. Moreover, it has been even suggested that motivational benefits of gamification should be coupled with the collaborative and participative capabilities that are provided by social networks (De-Marcos et al., 2014).

It has been also claimed that games have encouraged management scholars to propose the gamification concept for motivating employees (Zichermann and Cunningham, 2011). Based on this it is possible to hypothesize that using gamification concept coupled with proper software applications can help people to improve their healthy life or at least engagement in using these applications. Gamification concept is therefore also changing the perception and the value of IS.

Based on the literature review and practical observations the following research problem was exposed: Certain digital services may present IS as an enabler and important element of a better future by improving the healthy lifestyle with a special emphasis on the self-participation.
3 RESEARCH METHODOLOGY

The research question was empirically tested on Slovenian students. A questionnaire was therefore developed and particular software applications intended for increasing a healthier way of living were included in the questionnaire, namely MyFitnessPal and OPEN platform for clinical nutrition. The questionnaire was, among other indicators that are not relevant for this research, composed of several items measuring the physical activity of participants, items measuring motivation for healthy way of living, items describing MyFitnessPal application and OPEN platform for clinical nutrition and their possibilities and also items measuring participants’ satisfaction after a quick overview of selected applications.

Data collection started in 2014 and will be repeated in 2015. Approximately 500 students were invited to participate in the research in order to discover factors that are important for improving motivation towards using these solutions. A smaller number of students have already been invited to participate in the detailed testing of both applications with the intention to increase awareness of healthy lifestyles and to obtain valuable feedback regarding the motivational factors for a long-term use.

The research thus involves interviews with selected students in Slovenia and online participation of larger number of undergraduate students. The purpose of developing a survey is to present reasons for taking or not taking care for a healthier way of living and to test to which extent IS solutions can open the awareness towards healthier lifestyle. The research has two main phases. The first phase refers to completing the survey, while the second phase refers to the results obtained from analysing the smaller group of students and their detailed usage of selected IS solutions.

4 DATA ANALYSIS

An exploratory factor analysis using SPSS will be conducted to define factors that are important for considering a healthier way of living and to present to which extent particular digital services can open the awareness towards healthier lifestyle. However, some preliminary results are evident from the tables below. Table 1 presents the profile of the respondents which is influenced by the whole student population where female students are predominating.

<table>
<thead>
<tr>
<th>Students</th>
<th>Number</th>
<th>Percentages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>235</td>
<td>61%</td>
</tr>
<tr>
<td>Male</td>
<td>144</td>
<td>39%</td>
</tr>
<tr>
<td>Total</td>
<td>379</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 1. Profile of respondents.

Figure 1 presents additional information of the respondents’ profile, namely their physical activity in the free time (recreational or professional). Physical activity is included in the profile since it is an important element when considering how particular digital services influence the awareness of healthier lifestyle. From the figure below it is evident that the majority of respondents are physically active; however 5% of the respondents are not active, where the main reason for not being active is either lack of time or lack of will.
Beside physical activity proper nutrition is also an important element of healthy way of living and should also be considered when examining the influence of particular digital services on the awareness towards healthier lifestyle. Figure 2 shows that the majority of respondents are aware of the proper nutrition and take care for it; however 35% of students do not care for proper nutrition.

The main reasons for caring about proper nutrition are practicing sports activities, the awareness of the harmful effects of malnutrition and maintaining proper weight. Contrary, the main reasons for not caring about proper nutrition are lack of time to prepare proper victuals, lack of information regarding proper nutrition and lack of relevant software applications that would encourage people towards proper nutrition.

The following tables therefore refer to testing the particular digital services that are intended to improve healthier way of living. Table 2 presents the main reasons for possible using one of the tested digital services in the near future and the main motivation elements for using it. As it is evident from the table, caring for maintaining appropriate weight and easiness to use are the most important reasons for using one of the presented digital services, while allergies to certain ingredients are not seen to be an influential reason. The latter observation was not anticipated since the OPEN platform for clinical nutrition was originally developed especially for helping people with particular allergies or diseases. However, respondents without any allergies to certain ingredients may not perceive it as important since there is no need for it.
Reason | Number | Percentages
--- | --- | ---
Malaise. | 15 | 13%
Allergies to certain ingredients. | 5 | 4%
Digital service is easy to use. | 51 | 44%
Caring for maintaining appropriate weight. | 89 | 77%
Improving mental and physical capabilities. | 62 | 53%
Other. | 7 | 6%

**Table 2.** The main possible reasons for using one of presented digital services.

Contrary, Table 3 presents the main reasons for not using one of the digital services in the near future. The majority of students who will probably not use one of the presented tools in the near future selected that there is no need to plan the diet or no sense in using these tools for planning the meals as the main reasons for their declination.

<table>
<thead>
<tr>
<th>Reason</th>
<th>Number</th>
<th>Percentages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Using other tools or digital services.</td>
<td>17</td>
<td>9%</td>
</tr>
<tr>
<td>Not seeing the point in using tools for planning meals.</td>
<td>79</td>
<td>40%</td>
</tr>
<tr>
<td>No need to plan the diet.</td>
<td>88</td>
<td>44%</td>
</tr>
<tr>
<td>My diet seems to me totally acceptable.</td>
<td>85</td>
<td>43%</td>
</tr>
<tr>
<td>I do not care how and what I eat.</td>
<td>3</td>
<td>2%</td>
</tr>
<tr>
<td>Other.</td>
<td>19</td>
<td>10%</td>
</tr>
</tbody>
</table>

**Table 3.** The main reasons for not using one of the presented digital services.

Students selecting that they will probably not use one of the presented digital services were additionally asked for presenting the main elements that could motivate them to use these tools for planning the meals. Their answers are presented in the Table 4.

As it is evident from the table below positive experiences of their friends are important influential factor. The latter may be explained by the participating persons in the survey. Respondents were namely younger people and for them friends are important social and influential category. Beside positive experiences, also possible disease is seen as an important factor for using one of the presented tools. Moreover, simple user interface and possible sports activities are additional important factors to consider.

Contrary, competing-orientated tools (collecting points, bonuses…) as an example of gamification elements are not one of the most important factors, however in the combination with positive experiences of their friends may become an important motivational factor. Nevertheless, gamification is related to the social networks where experiences of others are central. Therefore, proper development of these tools that would include gamification elements and sharing some of them on the social networks may even improve the positive experience. Consequently, it will increase the motivation for possible usage even within the groups that would otherwise not care about their way of living.
Table 4. The main possible motivational elements for using one of the presented tools.

<table>
<thead>
<tr>
<th>Motivational factor</th>
<th>Number</th>
<th>Percentages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nothing, I could not be persuaded to use.</td>
<td>30</td>
<td>15%</td>
</tr>
<tr>
<td>Any disease.</td>
<td>77</td>
<td>39%</td>
</tr>
<tr>
<td>Possible sports activities.</td>
<td>46</td>
<td>23%</td>
</tr>
<tr>
<td>Positive experiences of my friends.</td>
<td>84</td>
<td>42%</td>
</tr>
<tr>
<td>Simple user interface.</td>
<td>47</td>
<td>24%</td>
</tr>
<tr>
<td>More advertising.</td>
<td>29</td>
<td>15%</td>
</tr>
<tr>
<td>Competing-orientated tools (collecting points, bonus ...).</td>
<td>29</td>
<td>15%</td>
</tr>
<tr>
<td>Other.</td>
<td>11</td>
<td>6%</td>
</tr>
</tbody>
</table>

5 CONCLUSION

In order to use incoming technology and new solutions it is important to improve the position of IS by presenting and emphasizing its value. Since presenting the business value of IS in the last few decades was not as successful as anticipated it is necessary to present the value from a different perspective.

The research will thus enlighten the value of IS by presenting how IS solutions can help people to improve their healthy life and will present IS as a tool and essential part for developing a better future. Instead of merely emphasizing the business value of IS, this research will focus on presenting the value of IS by emphasizing the common welfare that can be achieved through different IS solutions.

The research question will be empirically tested using the data from participating students. This will enable to identify motivational factors for considering a healthier way of living, and therefore to improve their healthy life. All observations will also be used to promote OPEN platform for clinical nutrition and to increase a general awareness regarding the healthy life.

Research and study results have also some limitations since they are limited to students in a single country; however the research is easily applied to other interested groups. Further, research data were mostly obtained by online questionnaire where certain level of bias exist (Dillman et al., 2008).

5.1 Future research

Since this is a research in progress a future research refers to defining factors that are important for considering a healthier way of living and developing a model that could show to which extent particular digital services can increase the awareness towards healthier lifestyle. Future research is also needed in different regions and populations in order to cross-validate the research findings.

References


A GAME-BASED MOBILE APPLICATION ON ENGLISH VOCABULARY ACQUISITION: CROSS-SITUATIONAL LEARNING APPROACH

Sevgi Özkan Yıldırım, Department of Information Systems, METU, sevgiozk@metu.edu.tr

Abstract

The advances in Information and Communication Technologies offer a platform for educational activities. For example digital games are used in educational context as motivating learning tools. Language learning is one of the educational fields, which has been affected from technological improvements. Digital games can be effective learning tools while learning English as a second language. In the scope of this study a mobile vocabulary learning game will be developed and evaluated. The game will be designed by taking cross-situational learning as a theoretical background. The game will be evaluated by conducting experiments and survey. There will be two experiments including pre-test and post-test phases. In each phase, the participants will use the game, after that they will take vocabulary quizzes and fill survey to evaluate the game. According to the results of the first phase, the game will be maintained and then the post-test will be conducted. At the end of the study, the game will be compared other vocabulary games in effectiveness, usability and contributions to the subject area.

Keywords: mobile learning game, vocabulary acquisition, cross-situational learning, vocabulary game

1 INTRODUCTION: MOBILE GAME-BASED LEARNING

1.1 Introduction

With the great improvements in Information and Communication Technologies (ICT), it is inevitable to use ICT applications in our lives. Education is one of the affected areas from technology. E-learning is the reflection of use of technology in learning activities. E-learning is a wide concept and contains different applications: online learning management systems, simulations, quizzes, educational games are some examples. Computer game-based learning applications are tools used in different fields of education to make the learning and teaching process more enjoyable. According to the researches in education domain, computer games provide an enjoyable and comfortable learning environment to the students and improve their problem solving skills (Ebner & Holzinger, 2007). In addition to this, games provides an effective learning environment by enabling students to use their prior knowledge, giving immediate feedback and giving active roles to the students (Oblinger, 2004). Also computer games are seen as motivating learning tools by the learners (Kirriemuir & McFarlane, 2004).

Language teaching is one of the educational contexts, which has been affected from the technological improvements. Computer aided language learning (CALL) is using technology especially multimedia applications to support different areas of language learning, such as grammar, writing, and vocabulary acquisition. Computer games are one of the examples of CALL applications in language learning (Levy, 2009).
Vocabulary acquisition while learning a second language is seen as a difficult part of learning the language. Vocabulary acquisition is important, since without vocabulary, the language cannot be used effectively. In order to use a second language, the learner needs to know 3000 words to understand written texts and spoken expressions (Laufer, 1997). This means that learners of a language should extend their vocabulary to communicate meaningfully that language. To enhance learners’ vocabulary acquisition in learning a second language, games can be used as effective vocabulary learning tools: games can take the attention of students and motivate them because they like playing games (Yip & Kwan, 2006).

In the scope of this study, a mobile game-based application to be used in teaching English as a second language will be developed and assessed. The second language is chosen as English, because in Turkey, English is one of the compulsory courses in most of the primary and high schools and in some universities as well as English is one of the mostly used languages through all over the world.

The game will be designed and developed by taking the cross-situational learning as a theoretical framework. Cross-situational learning is a strategy in which a word and its referent are tracked in multiple trials and the correct mapping is fixed by using the statistical information gathered from these trials. Cross-situational learning is proposed as an effective learning strategy to learn the pairs of word and its referent. The referent may be meaning of that word or an object that refers to the word. The claim is that human learners employ cross-situational learning while learning words.

The game will be a matching game. The players will match an English word with a picture of an object with its Turkish equivalent. The game will be developed to be run on mobile devices such as tablets in order to benefit from the advantages of mobile technology. Also the game will be able to be played online by the players.

1.2 Purpose of the Study

In this study a game-based mobile application will be developed to be used in teaching English as a second language by aiming:

1. To improve the students vocabulary acquisition while learning English as a second language.
2. To contribute to the use of instructional technologies in language learning.
3. To apply Cross-Situational learning in a learning game while learning English as a second language.
4. To validate the cross-situational learning via a vocabulary learning game.
5. To contribute to the game-based learning literature.
6. To contribute to the mobile learning literature.

1.3 Problem Statement

Vocabulary acquisition while learning English as a second language is seen difficult and boring by students. This difficult part of language learning can be enhanced by integrating technology to the learning process. Especially digital games can be helpful to make this boring part of learning more enjoyable since digital games are seen as motivating learning tools by the students. Mobile technologies offer a good platform for game applications. In this context, a mobile game-based vocabulary acquisition tool will be developed by considering cross-situational learning as a theoretical
framework which is an effective learning strategy to map the words with their referents. The game will be evaluated by conducting experiments and comparing with other types vocabulary games.

1.4 Research Questions
1. Is use of games in vocabulary acquisition effective while learning English as a second language?
2. What is the effect of using game in learning English vocabularies?
3. Can the cross-situational learning be proved in learning English as second language?
4. Is the cross-situational learning an effective learning strategy while learning English as a second language?
5. Do the students like playing games while learning English as a second language?
6. What are the advantages of mobile games over traditional games and PC games in English learning?
7. Do students prefer playing games to traditional classroom activities in vocabulary acquisition?

1.5 Study Plan
Study plan includes the tasks to be completed in the scope of the study. The Figure 1 shows these tasks. The first part of the study is literature review which will continue until the end of the study. This part will cover the topics game-based learning, cross-situational learning and games in vocabulary acquisition in English learning. The second part of the study will cover game development which consists of analysis, design, implementation and testing sub steps. These sub steps cover the software development process of the game. In the third step, the game will be evaluated by conducting experiments and survey. The effectiveness of the game and the theoretical approach will be evaluated. Lastly, the game will be assessed by comparing to the other games and by its contribution to the related area.

2 Overview of the Related Literature

2.1 Digital Game-based Learning
First, a definition for game will be given, than games will be mentioned as instructional tools. There are different definitions for game in the literature. The following is a general definition for “game”:

“A game is a rule-based formal system with a variable and quantifiable outcome, where different outcomes are assigned different values, the player exerts effort in order to influence the outcome, the player feels attached to the outcome, and the consequences of the activity are optional and negotiable ” (Juul, 2003).

According to this definition, there are different parameters of a game such as effort, rule, and outcome. However, according to the Wittgenstein (1953), all games do not have common properties; they can be in the same semantic category. Also Garris, Ahlers and Driskell (2002) gave different characteristics of a game defined in the literature, such as interaction, challenge, control, complexity and so on. It is hard to generalize the characteristics of games, so that researchers categorized the characteristics of the games. Also the games are categorized by Gros (2007) according to genre types: actions games, adventure games, fighting games, RPG, simulations, sport games, strategy games. Here, actually we are dealing with the two categories of games in addition to their common characteristics and genre types: games for entertainment and games for learning. In addition to entertainment purpose, games have educational values (such games are called as learning games). Games are more than entertainment when we use them in educational context. According to Piaget (1962) playing games make the children close to the environment in which they live and imagine.
Digital game-based learning has emerged after technology had become central to our lives and use of computer in education has not motivated the students anymore because of the wide use of computers in anywhere (Kiili, 2005). The rationale behind learning games is adding fun to the learning activities and motivating the students. According to Klopfer, Osterweil, and Salen (2009) digital learning games are different from games for entertainment in acquisition of knowledge and they foster learning in academic context. When games are used in classrooms, they add fun to the learning and students feel more flexible and comfortable (Uzun, 2009). According to Robertson and Howells (2008), digital games provide a powerful learning environment. Squire (2002) said that games have educational potential from both cognitive and social perspectives. Computer games enable students to repeat their prior knowledge, give immediate feedback and provide an interactive learning environment to them (Ebner & Holzinger, 2007). According to the Kirriemuir & McFarlane (2004), there are two key themes for developing digital games for education which are motivational power of the games for making fun and seeing games as powerful learning tools. As seen, digital games for education has been accepted as motivational learning tools.
With advances in technology, mobile technology becomes an effective platform for games. In addition to the PC game applications, mobile gaming also has been accepted as powerful tool to support educational activities such as learning and teaching (Facer et al., 2004). In the study (Thornton & Houser, 2005) which is related to the use of mobile devices in teaching, the results showed that students evaluates use of mobile devices positively and mobile devices can be powerful tools to support a wide range of teaching activities. Games can be one of the applications running on mobile devices to support educational activities. Schwabe & Göth (2005) conducted a study with a mobile game to support learning. They have concluded that the mobile technologies and mobile games may be used to support learning; however the applications should be designed carefully.

2.2 Language Learning: Vocabulary Acquisition

In language learning, there are different areas and skills which are grammar, vocabulary, reading, writing, pronunciation, listening, speaking and culture (Levy, 2009). Learning vocabulary is seen as an important part of learning a second language, because according to Laufer (1997), the learners should know 3000 words to understand written and spoken texts in a language. This implies that the learners of a language should extend their vocabulary. Also according to McCarthy (1990), even if you know grammar well, you cannot communicate meaningfully in a language without having enough words. Vocabulary acquisition is seen as the most difficult part of learning a second language and students usually think that learning new vocabularies is boring (Lo & Tseng, 2011).

2.3 Language learning with Technology Support

With the advances in technology and its integration into educational context facilitates its use in language teaching. Computer aided language learning (CALL) is the use of computers and digital applications especially multimedia applications to support language learning. There are different CALL applications to support different areas in language learning (Levy, 2009). Vocabulary acquisition is one of the focuses of CALL applications and there are different CALL applications to support vocabulary acquisition. For example WordChamp (an online vocabulary building tool), Lexical Tutor are some of the examples of these applications (Levy, 2009).

Also, digital games are one of the examples of technological applications in language teaching. In foreign language teaching, games are seen as an effective and motivating tool and take attention of students easily (Gaudart, 1999; Muhanna, 2012). Games can be used both to practice the language (simulation games) and to extend the vocabulary of the students (vocabulary games). Vocabulary games can help to better vocabulary learning and to gain more vocabularies when compared to face-to-face classes. Both students and teachers accept games as effective vocabulary learning tools, since games can take attention of students and motivate the students (Yip & Kwan, 2006).

Levy (2009) said that mobile devices can be used to support vocabulary acquisition since with the advances in technology, they functions like computers and they have features like SMS, internet connection which can be useful to support vocabulary acquisition, however their disadvantages like small screens should be considered while designing an application. Games on mobile devices should be considered to both take advantages of games and mobile devices in language learning.

2.4 Digital games for English Vocabulary Learning

In the literature there are different games that designed to support vocabulary acquisition. Hung and Young (2007) classified the English vocabulary games in the literature, which are given in Table 1. The first 6 games were classified by Hung and Young (2007), The WiCFG game in the table is their research and the Bingo Game was researched from Lo & Tseng (2011). As seen from the Table 1,
there are different games in the literature aiming to enhance students’ vocabulary acquisition. Over these games, WiCFG runs on mobile devices.

<table>
<thead>
<tr>
<th>Game</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-word puzzle</td>
<td>Guess the correct word in grids by following the vertical and horizontal hints. Test word ability.</td>
</tr>
<tr>
<td>Word Search</td>
<td>Find out the word from a group of English letters.</td>
</tr>
<tr>
<td>Quiz</td>
<td>An exam to test the word’s meaning by selecting the right answer.</td>
</tr>
<tr>
<td>Hangman</td>
<td>Guess the correct letter will appear the right position of the word; otherwise lose a chance to reconstruct the word.</td>
</tr>
<tr>
<td>Match game</td>
<td>Find the same pair of cards by turning over two cards once in several cards.</td>
</tr>
<tr>
<td>Word Jumble</td>
<td>Put characters in an order into a meaningful word after arranging.</td>
</tr>
<tr>
<td>Bingo Game</td>
<td>The game presents randomly vocabularies and the player should select the corresponding word in the other language from the game board. If the user select the correct word than the cell of the word becomes green otherwise grey. To win the game, the player should form a line from green cells.</td>
</tr>
<tr>
<td>WiCFG</td>
<td>Students create words from letter and they create word maps, Competition based between groups, there is one teacher side and one student side, teacher monitors them.</td>
</tr>
</tbody>
</table>

Table 1 The vocabulary games in the literature

2.5 Cross-situational learning

In real life, there are numerous word-objects mappings in language learning environment and the question is how humans learn the correct referent of a word in such environments. Researches concentrated on how word learning occurs in a single trial, that’s how a word is mapped to the correct referent in a single trial. According to the literature, attentional (Smith, 2000), linguistic (Gleitman, 1990), social (Baldwin, 1993; Tomasello, 2000) and representational (Markman, 1990) constraints governed by user enable them to fast mapping words to the referents in a single trial (Yu & Smith, 2007). However, in real life situations fast mapping of words to the correct referents may not occur in a single trial, since there are too many words and too many potential referents and limited cues about the correct mappings. In such cases, learner may not decide the correct map in a single trial and they
may store mappings from multiple trials and use statistical information from multiple trials to find out the correct word-referent mapping (Yu & Smith, 2007). Cross-situational word learning is proposed as an effective learning strategy for learning pairs of word-referent. In cross-situational learning strategy, a word and its referent are tracked in several trials and the word referent is fixed by using the statistical information gathered from the multiple trials.

The first experimental study was conducted by Yu and Smith (2007) in order to support that human learners exercise cross-situational learning while learning vocabularies. In order to verify this claim they conducted word learning experiments with adult subjects. They researched in this study that word-referent mappings can be acquired by the subjects across multiple exposures to multiple ambiguous contexts where there are multiple objects and words by computing distributional statistics or word and object occurrences. In this study, three experiments were conducted with human participants. At the end of the study, they support the claim that human learners employ cross-situational learning during word learning and cross-situational learning is an efficient learning strategy: this result can be used in the further studies.

Another study conducted by Smith et al. (2009) repeated the experiment of Yu and Smith (2007) and they concluded that cross-situational learning depends on the presentations of the words and their referents and restricted. There might be situations that cross-situational learning may not be efficient and it should be tested.

3 METHODOLOGY

3.1 Description of the Game

In this study, a game based application will be developed and assessed. The game will be a vocabulary game which aims to improve students’ vocabulary acquisition while they learn English as a second language. Vocabulary learning in the application will be based on cross-situational learning (mentioned in section 2.5).

The general structure of the game-based application will be as follows:

- The game will be a matching game: the students will match picture to vocabulary, vocabulary to picture or vocabulary to vocabulary in the game. For example an English vocabulary will be matched with a picture or with its Turkish equivalent.
- The vocabularies to be learned via the application will be selected according to two criteria: the vocabularies that they have learned before and the new vocabularies to them by aiming to enable students to repeat what they have learned before and to learn new vocabularies.
- There will be a master user in the application. The teacher will be able to be defined as a master user or virtual user from the application will be the master user.
- There will be a vocabulary pool in the game which will be formed from lesson curriculum.
- A game in the application will be composed of rounds. In a game there will be predefined number of vocabularies (such as there will be 20 vocabularies for a game) and these vocabularies will be selected from vocabulary pool by the master user.
- In a round a predefined number of vocabularies will be shown to the students (such as 5 vocabularies will be shown in a round). These vocabularies will be selected randomly from the vocabularies that are defined for the specified game and the students will try to match the vocabularies with the corresponding pictures or Turkish equivalent.
The active users in the game will propose their matching. They will see the matching of the other users in the game in a round anonymously (without seeing the owner of the matches). When they see other matches, they will not be able to change their matching if they match the word with its equivalent. In the following rounds, the students can face with the same vocabularies and it is expected that the students will remember the matching of other users in the previous rounds.

- A game will finish when the students reach a predefined percentage of achievement (such as % 75 correct matching) in the specified game.
- The students will be able to play the game individually or in groups.

### 3.2 Development of the game

The development of the game will be conducted by following the stages below:

- Analysis: the requirements for the application will be defined. What the game will do will be identified.
- Design: this stage will cover the design of software, that’s how the requirements defined will be done.
- Implementation: the coding of the application will be conducted.
- Testing: software testing will be performed.

### 3.3 Evaluation of the Game

To evaluate the game, an experiment will be designed and conducted with the real users. Also a survey will be conducted with the users.

#### 3.3.1 Experiment and Survey

There will be two phases in the experiment: Pre-test and post test phases. For the pre-test phase, vocabularies will be selected and the participants will use the game to learn the vocabularies. The vocabularies will be selected from the curriculum. After treatment, a vocabulary quiz and a survey will be conducted. Beside the treatment group, there will be a control group (they will not use the game) and they will take the same quiz with the treatment group. The treatment group will fill the survey related to the game. The survey will include questions about usability, effectiveness, advantages and design issues related to the game. Also there will be open-ended questions in the survey to take the opinions of the students about the game. After pre-test, the game will be maintained according to the survey results then the post-test will be employed. In the post test, the same procedure of the pre-test will be performed.

#### 3.3.2 Participants

The participants of the study will be selected from students who are learning English as a second language. The selection of the participants will be done according to the result of the project (mentioned in section 3.5.). If the project was conducted with the scope of FATİH Project, the participants will be selected from public schools which will be one of the pilot school in the project.

### 3.4 Assessment of the Game

The game will be compared with the other games in the same category (vocabulary acquisition) by considering its effectiveness, usability, contributions to the subject area. The evaluation criteria based on the literature and with expert opinion approval (i.e. two experts in the field of game based education and game based language learning) are formed as below:

1. Trigger motivation
2. Fun and interesting
3. Able to gain new knowledge
4. Facilitating students’ learning
5. Simple and understandable rules
6. Simple interface and user-friendly
7. Different levels of difficulty

4 CONCLUSION
This paper aims to provide an in-depth description of a research-in-progress in mobile game-based language learning. In that regard, the paper presents a detailed research plan including problem statement, review of the literature and methodology. The research is in progress by the time of the EMCIS 2015 Conference submissions and those who are interested can contact the author for the outcomes of the study.

5 REFERENCES


SOCIAL MEDIA USE BY THE UK POLITICAL PARTIES

Viktorija Panfilova, University of Westminster, UK, v.panfilova@my.westminster.ac.uk
Nadia Amin, University of Westminster, UK, aminn@westminster.ac.uk

Abstract

The Web 2.0 elements are seen as a significant facilitator in political communication. In the run up for the UK 2015 General Elections, major political parties are adopting various social media strategies to promote their political campaigns, engage with citizens and attract more voters. The study examined current use of social media by the UK political parties and discussed the impact of online communications during such political campaigns as Barack Obama’s President Election in 2008 and Scottish Referendum in 2014. It was found that UK politicians extensively use such networking sites as Facebook, Twitter and YouTube. However, politicians have not fully utilised social media capabilities and often overlook alternative social media pages. Many political parties still fail to engage in a conversation with their followers subsequently leading to a one-way communication. This report aims to provide the background and set the scene for future research in the area of online communications by political organisations.

Keywords: Social Media, UK political parties, Online activity, Online engagement.

1 INTRODUCTION

In May 2015 the UK will hold the next General Election that will determine which political party will be running the country for the next five years. In the run up for elections, every political party strives to join together with other people who share the same ideas and views; every party aims to win as many votes as possible to secure the seats in the House of Commons and put their promises into actions. Over the last decade governmental organisations have made a considerable effort in utilising the benefits of the Web 2.0 to enhance communication with citizens and promote political campaigns (Charalabidis et al., 2013). More and more people cite social media as their first source for political news (Gurevitch et al., 2009); over 25% of American citizens used internet during 2012 presidential campaign; 15% of German voters refer to social media at least once a week to read the campaign materials and the number of UK citizens signing e-petition doubled over the last seven years (EPRS, 2014). Politicians now more than ever recognise the need for an effective and sustainable social media strategy. Therefore, this report aims to evaluate current use of social media in the UK politics and set the scene for future research in the area.

The following part of the report provides a synopsis of social media channels adoption in political communication. The paper proceeds with studies of political use of three major social networking sites, Facebook, Twitter and YouTube. It provides best and worst practices of online political campaigns in the UK as well as US and evaluates adoption of social media in politics. The paper concludes by outlining future research in this area.

2 OVERVIEW OF SOCIAL MEDIA USE IN POLITICAL COMMUNICATIONS

Over the last decade social networking sites have been dominating digital arena and consequently producing a shift in the way constituents speak to their political representatives. Academics have a
divided opinion in regards to the influence of social media over political environment. According to Leadbeater (2008) social media platforms create online democracy, allowing users to collaborate and become source of ideas for democratic practises. In addition, von Hippel (2005) outlines that social networking challenges the control of media production by state institutions and enables constituents to share alternative opinions and thoughts. Such benefits as user flexibility and availability of information are often cited as advantages that social media bring to political communication (Papacharissi, 2010). In contrast, Leston-Bandeira and Bender (2013) argue that availability of constituents on social media does not mean that political representatives are able to engage in conversations and seize the attention of online users. Similarly, Heinonen (2011) outlines that online users are not as active as it is generally believed and only little number of consumers are involved in online participation and content production. Finally, such factors as marketing budget and availability of resources and knowledge can be a significant predictor in party’s visibility in the social media arena. Hence, making online networking sites a tool for political manipulation. Opinion mining tools, known as sentiment analysis are being employed today by political campaigns to track voters’ interest by monitoring social media messages in real time (Conover et al., 2011). Social media has undoubtedly changed the ways political organisations communicate their objectives and strategies. However, we are yet to know if these online tools are used effectively enough.

3 SOCIAL MEDIA PRACTICES IN THE UK POLITICS

Nowadays the internet and social media are seen as facilitators for the governmental communications with citizens (Barlett et al., 2013). The media is seen as a predominant tool that influences the lives of people. Despite the turnout for the elections falling over the last 20 years, people are interested in communicating with political members and see social media as the number one tool for political updates (Barlett et al., 2013). This trend produced a shift in the ways Members of Parliament (MPs) talk to their constituents (Zittel, 2003). This report analysed the use of six social media sites by the UK political parties and political leaders and presents a snapshot of the current online activity by political organisations as shown in Table 1.

3.1 Facebook

In 2013, 29% of parliamentarians all over the world had a Facebook page. Whilst Canadian and US politicians are leading in digital communications only 35% or European politicians use Facebook to communicate with constituents (Leston-Bandeira and Bender, 2013). All major political parties in the UK have official Facebook page; and all political leaders have a separate official page. The ability to communicate with citizens through Facebook ‘wall’ is considered one of the most valuable networking features of the site (Westling, 2007). All Facebook pages of the UK political parties are consistently updated with visuals and videos. However, as the research shows, almost a quarter of updates are related to parliamentary activity and offer little engagement (Figure 1).

In contrast, during the Scottish Referendum, over 52% of updates by Scottish Parliament were focused on engagement. The “Yes, Scotland!” page (Facebook, 2014a) has 374,863 likes and triggered the creation of many Facebook communities, e.g. “Yes to an independent Scotland” 73,514 likes (Facebook, 2014b), “Vote Yes for Scottish Independence” 7,629 likes (Facebook, 2014c). Overall, the campaign focused on a community engagement and extensively used imagery on social media to grasp the attention of the society. In addition, updates with data and statistics increased online engagement: YouGov update had 10,500+ likes; 1,000+ comments and almost 8,000 shares (Facebook, 2014d). Over 10m Facebook interactions happened during the campaign (Curtis, 2014) and high engagement allowed reaching out to wider Scottish community, resulting in a record-high turnout of 84.5% (Ridge, 2014).
3.2 Twitter

In 2013 73% of MPs had a Twitter account (Heaven, 2013) and the number of Twitter users is increasing amongst political parties and political leaders. Most active page had on average 5 updates per day with at least one retweeted update of a party leader and one update with relevant tags. Most popular updates include visual representation of statistics. For example, ‘More business in the UK’ tweet received 173 retweets and 137 favourites (Cameron, 2014a), and mentions of non-politically related events ‘Formula 1 apprentices at No10’ tweet received 291 retweets, 321 favourites (Cameron, 2014b). Provocative tweets also get high publicity: ‘What EU spend money on’ tweet had 242 retweets and 114 favourites (Farage, 2014). However, to achieve a continuous success with online content promotion, the style of updates should be consistent with the overall online communications strategy (Lilleker and Koc-Michalska, 2013). Poor or untimely communications can lead to devastating effects. Hence, Labour MP was asked to resign after sending a tweet that contradicted the general vision statement of the party. Thornberry’s (2014) tweet was retweeted over 1,000 times and received 533 favourites damaging the image of the Labour party.

In contrast, the President of the US, Barack Obama, has successfully utilised social media strategy during the 2008 elections campaign. Obama’s Twitter activity included constant stream of news, promotion of volunteer opportunities and updates on Obama’s speeches and locations (Greengard, 2009). This allowed Obama receive over 112,000 online supporters, comparing to opponent’s 14,000 followers. Strategic use of social media helped Obama develop relationships with American citizens and win 70% of votes among citizens under the age of 25 (Fraser and Dutta, 2008). Britain’s first Internet election is considered to have happened in 2010. 70% of British public used social media to follow the campaign (Wring and Ward, 2010) and more than 500,000 tweets were sent by political parties throughout the election campaign (Ampofo et al., 2011). However, the Internet did not have a strong effect similar to Obama’s campaign and only created a background to the mainstream media (Wring and Ward, 2010).
3.3 YouTube

The upload of video and audio files in political communication has been on the rise since 2009 (Williamson, 2009). YouTube is now considered a credible source for political news and information (McKinney & Rill, 2009). Currently, over 100 hours of video content is uploaded on YouTube every hour and over 6 billion hours of videos is being watched every month (YouTube, 2014). The research shows that all major political parties have presence on YouTube. Most successful channels are regularly maintained with at least one video upload per week and clear playlist structure. In contrast, least successful channels have not been consistently updated and mainly upload videos during the campaigning time. This results in inconsistent channel structure and a failure to engage with public.

According to Pew Research Center (2008), the web is the primary news source for adults under the age of 30. Thus in an attempt to reach younger audience politicians recognise the importance of video streaming. In the electoral campaign of 2008, Obama’s YouTube channel posted over 800 videos, resulting in over 40 million views. The McCain’s campaign had just over 100 videos with 20 million views (Boynton, 2008). It was found however that the most frequently viewed videos are usually those created by citizens or celebrities. For example, Obama’s video of Chicago night speech has been viewed 4 million times. However, Will. i. am. produced ‘Yes We Can’ video supporting Obama had been viewed 20 million times (Boynton, 2008). In the UK the YouTube was extensively used in the London Mayoral election in 2008, where Boris Johnson and Ken Livingstone used the social media site for online campaigning. Both used YouTube to attract followers and ‘Boris Johnson London Mayor First Interview’ gained 7,385 views, whereas ‘Ken Livingstone on Boris Johnson’ video received 3,335 views (Wheeler, 2008).

It should be considered that YouTube holds a threat for non-official or non-desirable content production (Lim and Golan, 2011). For instance, Casettboy’s video clip about UK’s political leaders had almost 343,000 views (YouTube, 2013), whereas an average number of views for the UK’s Conservative party videos is just over 17,000. Former president Bush’s unpopular comments about riots in Los-Angeles rapidly spread through social media channels and created mass negative responses via the media (Church, 2010). The spread of non-official and non-desirable content creates a threat for politicians that should be considered thoroughly when implementing online strategy (Lim and Golan, 2011).

3.4 Instagram

Instagram has over 300 million registered users and over 30 billion photos were shared so far (Instagram, 2014). With a great opportunity to communicate with a large audience, many businesses are already using Instagram to promote their brand (Instagram, 2014a). However, from the political perspective, the social network has not been fully adopted. Only one political party (Labour) have a running Instagram account for the party and for the leader. The follower to following ratio was positive for both accounts, demonstrating that there is a great potential for communication with audience through the use of imagery.

3.5 LinkedIn

The social network is widely used to communicate with like-minded professionals. LinkedIn is extensively used by David Cameron as the PM of the UK. This includes publications of articles and updates on latest governmental news. Despite being a LinkedIn Influencer, David Cameron, similarly to other politicians, does not have an extensive personal profile. Most of the politicians have little information about themselves and have little amount of connections. In contrast, Nick Clegg had 500+ connections and a detailed profile. Therefore, it could be assumed that the profile is used for internal communications, to ‘connect’ with party representatives.
3.6  **Google+**

Google+ was launched in 2011 and serves as an ‘identification management tool’ (Van Den Beld, 2012). Most importantly, the social network incorporates results into Google search and allows users and their content to be more searchable (Van Den Beld, 2012). The research shows that most of the political leaders and political parties have Google+ account. Most consistently updated profile of the Conservative party had 2 times more followers in comparison to other political parties. However, channels without updates still had a considerable number of followers, therefore suggesting that social network could be a popular source for online communications.

4  **Conclusions and Future Research**

This report examined the use of Web2.0 tools within the political arena. It can be concluded that Social Media has a distinctive role in a political environment both in the UK and overseas. Current research has shown that the existing media strategy for politicians includes extensive use of online websites and communications through such channels as Facebook, Twitter and YouTube. The examination of the social media trinity has revealed that the UK politicians do not engage with public often enough. However, creative updates and the use of visuals help increasing the level of online interactions. Strategically applied social media strategies help politicians support their political vision. As an example, this report outlined Barack Obama’s 2008 presidency campaign (Greengard, 2009). Online transparency allows two-way communication. Therefore, most of the politicians see social media as an essential part of their working lives and are extensively using Twitter and Facebook pages. The research identified that such networking sites as Instagram, LinkedIn and Google+ are currently overlooked by political parties in the UK and are not used extensively enough.

This preliminary study has opened up several agendas for future research. The use of political online communications is on the rise. The role of big data analytics will continue to grow in importance as the recent Obama’s analytics driven campaign has shown (Lampitt, 2013). It will be interesting to see how the analytic database and predictive models will drive the UK political campaigning. Current findings provide groundings to further analysis of the long-term use of social media by political parties. In addition, further research could examine the impact of social media on the UK 2015 General Elections in May and use current findings as a supporting background.
<table>
<thead>
<tr>
<th></th>
<th>Conservative</th>
<th>Labour</th>
<th>Liberal Democrat</th>
<th>UKIP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Leader’s Facebook</strong></td>
<td>FB page: 397,246 page likes; On average 3000+ likes; 400+ comments; 100+ shares.</td>
<td>FB page: 56,695 likes; on average 500+ likes, 200+ comments; 100+ shares.</td>
<td>FB page: 86,886 likes, average 150+ likes; 60+ comments; 20+ shares.</td>
<td>FB page: 56,162 likes; on average 850+ likes; 100+ comments; 75+ shares.</td>
</tr>
<tr>
<td><strong>Leader’s Twitter</strong></td>
<td>Twitter page: 1,287 tweets; 371 following; 838K followers. UK PM account: 7,429 tweets; 2.86M followers, 1,398 favourites.</td>
<td>Twitter page: 4,030 tweets; 1,456 following, 355K followers; consistent updates, high retweet levels.</td>
<td>Twitter page: 1,711 tweets; 8,109 following; 195K followers; consistent updates, updates get up to 30 retweets and favourites.</td>
<td>Twitter page: 6,903 tweets; 1,250 following; 169K followers. On average 80+ retweets, 40+ favourites.</td>
</tr>
<tr>
<td><strong>Leaders’ Instagram</strong></td>
<td>No profile</td>
<td>Instagram page: 24 posts; 1,697 followers; 23 following</td>
<td>No profile</td>
<td>No profile</td>
</tr>
<tr>
<td><strong>Leader’s LinkedIn</strong></td>
<td>LinkedIn page: 1,501,033 followers LinkedIn influencer: 78 posts Constant online activity – 200+ likes; 20+ comments</td>
<td>No official LinkedIn page: set up as a company page, 149 followers.</td>
<td>LinkedIn page: 671 connections. No recent activity. Used as a personal page, rather than a tool for communication.</td>
<td>LinkedIn page: 0 connections, profile not updated.</td>
</tr>
<tr>
<td><strong>Leader’s YouTube</strong></td>
<td>No profile</td>
<td>No profile</td>
<td>No profile</td>
<td>No profile</td>
</tr>
<tr>
<td><strong>Party’s Facebook</strong></td>
<td>FB page: 307,241 likes On average 200+ shares, 230+ comments, 1000+ likes</td>
<td>FB page: 196,961 likes; on average 700+ likes; 200+ comments; 100+ shares.</td>
<td>FB page: 105,479 likes; on average 250+ likes; 70+ comments; 70+ shares.</td>
<td>FB page: 296,912 likes; on average 1000+ likes; 250+ shares; 400+ comments.</td>
</tr>
<tr>
<td><strong>Party’s Twitter</strong></td>
<td>Twitter page: joined 2008; 6,227 tweets; 1,389 following; 127K followers; on average 40+ retweets, 17 favourites.</td>
<td>Twitter page: joined 2008; 10.2K tweets; 15.3K following; 160K followers; Every update gets retweeted and favoured but no more than 100 times.</td>
<td>Twitter page: joined 2007; 9,913 tweets; 4,747 following; 74.8K followers. Consistent updates, a lot of visuals. Low engagement, up to 20 retweets, up to 10 favourites.</td>
<td>Twitter page: joined 2011; 14.4K tweets; 4,940 following; 73.8K followers; Most updated get 50+ likes and retweets.</td>
</tr>
<tr>
<td><strong>Party’s Instagram</strong></td>
<td>Instagram profile: 0 posts; 28 followers; 0 following.</td>
<td>Instagram profile: 218 posts, 855 followers, 388 following.</td>
<td>Instagram profile: 0 posts; 44 followers; 0 following.</td>
<td>No profile</td>
</tr>
<tr>
<td><strong>Party’s YouTube</strong></td>
<td>YouTube page: 8,210 subscribers; 1,377,461 views; 5 playlists; 54 videos.</td>
<td>YouTube channel: 7,860 subscribers; 3,684,286 views; 11 playlists; 100+ videos.</td>
<td>YouTube channel: 3,534 subscribers; 1,268,913 views; 30 playlists; 490 videos.</td>
<td>YouTube channel: 6,410 subscribers; 998,830 views; 1 playlist; 6 videos.</td>
</tr>
<tr>
<td><strong>Party’s LinkedIn</strong></td>
<td>LinkedIn page: 4,106 followers; Page is not currently updated.</td>
<td>LinkedIn page: empty company page, 1,405 followers.</td>
<td>LinkedIn page: page is used to promote jobs; 1,461 followers.</td>
<td>LinkedIn page: page used to share updates; 161 followers.</td>
</tr>
</tbody>
</table>

Table 1. Social Media use by Political parties and Political leaders in the UK. Reviewed in November 2014.
References


Farage, N. 2014. While the eurozone struggles and British taxpayers are asked for more cash for the EU, this is what they spend it on. [Twitter] 1st December. Available from:


*Parliamentary Affairs.* 62(3), 514-527.

*Parliamentary Affairs.* 63(4), 802-817.


[Accessed 27th December 2014].

Zittel, T. 2003. Political representation in the networked society: the Americanisation of European  


An Innovative Decision Making Framework for E-banking Integration

Nikolaos Basias, Department of Digital Systems, University of Piraeus, basias@unipi.gr
Marinos Themistocleous, Department of Digital Systems, University of Piraeus, mthemist@unipi.gr
Vincenzo Morabito, Bocconi School of Management, Bocconi University and SDA, Milan, vincenzo.morabito@unibocconi.it

Abstract

The efforts undertaken by banks to improve their services through Information Technology (IT) have resulted in the development of many incompatible Information Systems (IS) and this has initiated integration problems as disparate IT solutions could not efficiently collaborate. System integration is crucial for e-banking success and involves integrating existing often disparate systems. In addressing integration issues, Service Oriented Architecture (SOA) has emerged and adopted by many banks. Nevertheless, numerous banks fail to achieve successful integration in e-banking for various reasons. One of the main reasons is attributed to the lack of a methodological framework that would: (a) explain factors affecting integration in e-banking and (b) define the SOA adoption process in e-banking. Such a framework might be used by banks as an innovative decision making tool to support the mechanism of effective decision-making. As a result, the aim of this paper is to provide: (a) an innovative decision making framework for integration in e-banking and (b) useful findings. In doing so, this paper extends the body of knowledge and provides new insights related to integration and SOA adoption in e-banking. The research presented in this paper has been partly funded by the Research Center of University of Piraeus.

Keywords: E-banking, Decision Making, Integration, SOA.

1. INTRODUCTION

Banks seek answers to the impact of integration, as it will help them realizing the benefits, the barriers, the risks, the costs and changes that are associated with integration and SOA adoption in e-banking. In this research a factor-oriented approach is introduced to investigate SOA adoption in e-banking and to provide an innovative decision making framework for integration in e-banking. In doing so, the author seeks to: (a) highlight the importance of e-banking, (b) highlight the importance of integration, (c) identify and analyse influential factors of integration and SOA adoption in e-banking, (d) identify and analyse a categorization of factors and (e) identify and analyse different phases of the integration process and SOA adoption.

The first key issue in this research is e-banking. E-banking services have been adopted and used by the majority of banks. Currently, the one third of European uses e-banking with estimates predicting an average adoption rate of 60% in the EU by 2020 (Deutsche Bank Research, 2010). Such figures illustrate that e-banking is an interesting fast growing area. International organizations highlight the
importance of e-banking matters and invest a significant amount of money on research for high-quality independent analysis (Deutsche Bank Research, 2010 & World Bank, 2008).

The second key issue in this paper is Service Oriented Architecture (SOA) that remains a research area of high importance (Luthria et al., 2009, Marks et al., 2008, Eckert et al., 2009, Lee et al., 2010, Basias et al., 2013). Systems integration, channel integration, applications integration (ERP, CRM, and Mobile Apps) and data integration have been major problems for e-banking with banks investing significant amounts of money in overcoming these problems (Themistocleous et al., 2014).

During the last decade, banks have turned to the adoption of SOA to solve their technical problems. SOA paradigm facilitates organizations to automate their business processes and build a maintainable, flexible, secure and manageable Information Technology (IT) infrastructure that is based on global accepted standards such as Web Service Description Language (WSDL), eXtensible Markup Language (XML), Simple Object Access Protocol (SOAP) and Representational State Transfer (Lawler et al., 2008). Service Oriented Architecture supports the automation of business processes using small reusable pieces of code that are called web services (Koumaditis et al., 2009).

Web services can be combined in different ways leading to the implementation of flexible solutions that are easily maintainable. SOA improves efficiency and reduces operating costs by promoting a faster flow of information throughout banks. Furthermore, SOA optimizes business processes and adds corporate business value (Basias et al., 2014, Luthria et al., 2009, Marks, 2008, Eckert et al., 2009). Despite its significant benefits, SOA cannot be considered as a panacea as there are technical and business issues that should be addressed by organizations and banks. For instance, as mentioned before the absence of a methodological innovative framework that would explain influential factors related to SOA adoption in e-banking forms an important research problem that requires deeper investigation.

2. LITERATURE REVIEW

To situate the current study and to provide context related to SOA adoption in e-banking we review the normative literature. The literature review is conducted through an extensive search on books, journals, conference proceedings and working papers published after 2005 using libraries and database search engines like AISel, IEEE Xplore, ProQuest, Science Direct, Google Books and Google Scholar. In a first step we use keywords like: “Integration in e-banking” and “SOA adoption in e-banking” and no matches are found for the criteria specified. We request for related literature to SOA adoption in e-banking via AIS world mailing list without any success.

For that reason, we use keywords like: “e-banking” and “SOA adoption” to investigate this area and to identify and select from a large amount of hits papers and books for deeper investigation and analysis. A review with a procedure outlined by Webster & Watson, (2002) is used in this research to describe, summarize, evaluate and clarify the adoption of SOA in e-banking. In addition, the researcher analyses and compares frameworks related to SOA in order to investigate this issue.

The critical analysis of the literature review indicates that: (a) integration is one of the most important factors for e-banking success and banks continue to face integration problems related to e-banking (Shah et al., 2007, Baskerville et al., 2010), (b) SOA might be a solution for banks to solve integration problems related to e-banking (Baskerville et al., 2010, Rosen et al., 2008), (c) one of the main reasons why banks often fail to achieve the benefits from integration efforts in e-banking is the lack of a methodological framework (Schonewille, 2010), (d) there is no study on analyzing the critical success factors of SOA adoption in e-banking, (e) there is no methodological framework to explain critical success factors affecting SOA adoption in e-banking, (f) the analysis of models and frameworks indicates that several factors and classifications of factors are not similar among researchers, (g) while some adoption issues are relevant for any organization, others depend on the industry and type of organization and some of these differences may be pertinent to the adoption of SOA in e-banking, (h) e-banking has different requirements and business drivers and may follow a different adoption path, (i) the SOA adoption path in e-banking is not clearly defined and there is a
confusion about the SOA adoption content and (j) a methodological framework that explain critical success factors for SOA adoption in e-banking is essential for banks to make right decisions.

3. PROPOSED INNOVATIVE DECISION MAKING FRAMEWORK FOR E-BANKING INTEGRATION

Based on the extensive critical review we build a decision making framework for SOA adoption in e-banking. The conceptual framework illustrates the relationship between e-banking and SOA adoption, providing a research framework for examining factors affecting SOA adoption in e-banking. In addition, it illustrates the stages of SOA adoption in e-banking and a classification of factors. The proposed framework suggests that fifteen factors may influence banks’ decision making process towards the adoption of SOA in e-banking. We classify the fifteen most important influential factors of SOA adoption in e-banking into Performance, Human, Business and Technical factors and propose three stages of SOA adoption.

Performance Factors focus on factors that support organizations to make their decisions regarding the adoption of SOA in e-banking and deal with the assessment of factors like barriers, benefits and costs that have an impact on Organizational performance.

Human Factors refer to those factors that are related to the employees like fatigue, skills, stress and employees’ resistance to change. These factors are considered as important as they can affect employees’ stance towards the implementation of SOA in e-banking and therefore affect the decision making process.

Business Factors deal with at least five business issues that shape the decisions for the adoption of an SOA solution in e-banking. These factors include IT/business alignment, communication, goal, strategy and risk.

Technical Factors are equally important to the aforementioned three factors’ categories as they are vital for the decision making process. Since the new solution will affect the existing IT infrastructure, technical factors related to support (e.g. consultants, vendors), the IT infrastructure and the overall system security should be assessed.

The SOA adoption Lifecycle in e-banking implemented in our framework has three stages: (a) SOA initiation, (b) SOA planning and (c) decision of SOA adoption. SOA initiation: The SOA initiation phase is the first phase in the project. In this stage a business problem or opportunity related to SOA adoption in e-banking is identified and a business case which provides various solution options is defined. SOA planning: After defining the scope of the SOA adoption project the second stage includes a detailed planning phase that involves among others: (a) a project plan, (b) a resource plan, (c) a financial plan, (d) a risk plan and (e) a communications plan. Decision of SOA adoption: The third and last stage refers to the analysis of the previous stages that leads to the decision to adopt or not SOA. The whole proposed innovative decision making framework for SOA adoption in e-banking (factors, taxonomy of factors, stages of SOA adoption) is illustrated bellow (Figure 1).
4. RESEARCH METHODOLOGY

This research was conducted in an interpretive manner and employed a multiple case study strategy (Yin, 2003). Qualitative research was used to study SOA adoption in e-banking in its natural setting and learn from practice. As SOA remains an important research area under investigation (Basias et al., 2012, Luthria et al., 2009, Lee et al., 2010) the qualitative research method seems to be an
appropriate approach to investigate and in depth analyze SOA adoption in e-banking (Miles & Huberman, 1994, Basias et al., 2015). The main focus in qualitative research is to understand, explain, explore, discover and clarify situations, feelings, perceptions, attitudes, values, beliefs and experiences. Factors that influence SOA adoption cannot be separated from its organizational, technical and cultural context and there is therefore a need for a qualitative research approach that allows us to understand the process of SOA adoption and the factors that influence SOA adoption in e-banking.

In this research a multiple case study strategy is used to test the proposed framework since such an approach seems more appropriate for the banking sector (Eckert et al., 2009). A case study offers a ‘holistic’ view of the processes involved, as well as a realization of the topic under research (Yin, 2011). According to Baskerville, Cavallari, Madsen, Heje, Sorrentino & Virili (2010), case studies facilitate multi-perspective analyses that lead to a holistic understanding of cultural systems of action, providing the insight that satisfies exploratory questions. Considerable work on SOA takes the form of case studies to identify SOA success (Lee et al., 2010, Basias et al., 2013). The need for rich empirical data related to SOA adoption in e-banking indicates that the use of a case study approach is appropriate, since it allows examining in depth processes (Miles & Huberman, 1994, Basias et al., 2015). An important aspect of a case study is the use of multiple methods to collect data that leads to obtain rich empirical data for this research.

The proposed decision making framework for e-banking integration is tested in three banks. In particular it is tested in a European Bank, in an American Bank and in a Bank of Southeastern Europe (transition economy). Various data collection methods such as interviews, documentation and observation are used in this research. An interview agenda is developed to ensure desired coverage of the areas of enquiry and comparability of information across respondents. A predefined interview protocol is used to collect the data required for this research.

Interviews were conducted with the banks’ employees who played a key role in an SOA adoption project in e-banking in three banks. In doing so, we interviewed four professionals in each bank including the: (a) SOA adoption project manager, (b) IT manager, (c) a manager from business and (d) a member of the IT strategy board. Interviews lasted for around 40-60 minutes, they were digitally recorded and transcripts were prepared as soon as possible after each individual interview.

Telephone, Skype and e-mail communication was also used to clarify and probe unclear issues that, in some cases, occurred subsequent to transcribing the interviews. In addition to this, the data were crosschecked several times to overcome the contradiction associated with data gathering across multiple sources. Interviews were the main data source that we used to collect data and to capture the verbatim. In most of the cases structured and/or semi-structured interviews mainly took place in the office of the interviewees. Structured interviews were based on the interview agenda designed for this research. Using this agenda, the interviewees replied to specific questions related to SOA adoption in e-banking. The interview agenda consists of sections and each section has multiple questions, to guide us during the structured interviews.

Semi-structured interviews took place without the use of an interview agenda and were conducted during breaks. These involved a series of open-ended questions based on the topic areas we wanted to cover. The open-ended nature of the questions supports them to define the topic under investigation. Moreover, it provides the opportunity for both interviewer and interviewee to discuss the topic in more detail. Using this type of interview we attempted to clarify issues that derived from structured interviews and to collect some important data regarding to SOA adoption in e-banking. In addition, during lunch or coffee breaks, we had the opportunity to collect data through unstructured discussions with employees from the three banks.

Empirical data derived from the three cases related to SOA adoption in e-banking are triangulated and then analyzed to draw empirical conclusions. The purpose of triangulation in qualitative research is to increase the credibility and validity of the results.
5. REVISION OF THE DECISION MAKING FRAMEWORK FOR E-BANKING INTEGRATION

The proposed framework is tested through multiple case studies in three banks that had recently experience in SOA adoption in e-banking. In the first case we collected data from a European Bank, in the second case from an American Bank and in the third case from a Bank of Southeastern Europe (transition economy).

Based on the empirical findings it appears that the factors of our proposed framework are verified. In addition, interviewees from the three banks agreed that they can use the proposed four factors’ categories to better investigate SOA adoption in e-banking. It appears that the proposed factors’ categorization supports the analysis of the factors and helps organizations to enhance their understanding. All interviewees agreed that such a methodological framework would facilitate them to understand the factors affecting SOA adoption in e-banking and to avoid problems. In addition, the SOA adoption lifecycle in e-banking proposed in our framework was confirmed and clarify much of the confusion in this area.

A number of parameters have been extrapolated from the empirical data and identified as factors that were taken into consideration when these banks adopted SOA solutions in e-banking. New factors like: (a) culture, (b) organizational culture, (c) SOA governance, (d) top management support and (e) external pressures were reported by interviewees and confirmed by empirical data.

The analysis indicates that most of the above individual factors are not independent. It is reported that there might be interrelationships among: (a) factors, (b) factors’ categories and (c) factors and factors’ categories. This finding is in line with other adoption studies on integration technologies that support that the one factor can impact the other.

Based on the findings of this research we alter our framework and we present the revised one in Figure 2. We added the five new identified factors namely: Culture, Top Management Support, SOA Governance, External Pressures and Organizational Culture to the relevant categories and we depict them using a circle. The revised innovative decision making framework for integration illustrates also possible interrelations among the factors’ categories.
Figure 2: Revised Decision Making Framework for E-banking Integration
6. MAIN FINDINGS

The main findings derived from the work presented in this paper are presented below:

Finding 1: The Significance of Integration in E-banking

This research highlights the importance of integration in e-banking. The importance of integration is mentioned in the literature and verified by empirical data. As mentioned by all interviewees integration is one of the most important factors for the success of e-banking.

Finding 2: The Need for a Decision Making Framework for Integration and SOA Adoption in E-banking

This research indicates that one of the main reasons for unsuccessful integration and the inability of banks to realize the benefits of SOA adoption in e-banking is the lack of a decision making framework. The lack of a relevant decision making framework for SOA adoption in e-banking and the need for such a framework is verified by empirical data.

Finding 3: Identification of Influential Factors of Integration and SOA Adoption in E-banking

As verified by this research influential factors are significant for decisions related to integration and SOA adoption in e-banking. The necessity and importance of the proposed most important factors is verified by empirical data. As underlined by all interviewees’ non-examination of important integration factors might lead to significant problems related to SOA adoption in e-banking and the failure of the overall SOA adoption project.

This research identifies and analyses for the first time the most important factors that influence the decision making process for integration and SOA adoption in e-banking. The most important factors are: (a) barriers, (b) benefits, (c) cost, (d) fatigue, (e) skills, (f) stress, (g) resistance to change, (h) IT/Business Alignment, (i) communication, (j) goal, (k) risk, (l) strategy, (m) IT infrastructure, (n) security, (o) support, (p) culture, (q) SOA governance, (r) top management support, (s) external pressures and (t) organizational culture. The empirical data and the analysis indicate that the above factors are the most important factors for successful SOA adoption in e-banking. For that reason these factors should be examined carefully by banks.

Finding 4: Classification of Influential Factors of Integration and SOA Adoption in E-banking

This research highlights the need for a suitable taxonomy of influential factors of Integration and SOA adoption in e-banking for a better management and presentation of SOA projects. In addition, this dissertation identifies and tests four appropriate categories for the most important factors that influence the adoption process in e-banking. We classify the most important influential factors of SOA adoption in e-banking into human, business, technical and performance factors. The importance and relevance of the proposed classification is indicated by all interviewees. This taxonomy has a positive impact on SOA adoption projects in banks. In addition, a classification of barriers and benefits in financial, operational, organizational and technical barriers / benefits and cost in direct, indirect human and indirect organizational cost have a positive effect on SOA adoption in e-banking.

Finding 5: Integration and SOA Adoption Lifecycle in E-banking

For the first time this research identifies three phases of SOA adoption in e-banking. The importance and relevance of the proposed three phases are adopted by the three banks under examination and
verified by empirical data. The phases of SOA adoption in e-banking are: (a) initiation of SOA, (b) planning of SOA and (c) decision of SOA adoption.

Finding 6: Hierarchy of Influential Factors of SOA Adoption

This research presents a ranking of the most important factors for successful SOA adoption in e-banking. The empirical data indicates some slight variations in the hierarchy of factors among Banks. The European Bank and the American Bank highlight as the most important factor for SOA adoption in e-banking security while the Bank of Southeastern Europe highlights cost. Security appears at the third place at the Bank of Southeastern Europe while cost appears at the second place at the European Bank and at the fourth place at the American Bank.

These small variations among banks are expected and empirical data indicates that these differences are based among others on: (a) size, (b) the external environment, (c) culture, (d) experience, (e) the adoption level of technologies and (f) the financial situation of the banks. However, after analyzing reach empirical data from three banks this paper presents a general hierarchy for influential factors for Integration & SOA adoption in e-banking. The five most important factors for successful SOA adoption in e-banking are: (a) security, (b) cost, (c) benefits, (d) strategy and (e) barriers.

Finding 7: Interrelationship of Influential Factors of Integration

The empirical data indicates that influential factors should not be considered as independent entities. These factors are components of the methodological decision making framework for integration in e-banking and should be considered in a holistic approach. Empirical data suggest that most of the factors are not independent and affect and/or are affected by other factors from the same and/or other category of factors. For instance, cost is associated with benefits and barriers while resistance to change is connected to culture. Stress and fatigue affect resistance to change and strategy is associated with most of the factors.

Finding 8: Synthesis of an Innovative Decision Making Framework for E-banking Integration

This research develops, tests, revises and presents for the first time an innovative methodological decision making framework for integration in e-banking. The new proposed methodological framework introduces innovative elements in three levels. The first level identifies the most important factors that should be considered by banks in the decision making process of integration and SOA adoption in e-banking. The second level introduces a new classification of important factors of integration and SOA adoption in e-banking and the third level presents a SOA adoption lifecycle with three phases. The phases for SOA adoption in e-banking are: (a) initiation of SOA, (b) planning of SOA and (c) decision of SOA adoption.

The author points out that the three main components of the methodological framework and their interrelationships should be examined by third parties with a holistic approach and therefore the elements are incorporated into an innovative decision making framework for integration in e-banking. As indicated by the empirical data and all the interviewees from the three banks under examination the proposed methodological framework greatly facilitate the SOA adoption process in e-banking. In addition: (a) it helps banks to make robust and fast decisions about integration and SOA adoption, (b) it adds a significant new dimension to existing integration and SOA frameworks examining for the first time SOA adoption in e-banking and (c) it enriches the academic literature.
7. REFERENCES


E-Government

Chairs: Dr. Karim Al-Yafi
Table of Contents

Track – E-Government


2. [PID-42] Ryma Abassi and Sihem Guemara El Fatmi, “PRIVACY PRESERVATION IN A PATTERN-AWARE AUTHENTICATION SCHEME FOR VEHICULAR AD HOC NETWORKS”


POLICY IMPACT EVALUATION THROUGH PROSPERITY METRICS AND OPEN DATA SOURCES

Panagiotis Kokkinakos, National Technical University of Athens, Greece, pkokkinakos@epu.ntua.gr
Costas Koutras, National Technical University of Athens, Greece, ckoutras@epu.ntua.gr
Ourania Markaki, National Technical University of Athens, Greece, omarkaki@epu.ntua.gr,
Sotirios Koussouris, National Technical University of Athens, Greece, skous@epu.ntua.gr
John Psarras, National Technical University of Athens, Greece, john@epu.ntua.gr

Abstract

The purpose of the paper at hand is to study and provide an overview of (the theory and practice of) prosperity indicators and the data sources associated to their calculation. The document also aims to summarise the basics of social indicators, their basic methodological principles and their typology, as they are considered a main tool for efficiently and effectively assessing the impact of governmental policies. The basics of Open Data and their exploitation are also reported. As the work described in the paper at hand is highly connected to the Policy Compass FP7 project, a separate section is devoted to the origins and the current interests of social indicators exploitation at the two project’s pilot sites (Cambridgeshire (UK) and St. Petersburg (Russian Federation)). Finally, a number of important indicators’ initiatives is analysed putting particular focus on the sources for open data associated to their calculation. A set of interesting conclusions is provided, describing also the projected added value that accrues from an innovative ICT-based approach put forward by the Policy Compass project.

Keywords: Policy Evaluation, Prosperity Metrics, Open Data.

1 INTRODUCTION

Prosperity indicators, following the view that “Indicators and performance measures have become an important element in policy initiatives relating to sustainability and to the re-invention of government” (Innes & Booher, 2000), are nowadays considered one of the main instruments towards measuring and assessing the actual impact that policies and policy measures have in the areas they target but also in society in general. The term ‘indicator’ is regularly conceived as a sort of ‘statistical measure’ that can adequately capture crucial aspects of a phenomenon that should be monitored, in particular when a specific policy measure is enforced to affect it. Perhaps the simplest and most general definition is that of (Innes, 1990): an indicator is “a set of rules for gathering and organizing data so they can be assigned meaning”, while a number of additional indicators also exist (Carlisle, 1972), (Bauer, 1966).

The idea of employing quantitative indicators in order to evaluate policy implementation goes back to the ‘40s, when the US economy was being evaluated in terms of the Monthly Economic Indicators (Wong, 2006). Subsequently, this led to an explosion of indicators for social change in the ‘60s, when actually the term ‘social indicators’ was coined with great clarity and its definition is one of the most readable ones, even today: “Social indicators – statistics, statistical series, and all other forms of evidence that enable us to assess where we stand and are going with respect to our values and goals, and to evaluate specific programs and determine their impact.” (Wong, 2006). As is to be expected with most scientific and methodological breakthroughs, the initial explosion of interest and enthusiasm on ‘social indicators’ was followed by a wave of scepticism and partial disappointment on the actual effect and the scope of the method. However, the early ‘90s witnessed a noisy comeback of the ‘indicators methodology’, through the phenomenon called the ‘community indicators movement’ (Innes & Booher, 2000).
The new wave of interest has been significantly motivated by the global questions on environmental matters and has led to a series of approaches, typically associated to the keywords ‘indicators for quality of life’, ‘sustainability indicators’, sometimes combined with other widely used terms in public discourse, such as ‘economic competitiveness’, etc. (Sawicki, 2002).

Along the above lines, the paper at hand aims to study and provide an overview of prosperity indicators and the data sources associated to their calculation, touching also on the broad theme of Open Data which open new horizons for the calculation and exploitation of Social Indicators. The following section deals with open data and how these are utilised for the calculation of prosperity indicators. Section 3 aims to present a quick overview of the main methodological concerns regarding prosperity indicators structuring and formalisation while section 4 presents the utilisation of indicators for policy making. Section 5 presents in brief the Policy Compass FP7 project approach (and the respective envisioned pilot applications), while section 6 concludes the document.

2 OPEN DATA AND PROSPERITY INDICATORS

2.1 Definition, Benefits and Applicability of Open Data

According to the definition given on Wikipedia, “open data is the idea that certain data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control”\(^1\). The open data movement shares the same concept with the “open” movements, such as open source or open access. According to the Open Data Institute, “Open standards are standards that are developed through a fair, transparent, collaborative process, available under a royalty-free license”\(^2\) and highlights the importance of the quality of such data by listing the following important attributes:

- Easily linkable, thus allowing easy sharing and discussions over the data.
- Standardised and of structured format, in order to facilitate processing.
- Characterised by guaranteed availability, so that others can rely on it.
- Easily traceable, so that others can trust it.

One of the most significant types of open data is government data. Government data is considered as public by law and should be available to anyone. Furthermore, observations and results of scientific activities is another common type of open data, allowing scientific data to be available to anyone in order to analyse and reuse. Private companies have also started opening their data, in an effort to revolutionise the way they compete. The availability of open data for public use is of great value and growing importance. Especially at the moment of the global economic crisis, opening data can be of great importance to society, the government and the economy in general (GovTech, 2014).

There is no doubt that governments are more transparent and accountable by opening their data, providing information to everyone about where government money is spent and what the government is doing. Nevertheless, transparency does not directly imply accountability (Robinson & Yu, 2012).

Opening-up data, also helps democracy. It allows citizens of the democracy control government and be able to significantly reduce corruption. According to a relevant analysis there is a strong relationship between democracy and transparency, between democracy and open data (CDC, 2014). Depending on the nature and the availability of open data, there are different ways of possible applications; open data portals with visualisation and sharing capabilities, as well as unlimited web and mobile applications and services on different subjects.

2.2 Open Data in Defining Indicators

One of the major concerns in the construction and exploitation of indicators has been the access to the relevant data and the difficulties in the collection and reliability of the data needed in order to calculate and interpret social metrics. This was one of the ‘pitfalls’ identified in the scepticism of the ‘70s. The revolution of the World Wide Web and the Open Data Movement, conceived as “the idea that certain data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control”\(^1\) arguably opens a new arena of experimentation with social indicators.

---

1 http://en.wikipedia.org/wiki/Open_data
2 http://theodi.org/guides/what-open-data
We do expect it will be possible to define and exploit indicator sets, with the ultimate goal of influencing the policy-making process at the local (and perhaps the global) level. Many methodological questions abound and some of them are reviewed below in the document at hand.

3 ON THE METHODOLOGY OF DEFINING SOCIAL INDICATORS

With the advent of evidence-based policy and the accountability concerns of the recent years, it is not hard to observe an obsession with the heavy exploitation of indicators and all sorts of statistical measures in the public policy arena. As stated in (Jany-Catrice & Marlier, 2013), “These initiatives have to be viewed with a certain degree of circumspection. To what extent does the infatuation with indicators have its roots in a fashion for quantification, a form of ‘quantophrenia’? With notions as fuzzy as ‘sustainable development’, ‘quality of life’ or ‘well-being’, the indicators themselves eventually come to embody the concept. This is not a recent phenomenon. Historically, the discipline of economics has conceptualised and defined wealth and progress in a particular way, and in conjunction with establishing the instruments of measurement. This approach has tended to develop in societies in which quantified arguments and, more specifically, ‘numbers’ frequently take on all the trappings of an incontestable argument.”

Care should be taken in order to avoid turning policy-related discussions in a meaningless collection of statistics or number series, with questionable usability and unidentified direction. There exists a widespread distrust (see also the passage above) that all these attempts to measure very abstract concepts (quality of life, deprivation, welfare, environmental quality, etc.) are not always well defined and the measurement reflected is rather rarely supported by methodologically sound techniques and well-defined policy related frameworks (Innes & Booher, 2000).

The need for a solid formal methodological basis for indicator development appears to be self-evident. Several suggestions appear in recent literature, each one employing a discrete number of steps, from the early conception of the idea to the final description of the conceived index. During the ‘80s several approaches have been proposed; we have opted for a brief presentation of the four-step methodology from (Coombes & Wong 1994), for its simplicity and flexibility. This approach has undergone a few improvements deriving from the experience acquired in its exploitation.

The description below draws directly from (Wong, 2006). The steps of the methodology comprise:

- **Step 1 – Conceptual Consolidation:** Clarifying the basic concept to be represented by the analysis.
- **Step 2 – Analytical Structuring:** Providing an analytical framework within which indicators will be collated and analysed.
- **Step 3 – Identification of Indicators:** Translation of key factors identified in Step 2 into specific measurable indicators.
- **Step 4 – Synthesis of Indicators’ Values:** Synthesizing the identified indicators into composite index/indices or into analytical summary

On the other hand, from a purely non-technical, ‘political’ standpoint, many trends can be identified in indicator design, in particular with respect to their intended exploitation (Innes & Booher, 2000):

- Attempts to construct a single, usually composite, indicator in order to capture a quality-of-life dimension. There exist quite ‘successful’ indicators in this trend, such as the GDP\(^3\) or the GPI (Genuine Progress Indicator)\(^4\) from ‘Redefining Progress’ (Innes & Booher, 2000).
- Using multiple, separate, indicators for social problems, with the aim of capturing single important aspects of everyday life, such as crime rate, poverty level, air pollution, unemployment rate, etc.
- Preparing all-inclusive indicator reports, usually in significant numbers, intended for wide distribution and produced by experts or by groups of citizens in a collaborative fashion. The underlying idea is that decision-makers and policy-analysts will consult the reports in an iterative

---


\(^4\) [http://genuineprogress.net/genuine-progress-indicator/](http://genuineprogress.net/genuine-progress-indicator/)
fashion, distilling the information related to their area of expertise. In Innes & Booher (Innes & Booher, 2000), the European Common Indicator initiative is considered to be representative of this direction.

- The trend of ‘re-inventing government’ (Osborne & Gaebler, 1992), which “emphasises the development of measures of performance and customer satisfaction with government. This spreading re-invention movement differs from the other views of indicators in that it emphasises dialogue in the design and use of measures and in the interpretation of these in a complex, changing context. It does not advocate the use of indicators to guide a top down control system, but rather to facilitate the work of many players to make better choices, solve problems, and be better able to respond to context and change.” (Innes & Booher, 2000).

3.1 Is there a Way to Concretely Define a Good Indicator?

Another important question that comes up to mind is ‘Which is the characteristic profile of a useful indicator’? As also mentioned earlier in the document at hand, an indicator is a quantitative tool that provides an indication, used to give an order of magnitude to a given condition. Indicators provide information on the state of a particular social ‘system’, and are particularly useful to highlight trends that can assist on identifying causal relations among the components of the system. A combination of different indicators (a composite indicator) is sometimes useful for the description of complex social phenomena.

According to OECD, a well-defined and useful indicator should comprise (UNEP, 2014):

- Policy Relevance: the indicator needs to address issues that are of (actual or potential) public concern relevant to policymaking. In fact, the ultimate test of any single indicator’s relevance is whether it contributes to the policy process.
- Analytical Soundness: ensuring that the indicator is based on the best available science is a key feature to ensure that the indicator can be trusted.
- Measurability: the need to reflect reality on a timely and accurate basis, and be measurable at a reasonable cost, balancing the long-term nature of some environmental, economic and social effects and the cyclicity of others. Definitions and data need to allow meaningful comparison both across time and countries or regions.

4 INDICATORS AND POLICY MAKING

4.1 The Causal Component

Having been able to define indicators and having access to the data sources needed to calculate them in a consistent way, it is important to look back again at the fundamental problem of their usability. How are these indicators to be exploited? How are they to be linked to the decision-making process?

This seems to be a hard question, after all. Indicators should be integrated somehow in the policy lifecycle and this is the difficult part. As we exhibit below, it has been achieved in an integrated way in some application areas, e.g. sustainability and environmental planning. The key aspect in the idea behind turning indicator development into a decision-making process, is the causal dimension of indicator usage which is to be triggered through the usage of a causal model.

At the heart of all policy design and implementation is a need to understand why policies should be introduced and how well they are working. Various frameworks assist in identifying problems, assessing impacts and developing understanding of the underlying causes and effects. The difficulties of establishing a ‘cause-effect’ relationship in social phenomena are very obvious (Wong, 2006). At a different scale, a macroscopic one, the UNEP Working Paper comments on the importance of analysing causal relationships and examining indicators in an aggregated fashion (UNEP, 2014).
For the needs of the Policy Compass FP7 project\(^5\), Fuzzy Cognitive Maps (FCM), that apply fuzzy logic to reason about the behaviour of a system using a network of interconnected nodes, are the selected tool for the causal modelling and analysis.

### 4.1.1 FCMs: Advantages and Application to Policy Modelling

A cognitive map (CM) is a representation of causal relationships among the elements of a given object or problem. It describes an expert’s perceptions about a subjective world that does not necessarily reflect an objective reality. It is composed of:

- Concept nodes that represent the factors describing a target problem;
- Arrows that indicate the causal relationships between two concept nodes and finally
- Causality factors (coefficients) on each arrow indicating a positive (or negative) strength with which a node affects another node.

The causality value can be positive (‘+’) or negative (‘-’). The causality coefficient can be transformed into a real value between −1 and +1. Axelrod (1976) states that a simple CM with a causality coefficient ‘+’ and ‘-’ is sufficient for replicating human cognition because decision-makers typically do not use a more complicated set of relationships.

In general, FCM (introduced by (Kosko, 1986)) is a fuzzy extension of the CM that strongly resembles neural networks and they are used to compute the strength of impact of these concepts through forward and backward chaining. Till today there are over a hundred research papers which deal with FCMs, and the tool has been used to study real-world situations as varied as stock-investment analysis to supervisory system control, and child labour to community mobilisation against the AIDS epidemic (Kandasamy and Florentin, 2003).

The application of FCM to policy modelling and analysis has a long history. As a matter of fact, FCM have been applied to policy modelling and analysis in politics and public administration. In (Mendoza and Prabhu, 2006) for example, FCM was applied for participatory modelling and analysis for sustainable forest management. FCM also has been applied to public participation for Water Framework Directive in Greece (Mouratiadou and Moran, 2007).

FCMs do not only represent relationships between variables but also their strength as fuzzy value (hazy degrees of causality). Given that these fuzzy degrees of causality can be expressed either quantitatively or qualitatively (Kosko, 1986) and that FCM are not limited by the problem identified with model identification as seen in classical statistical methods (Craiger and Coovert, 1994), FCMs are useful for modelling systems that cannot be explained entirely mathematically or that need to represent both qualitative and quantitative information or model both tangible and intangible issues (Stylios and Groumpos, 1999). Thus FCMs are able to represent domain or process or problems that can be considered to be complex, vague and even incalculable (Khan et al. 2001). Therefore, FCM allow for the comprehensive modelling of qualitative knowledge governing strategic decision-making, together with the ability to provide a holistic picture of the problem or domain (Irani et al., 2002) FCMs can also allow systematic propagation, in particular forward and backward chaining because of their graph structure (Lee and Han, 2000). This feature gives the ability to simulate policy model through forward and backward chaining to identify the policy impact, which makes it ideal for the Policy Compass platform.

Another characteristic that makes FCMs more attractive to Policy Compass is its ability to encapsulate both quantitative and qualitative aspects of a complex domain or process and overcome limitations of traditional approaches. It allows capturing the dynamic aspect of system behaviour. It can also allow analysts explore different alternatives. For example (Tsadiras et al., 2001) simulate a policy model using FCM for alternative political scenarios, resulting in respective predictions. The exploration of alternative scenarios provides analysts with possibility to see the impact of policy changes on different concepts (policy factors). The analysis leads to identifying the crucial factors of proposed policy model and this made FCM one of the major tools for resource allocation. Given that FCMs allow for

---

\(^5\) [http://policycompass.eu/](http://policycompass.eu/)
exploration of different possibilities and have been considered and used successfully as discussed above, it is natural to adopt FCMs as basic policy modelling formalism in Policy Compass. Furthermore, FCMs are easy to use. Policy makers or citizens can use diagrams to represent cause-effect relationships among major concepts instead of text rules. There is a well-defined procedure for capturing domain knowledge, analyse validity of models, and simulate impacts of policy changes.

In summary, the motivations for using FCM approach on Policy Compass platform are the following:

- Easy to use and parameterize;
- Easy to build an abstract of a policy model including variable that needs analysing;
- Able to model casual relations that are not known;
- Easily understandable/transparent to non-experts and lay people;
- Able to model complex casual relation with feedback system;
- Easily and quickly combining different knowledge sources and running different policy options for policy impact analysis;
- FCM can be used as rich body of knowledge by combining views of experts or stakeholder from different information sources banding them in structural/understandable form.
- FCM dynamic system is capable of capturing the dynamic aspect of system behaviour.

The main advantage of this method based on the literature review is its simplicity. The Policy Compass platform brings policy makers and citizens to share their opinion to build a complex policy model very quickly and easily with the use of FCMs. This also provides an opportunity for policy makers and citizens to link different data sources into decision making to run different scenarios to determine perceived outcomes of proposed policies and share them with others.

### 4.2 Some Interesting Cases

In this section, we demonstrate some interesting cases of indicators that may provide an inspiration on how to work when confronted with the measurement of a specific social phenomenon. We have chosen to cope with aggregate indicators, in particular with interesting examples at the level of EU, the level of region or the level of a large city.

#### 4.2.1 European 2020 Headline Indicators

As a large-scale example described in the ‘big picture’ let us examine the methodology for monitoring the EU progress towards the goals set by the Europe 2020 strategy, adopted by the European Council on 17 June 2010. Thus is the EU's common agenda for the current decade. It pays special emphasis on ‘smart, sustainable and inclusive growth, aiming to remedy the structural weaknesses in Europe's economy, improving competitiveness and productivity.

The key objectives of the strategy are expressed in the form of five headline targets at the EU level:

- 75% of the population aged 20-64 to be employed;
- 3% of the EU's GDP to be invested in research and development (R & D);
- Climate change and energy targets:
  - To reduce greenhouse gas emissions by 20% compared with 1990;
  - To increase the share of renewable energy sources in final energy consumption to 20%;
  - To improve energy efficiency by 20%;
  - The share of early school leavers to be under 10 % and at least 40% of those aged 30-34 to have completed tertiary or equivalent education;
- Reduction of poverty by lifting at least 20 million people out of the risk of poverty or social exclusion.

#### 4.2.2 The EU Sustainable Development Indicators

Our second example is again drawn from the large-scale and is again about headline indicators: Sustainable development as an objective of the European Union.

The Sustainable Development Indicators (SDIs) are used to monitor the EU Sustainable Development Strategy (EU SDS). There exist ten themes and twelve headline indicators (see below) out of more
than 100 indicators. They are intended to give an overall picture of whether the European Union has achieved progress towards sustainable development in terms of the objectives and targets defined in the strategy. More accurate information is obtained if we look at all indicators within a theme. Sustainable development is conceived as ‘a better quality of life for both present and future generations’.

The major challenges are:

- Climate change and clean energy;
- Sustainable transport;
- Sustainable consumption and production;
- Conservation and management of natural resources;
- Public health;
- Social inclusion, demography and migration;
- Global poverty and sustainable development challenges.

The set of sustainable development indicators has been developed by the European Commission, in cooperation with Member States, EFTA and candidate countries. Eurostat produces a monitoring report based on this indicator set every two years.

4.2.3 A striking example: Deprivation Indices at the UK

It seems that one of the ‘oldest’, more mature and successful policy indicators examined in the literature, are the UK deprivation indices, a prime example of important application of social statistics. On the web many details of the indices are available. What is interesting to see here is the applicability of the method, which is apparent in the following picture and clearly shows the scope of this particular set of social indicators.

5 THE POLICY COMPASS APPROACH

As reported in the previous sections of the document at hand, it is widely recognised in the literature that the exploitation of quantitative techniques may assist in the attempt to pin down some of policy making facets and evaluate the results of planned or enforced policy measures. At the same time, most sources claim that the more the definition of such an indicator follows a standardised, highly structured and expert involving process, the more rigorous it should be, and as this process is lengthy and time consuming, it is highly improbable to arrive to accurate indicators by following other paths.

However, the latest advances in IT and open government have cultivated a highly encouraging environment towards the opposite direction: the advent of the web-based open data sets provides a fantastic observatory of experimentation in the calculation and exploitation of social indicators for policy evaluation and the strengthening of the democratic process in today’s societies. The power of the crowd, and the wisdom of social groups may prove to be more inclusive than this of certain experts, and a highly flexible, iterative and participatory process in defining metrics of prosperity could prove to provide better and closer to reality figures that could measure quality of life in a more understandable and realistic manner, not only focusing on pure economic data.

The Policy Compass project, intends to contribute in this research, along several axes, including:

- Exploring the limits of social computing with quantitative indicators for policy design and assessment, given access to meaningful data provided by the open data sources available.
- Enhancing the experimentation with various kinds of social indicators, ranging from the well-known and widely established metrics around the GDP and its variants, to the composite and headline indicators which can suitably apply to the regional or municipality level.

---

6 http://epp.eurostat.ec.europa.eu/portal/page/portal/sdi/indicators
• Experimenting on the cross-fertilisation of today’s ICT capabilities, with the intuition of the social motivation of describing societal welfare with well-defined, representative metrics.

• Further on the previous, import the cause/effect component of policy analysis directly into the indicator analysis process through the ‘injection’ of causal analysis tools.

The Policy Compass project aims to design and implement a user-friendly tool that will let the various interested stakeholders (e.g. citizens, policy makers, NGOs etc.) assess the actual impact of specific policies in local, regional, national and international level. Policy Compass will encompass various different functionalities including:

• Fuzzy Cognitive Maps (FCMs);
• Argumentation support;
• Visualisation of datasets either from various trusted sources or from the users themselves;
• Annotation of existing visualisations with specific events via an events’ registry etc.

5.1 Policy Compass in Cambridgeshire

In the UK the Skills Funding Agency (SFA) provides nationally a set of indicators for adult learning. The data for the adult skills indicators is sourced from the Annual Population Survey, which is a sample of data at local authority level. The Skills Funding Agency data provides education providers with performance indicators on current participation and achievement. This data is available at Local Authority level, by age, sector, training type and level. It provides more accurate information on provision and resident participation but not all data captured from participants is open data.

Adult Learning & Skills’ performance is judged against a number of Key Performance Indicators (KPIs) that are set out in the contracts awarded every year. Current KPIs include:

• Total number of starts (enrolments);
• Total number of withdrawals;
• % of timely achievements;
• % of learners achieving a job outcome;
• Learner satisfaction.

The Cambridgeshire County Council will use Policy Compass as a policy modelling tool for the Cambridgeshire Adult Learning Fund (CALF) budget. Policy modelling will allow the services to plan the provision and budgets of the fund. By using open data trends in learners and learning outcomes can be identified as well as areas where more work could be focussed. CCC will be using this tool to help demonstrate the impact of the CALF which has been available for some years and through which a number of learning programmes and projects have been run.

The two user groups (experts and community groups) will be made up of those individuals directly involved in Community Learning. The CALF Panel will be the expert group. The Panel are responsible for allocating the CALF budget to training providers and deciding which projects get funded. The Community group will be made up of members of Learner Advisory Panel (LAPs). LAPs are a new development for the Service and are designed to engage service users directly with the Service, giving feedback and helping plan provision at a local level.

An initial FCM designed for the corresponding policy model can be found in Figure 1. The FCM depicts the (positive) interconnections among the various (alternative) proposals to the proposed indicators and the interconnections among the latter:
5.2 Policy Compass in St. Petersburg

The Leningrad Region of the Russian Federation is one of two field trials for the Policy Compass project. The region’s Administration joined the project as an associate partner together with the e-Governance Center of the ITMO University – the only Russian project partners. The Leningrad region was chosen because of its strong interest and commitment to the project’s goals, as well as its long standing and successful co-operation with the ITMO University in the area of e-Government and Information Society development.

The focus of the aforementioned trial is the implementation of the Regional program “Development of the Information Society in Leningrad region in 2014-2018” which is the successor of a series of federal and regional programs devoted to the promotion of e-governance in various fields of public administration in the region in 2002-2013. The recognition of the importance of these programs for both the economy and society of the region is reflected in the level of interest and high importance given to this topic. There is a real need for harmonization of program goals, activities and indicators with the urgent and long-term interests of citizens and businesses. Among the decisive factors to include the Leningrad region as a pilot in the project were the presence of a rich history of related normative documents as well as precise socio-economic data which reflected the result and impacts of previous projects.

Despite the fact that in recent years a lot of Internet-based instruments for bilateral G2G, G2B and G2C interaction tools, and jointly-used IT systems and data warehouses have been implemented, their effective use has evolved slowly and their impacts are still insignificant.

Until recently, the lack of available information about the results, and user-friendly tools for analysing and forecasting the results of such Regional programs led to a loss of public interest in them. Under these circumstances, the Administration preferred the use of “comfortable” indicators that helped the preparation of positive reports; however, little actual benefit was brought to local citizens and business.

With the rapid development of open data and state-of-the-art tools to work with, this situation changed radically and the Policy Compass methodology and tools will help overcome problems such as:
The sometimes passive nature of citizens’ engagement with politics and policy making due to the lack of open socio-economic data and tools to analyze it;

Lack of civil servants motivation to meet the real needs of local citizens and business;

Lack of analytical support for policy-making processes.

Policy Compass offers the Leningrad Region’s citizens and policy-makers a novel and user friendly platform for analysing various factual indicators and the ability to compare them against the planned ones both in their region, as well as in other regions within the Russian Federation and around the world.

The Committee on Information and Telecommunications Administration of the Leningrad Region is the most interested and involved department in the trial structure. In the context of the pilot implementation, the Committee with the support of the e-Governance Center, will offer metrics and data sets, propose schemes of their interrelations in the form of fuzzy cognitive maps (FCMs) and try to use the Policy Compass tools to visualize, analyse, explain and improve expected outcomes of the previously mentioned Regional program.

To begin with, the simplified model will be used to analyse the relationship among aggregate indicators such as quality of life, quality of public services and functions, performance measures for the development of the information society etc. Furthermore, the gradually improved and wider set of tools for open government data analysis, results’ visualization and their collective discussion will provide appropriate conditions to involve a critical mass of citizens and policy makers in the processes of the open policy development. This should have a positive impact on the Program’s goals setting and prioritization, formation of optimal project portfolio and the selection of appropriate performance indicators. As a consequence, a radical increase in efficiency and effectiveness of the implementation of Regional programs and projects, as well as improving quality of life in the region are expected.

6 CONCLUSIONS AND NEXT STEPS

In the context of the paper at hand, the authors worked towards providing a review of an important and fast-growing field – the design and interpretation of social indicators. The reader of this paper has encountered the basic fact that Social Indicators attempt to measure social phenomena such as the ‘quality of life’ and ‘well-being’, which are hard to pin down by a set of metrics.

The Policy Compass project, intends to enhance the experimentation with various kinds of social indicators, exploiting contemporary ICT capabilities, with the ideas and intuition of the social (and political/economic) motivation of describing societal welfare with well-defined, representative metrics. The import of the cause/effect component of policy analysis is also among the main targets of the Policy Compass platform and methodology.

In the interests of completeness however, we should summarise here the basic questions identified in the context of this paper. Experimenting with indicators seems completely feasible, technically speaking. Is it scientifically legitimate? In the literature, one can find strong statements in favour of the position that it makes little sense to proceed in proposing alternatives to single, established, widely accepted and mature indicators, such as GDP. It seems very ambitious to try to replace such a mature indicator.

However, the deep questions on ‘what is a good society’ or a ‘good territory’ region or a ‘high quality city’, are put forward again and again. The purely economic response based on single indicators is consistently questioned (Jany-Catrice & Marlier, 2013). In particular, there exists a growing concern on the limits of GDP-like measures for measuring societal welfare. The ‘beyond GDP’ discussion is exemplified by the expert commission established by the German Federal Parliament and (in parallel) by the French Government.

On the other hand, it makes perfect sense to attempt constructing aggregate, headline or composite indicators, in order to measure important social phenomena at the regional level. A strictly scientific and disciplined approach can be pursued but also experimental approaches, with empirical weighting and evaluation schemes, are welcomed by the community of policy theoreticians. The major question...
here is of course, the ‘legitimacy’ of the proposed indicator, as it may be considered ad hoc or arbitrary in a subsequent political debate over the results exposed by the evolution of the indicator suggested. This is an interesting and largely unexplored issue (Jany-Catrice & Marlier, 2013).

It is to be stressed again that there do not exist ‘objective’ or ‘neutral’ indicators of any kind. The community does not pursue objective indicators, it rather attempts to construct useful ones. Experience shows that, at the local or regional level (which is the primary focus under the lens of Policy Compass) community participation should involve the peaceful co-existence of economic, social, and environmental goals around some general vision of well-being, and a vision for the future. The construction and monitoring of the indicators should preferably be a community participation process and should concern the setting of goals or benchmarks for monitoring progress of conventional policy along with social capital.

References

CDC (Congressional Data Coalition), 2014. ‘Democracy and open data: are the two linked?’ http://congressionaldata.org/democracy-and-open-data-are-the-two-linked/ (Retrieved 04/2015)


PRIVACY PRESERVATION IN A PATTERN-AWARE AUTHENTICATION SCHEME FOR VEHICULAR AD HOC NETWORKS

Ryma Abassi, Higher School of Communication, Sup’com, University of Carthage, Tunis, Tunisia, ryma.abassi@supcom.rnu.tn

Sihem Guemara El Fatmi, Higher School of Communication, Sup’com, University of Carthage, Tunis, Tunisia, sihem.guemara@supcom.rnu.tn

Abstract

Vehicular Ad hoc NETwork (VANET) is a promising approach in order to improve road safety. In fact, it allows the exchange of useful traffic information to vehicles timely allowing by the same an appropriate reaction. In such context, security becomes a key component in order to be sure that exchanged messages were not modified or inserted for example. However, due to VANET’s characteristics such as high mobility and scalability, classical security solutions are no more applicable as they are. For instance, a classical authentication will compromise privacy. Hence, in this paper, our main contribution concerns a pattern aware authentication scheme preserving privacy in a VANET environment. The proposed scheme relies on the use of temporary keys, certificates as well as challenge/response process. A security analysis showed the efficiency of our proposal concerning short term/long term unlinkability as well as the Sybil attack.

Keywords: VANET, Security, Authentication, Privacy.

1 INTRODUCTION

According to the WHO (WHO, 2013), about 1.24 million people die each year as a result of road traffic crashes. Moreover, without action, road traffic crashes are predicted to result in the deaths of around 1.9 million people annually by 2020. In order to alleviate the threats of these accidents and improve the driving experience, car manufactures and the telecommunication industry have made great efforts to equip each vehicle with wireless devices that allow them to communicate with each other as well as with the roadside infrastructure located in critical points of the road. That’s how the Vehicular Ad hoc NETwork (VANET) is born.

VANET is a self-organized network that can be formed by connecting vehicles aiming to improve driving safety and traffic management with internet access by drivers and programmers (Raw et al., 2013). They are utilized for a broad range of safety applications such as collision warnings and non safety applications such as road navigation. Two types of communication are provided in VANET: Vehicular to Vehicular (V2V) and Vehicular to Infrastructure (V2I). In the first communication type, vehicles communicate directly whereas in V2I, vehicles communicate through routers called Road Side Unit (RSU).

Due to the critical nature of some VANET applications, the security of the information in this kind of network is crucial. In fact, any successful attack can cause loss of lives or financial lose (Raw et al.,
2013). A security system for VANET should then satisfy several requirements such as authentication, verification of data consistency, availability, non repudiation and privacy to be able to thwart any attack (Raya and J.P Hubeaux, 2007). In fact, VANET packets contain life critical information and it is necessary to make sure that these packets are not inserted or modified by attackers. The liability of drivers should also be proved. For instance, in absence of adequate authentication mechanisms, attackers can launch different types of attacks from which masquerading, Sybil attack, GPS spoofing, etc. In masquerading, an attacker pretends to be a legitimate vehicle by using stolen identity. In Sybil attack a malicious vehicle uses multiple identities to send multiple messages and in GPS Spoofing the attacker tries to deceive the legitimate users by hiding its actual location, etc (Gillani et al., 2013).

Furthermore, VANET environment is quite different from other networks due to its high speed mobility nodes and distributed nature. Therefore security requirements in VANETs are also different from other networks (Gillani et al., 2013). They have to deal with mobility, scalability, privacy, heterogeneity, volatility, etc. For instance, authentication mechanisms should preserve the privacy of the authenticated vehicle: drivers may not wish to be tracked wherever they travel.

Several works have been proposed in order to secure VANET. They can be classified into group signature base schemes and pseudonym-based schemes. Group-signature-based schemes (Hu et al. 2011, Chaum and Heijst, 1991, Rivest et al., 2001, Yeh et al., 2011, Studer et al. 2009, Zhang and Xu, 2013) allow every legal vehicle representing an organization (group) to sign a message. Pseudonym-based schemes mainly rely on the public key cryptography (Calandriello et al., 2007, Huang et al., 2011, Lu et al., 2008, Paruchuri, and Durresi, 2010, Sun et al., 2010). They are usually simple and efficient, and can be applied in a variety of scenarios in VANET; however, their main drawback is the complicated management of pseudonyms.

The main contribution of this paper is then, the proposition of a pattern aware authentication scheme preserving privacy using an hybrid approach. In fact, our scheme adapts itself to the different pattern messages used in a VANET such as warning and beaconing in order to offer authentication while assuring the privacy of the authenticated vehicle. This is made mainly through the use of temporary keys changed randomly and certified through a trusted authority. More precisely, V2V warning messages are authenticated using signature, temporary public and private keys, certificates as well as challenge response process. V2V group communication messages are authenticated based on the group type: static group are affected with a group secret whereas dynamic group are based on a defined context, too. Let’s note that in the two cases, secret revocation and update are handled and a group head is in charge of the group authentication process. In the third scheme, V2V beaconing messages are authenticated using temporary keys, too. Finally, I2V and V2I warning messages are authenticated using the same main process based on temporary keys as well as a challenge/response process.

The rest of this paper is organized as follows. In Section 2 some existing works are recalled. Section 3 presents our motivation as well as needed concepts and assumptions. In Section 4, the main proposition is presented, a VANET authentication scheme preserving privacy. A security analysis on impersonation, long term unlinkability, traceability, etc. is achieved in Section 5. Finally, Section 6 concludes this paper.

2 RELATED WORK

Several works were achieved concerning authentication with respect to privacy. According to the approach employed in the scheme, they can be classified into two types: pseudonym-based scheme and group-signature-based scheme.

In (Hu et al., 2011), authors proposed Atcs, an anonymous and traceable communication scheme based on a signature scheme. Atcs provides anonymous, authenticated and traceable communication based on the efficient combination of (t, n)-threshold signature and Weil Pairing. Moreover, Atcs provides the authenticity of signed broadcasting messages to prevent internal attacks. In (Chaum and Heijst, 1991; Studer et al. 2009), a group signature scheme is used. According to this scheme, a public key for signature verification is generated for a group of signers in such a way that any member of the group
can create a valid signature; a verifier is unable to extract the identity of the specific signer. However, this scheme uses a group manager and the privacy of signers is dependent on this latter and there is a need of a maintenance that may take up the system resources. The group signature scheme was improved in (Rivest et al., 2001) by proposing the ring signature scheme. In fact, according to this scheme, there is no need of a group manager anymore and each signer will be able to create an ad hoc group membership even without the knowledge of other members. However, this solution doesn’t provide an effective mechanism to reveal message sender’s identities when necessary. Yeh et al. proposed a portable privacy-preserving authentication and access control protocol in vehicular ad hoc networks (PAACP) (Yeh et al., 2011). PAACP is based on the blind signature concept in order to ensure privacy. A traditional blind signature is similar to a digital signature except that it allows a person to get another person to sign a message without revealing the content of a message. In addition to the essential support of authentication, key establishment, and privacy preservation, PAACP is developed to provide sophisticated differentiated service access control, which will facilitate the deployment of a variety of non-safety applications. However, in the authorization phase, a PAACP is breakable and cannot maintain privacy in VANETs (Wu and Chen, 2014). Recently, Zhang and Lu proposed two novel privacy-preserving authentication schemes. The first one provides a posteriori and a priori countermeasures with a low computational cost in the verification phase and tight security proof. The second one can achieve batch verification on multiple messages (Zhang and Xu, 2013).

Pseudonymous (public keys certified by CA) authentication is proposed in (Calandriello et al., 2007). This latter requires that each node is equipped with multiple credentials, e.g., certified public keys that do not reveal the node identity. Upon receiving a message, a vehicle with the public key of the CA validates the certificate with the certificate revocation list. However, the use of pseudonyms faces a number of shortcomings such as the storage, update, and retrieval in addition to the pseudonym consistency and conflict. In (Huang et al., 2011), a privacy preservation scheme, named pseudonymous authentication-based conditional privacy (PACP) is proposed. PACP allows vehicles in a vehicular ad hoc network (VANET) to use pseudonyms instead of their true identity to obtain provably good privacy. ECPP, an Efficient Conditional Privacy Preservation was proposed in (Lu et al., 2008) to address the issue on anonymous authentication for safety messages with authority traceability. ECPP uses short time anonymous authentication between OBU and RSUs and is based on four parts: a system initialization, OBU short-time key generation, OBU message generation and sending and OBU fast tracking algorithm. However, their scheme is vulnerable to Sybil attacks since RSUs are used as the source of certificates. In (Paruchuri and Durresi, 2010), authors proposed a Protocol for Anonymous Authentication in Vehicular Networks (PAAVE) to address the issue of privacy preservation with authority traceability on VANET. This protocol is based on smart cards to generate on-the-fly anonymous keys between vehicles and RSU. However, the main limit of such approach is the need of additional equipment as well as its maintenance.

In this work, we propose a pattern aware authentication scheme for VANET preserving vehicles privacy. Our scheme is an hybrid one in that sense that it benefits from pseudonyms as well as group signatures concepts. In fact, our scheme adapts itself to the different pattern messages used in a VANET such as warning and beaconing through the use of temporary keys changed randomly and certified through a trusted authority.

3 Preliminaries
The motivation underlying our proposition is three-folds: (1) the growing importance of VANETs and their particularities, (2) the proliferation of attacks aiming such environment and (3) the need of security in order to fully benefit from VANETs.

In this section, we recall some basics concepts then we introduce the considered adversary model.

3.1 Basic Concepts
A VANET includes at least three types of entities (Xiong et al., 2010): the top trusted authorities (TAs), the immobile Road Side Units (RSUs) at the roadside, and the moving vehicles equipped with
on-board units (OBUs). TAs control the network, RSUs are deployed as an infrastructure to provide information or access to the internet while OBUs are installed on vehicles to provide wireless communication capability. This is depicted by Figure 1. Two types of communication are provided in VANET: Vehicular to Vehicular (V2V) and Vehicular to Infrastructure (V2I). In the first communication type, vehicles communicate directly whereas in V2I, vehicles communicate through RSU (Zhang et al., 2011).

![VANET architecture](image)

Figure 1. VANET architecture

In this work, considered systems are any typical VANET with a TA, some stationary RSUs deployed at the roadsides, and a large number of vehicles equipped with OBUs moving on the road.

3.2 Adversary Model

Figure 2 depicts a selfish attack in a VANET. In Figure 2 (a) when there is an emergency such as a fire, the fire engine broadcasts an emergency signal in order to empty the road and to reach the fire incident more rapidly. However, a vehicle may be tempted to empty a way maliciously and can use this technique by spoofing the identity of the fire truck such as depicted in Figure 2 (b).

![Authentication attack example](image)

(a) Normal case: a fire truck sends a liberation message in order to empty the road. Each vehicle receiving this message has to change the lane. (b) Attack case: a vehicle spoofs the identity of the fire truck in order to send a liberation message to other vehicles and to liberate the road maliciously.

Figure 2. Authentication attack example

To avoid this attack, authentication may be used. However, classical authentication techniques may cause a loss of privacy through the recuperation of private information such as the position, the vehicle owner’s identity, etc. Privacy is defined as “the expectation that confidential personal information disclosed in a private place will not be disclosed to third parties, when that disclosure would cause either embarrassment or emotional distress to a person of reasonable sensitivities” (Standler, 1997). This lack of privacy may be dangerous since using private information; several attacks became easier such as the fabrication attack where an attacker uses the RSU’s privacy information to achieve fabrication operation. The attacker can also, after learning the identity of a given vehicle, tracks this
latter and eavesdrops its communication. Hence, private information must be preserved against unlawful disclosure.

Based on the above analysis, we can see that anonymous authentication can be used in order to authenticate vehicles while preserving their privacy. It enables any vehicle to be authenticated without being identified. Hence, while providing authentication and privacy, our scheme has to deal with tracing malicious nodes and group handling, too.

4 PROPOSITION

Four different communication patterns can be identified in a VANET (de Fuentes et al., 2010): V2V warning propagation, V2V group communication, V2V beaconing and I2V/V2I warning. In the following, the proposed authentication scheme is built upon these patterns.

4.1 Notation and Assumptions

Before describing the proposed scheme, the notations used throughout this paper are listed in Table 1.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>V_i</td>
<td>Vehicle i</td>
</tr>
<tr>
<td>T_{pubi}</td>
<td>Temporary public key of vehicle i</td>
</tr>
<tr>
<td>T_{privi}</td>
<td>Temporary private key of vehicle i</td>
</tr>
<tr>
<td>Val</td>
<td>Random value</td>
</tr>
<tr>
<td>C_i</td>
<td>Cipher text of vehicle i</td>
</tr>
<tr>
<td>Resp</td>
<td>Boolean value = yes if vehicle is authenticated = no otherwise</td>
</tr>
<tr>
<td>G_x</td>
<td>V2V group communication</td>
</tr>
<tr>
<td>Cxt</td>
<td>The context of a group corresponding to an area and a time</td>
</tr>
<tr>
<td>Secret_{Gx}</td>
<td>A secret key exchanged between V2V group communication G_x</td>
</tr>
<tr>
<td>Sign</td>
<td>Beacon message signature</td>
</tr>
<tr>
<td>Certif</td>
<td>A certificate</td>
</tr>
<tr>
<td>ID</td>
<td>A vehicle unique ID</td>
</tr>
<tr>
<td>TID</td>
<td>A temporary vehicle ID</td>
</tr>
<tr>
<td>s</td>
<td>A group secret</td>
</tr>
</tbody>
</table>

Table1. Used notation

Moreover, the following assumptions are used:

- The transmission range of an RSU is longer than that of vehicles.
- Symmetric communication range is used: if a verifier vehicle V hears a verified vehicle P, then we can also assume that P can hear V.
- All vehicles use tamper-proof modules (TPMs) used to store sensitive information.
- The TAs as well the RSUs are safe and cannot be attackers.
- The inter-RSU communications are secured.

4.2 V2V Warning propagation

When a vehicle needs to communicate with another vehicle or a group of vehicles in order to declare an incident and/or request road liberation, a V2V warning propagation is used. The authentication process triggered in such case is based on four steps: an initialization, an authentication, an update and a revocation.
**Initialization.** Before a vehicle $v_i$ can join a VANET, it must register with the TA based on the process described in Figure.3. Each vehicle $v_i$ generates a number of key pairs $(T_{pub}v_i^1, T_{priv}v_i^1), ..., (T_{pub}v_i^n, T_{priv}v_i^n)$ and sends a registration request $\text{reg-req}(ID, (T_{pub}v_i^1, ..., T_{pub}v_i^n))$ to the TA. This latter associates it with a temporary identifier $TID$ and sends back the corresponding certificates. The initialization stops when $v_i$ chooses randomly a $T_{pub}v_i$ and the V2V warning messages exchange starts. In order to authenticate this kind of message, we propose to use a signature: when a vehicle detects an incident, it informs other vehicles by exchanging a warning message enriched with the following information:

<table>
<thead>
<tr>
<th>Type</th>
<th>Signature</th>
<th>Certif</th>
</tr>
</thead>
</table>

Where $type$ field identifies the message as follows: Type=1 for a warning message, Type=2 for a liberation message. The $signature$ field corresponds to obtained signature used in order to authenticate the sender. The $certif$ field is used in order to share the used public key of the authenticated vehicle.

**Warning message authentication.** Each vehicle $v_i$ possesses two keys, a temporary public key $T_{pub}i$ and a temporary private key $T_{priv}i$. As depicted by Figure.3,

1. After receiving a warning message from vehicle $v_i$
2. The vehicle $v_c$ verifies the validity of the received certificate.
3. If the certificate is valid,
   3.1 Vehicle $v_c$ sends an a random value $val$ to the vehicle $v_i$.
   3.2 Vehicle $v_i$ calculates its cipher $c_i = [\text{val}]_{T_{priv}v_i}$ and returns it to vehicle $v_c$.
   3.3 The vehicle $v_c$ retrieves $T_{pub}i$ is retrieved from the received certificate $certif$ and decrypts $c_i$ as follows $w = (c_i)_{T_{pub}i}$. According to $w$, $v_i$ can be either authenticated or ignored.
4. If the certificate is not valid, a notification is sent to the TA which associates the TID with the ID and decreases its score $score$. This latter influences the decision of the TA when a vehicle requests updating its certificates. In fact, under a given threshold, the TA will not accept such request. However, when the corresponding ID doesn’t exist no actions can be taken.

Let’s note that when a vehicle $v_i$ is authenticated, while its certificate is valid, vehicle $v_c$ will directly use the public key for communication. In fact, each vehicle maintains an authentication table.
containing respectively, the vehicle TID, the authenticated temporal public key and its time validity. Each time a vehicle needs to communicate, $v_c$ verifies that the vehicle’s TID is in its table and picks up the corresponding public key for message verification. However, once the time validity expires, the corresponding line is dropped and the vehicle has to be re-authenticated.

**Liberation message authentication.** When the type field is affected to the value ‘2’, this is a liberation message. Basically, this kind of message is sent by a specific set of vehicles needing to liberate the road such as ambulance, policy, fire engine, etc such as depicted by Figure. 2 (a).

In this case, privacy is no longer a requirement. Hence, each $vi$ receiving a liberation message, verifies the signature using the public key of the sender retrieved from the RSU playing in such case the role of a Trusted Authority.

**Temporary keys updating.** Periodically, temporal keys are updated. In fact, two updates are possible: (1) updating the used certificate by choosing another one from the set of generated certificates. (2) Regenerating another set of certificates when all certificates are already used.

In the first case, the vehicle will simply choose another pair of keys and the corresponding certificate generated in the initialization phase. However, when all pairs are already used, the second update type is triggered. This is achieved as follows:

1. The vehicle picks a new set of keys pairs ($T'_{pub}$, $T'_{priv}$) randomly.
2. The vehicle sends this set to the concerned TA encrypted with the currently used temporary private key, $T_{priv}$.
3. The TA verifies the validity of the signature: is it a valid vehicle? Is it a valid key? Once the verification done, certificates are generated, stored in a key management table in the TA and sent to the vehicle.

The key management table is used by the TA in order to maintain for each vehicle, the list of its previously used keys/certificates as well as the currently used keys/certificate.

**Temporary keys revocation.** Each TA maintains a revocation list containing removed certificates / keys. In fact, when a vehicle misbehaves and is detected by a given RSU, this latter informs the TA in order to revoke it. The TA copies then the corresponding entry in the key management table to the revocation list. An alert message is finally broadcasted in order to inform other TAs and RSUs.

### 4.3 V2V group communication

Under this pattern only some vehicles having some features can participate to the communication.

**Group formation and communication.** Vehicles that are moving in close proximity of each other can form a group. Formally, a group $G$ is composed by all vehicles having sensibly the same speed and position:

$$\mathcal{G} = \{ \ldots \} \cap \{ \ldots \}$$

where $\triangleleft$ and $\triangleright$ are $\pm$ and $\pm(\ldots)$

Each group is associated to a GroupHead (GH) corresponding to the vehicle having the average speed in the group.

The GH is in charge of monitoring the constituted group i.e. exchanging information with the TA, broadcasting information to the group members, etc. Let’s note that the constitution of this group can be static or dynamic and consequently the GH functions will vary according this characteristic.

**Static group.** in this case, vehicles having to join a group exchange a HELLO message containing a TID and their speed. Each vehicle receiving this message, verifies its own speed. Two cases are then conceivable: (1) if its speed is lesser than the received one, it responds with an ACK-HELLO message containing its TID (2) if its speed is greater, it simply waits for an ACK-HELLO from its neighbors. The vehicle that doesn’t receive any ACK-HELLO is the GH and broadcasts a GH message announcing that a group is constituted and that a GH was chosen.
Inter vehicles communications are then achieved. First, the GH chooses a secret $s$ then it sends this latter to each vehicle $vi$ belonging to the group $G$ encrypted by its public key $T_{pubi}$. The same procedure as in the previous subsection is then used in order to validate the GH’s public key.

When a new vehicle joins the group, the same secret is shared with it following the same procedure whereas when a vehicle quits the group, a secret update isn’t needed since the vehicle is not able to participate to the communication anymore. However, when the GH quits the group, the first vehicle detecting this event will broadcast a GH message informing the rest of the group that a new GH was defined. This latter will be in charge of new vehicles management and doesn’t need to change the secret. Let’s note that having that the group is static, these two group update actions (the arrival of a new vehicle, the departure of a vehicle) will be made rarely and consequently the potential overhead that may be generated will be limited.

**Dynamic group.** This group is constituted by vehicles that are on the same area in a given time and consequently having the same context $cxt$.

$$cxt = \ldots$$

In such case and due to the dynamicity of the group, it is difficult to choose a vehicle as GH; instead we propose that the GH’s role is played by the RSU. Hence,

1. The RSU chooses a random $val$; crypts it as follows $c_l = (val)_{T_{pubi}}$ and returns the obtained value $c_l$ to vehicle $vi$.

2. Upon receiving $c_l$, vehicle $vi$ decrypts it $w = (c_l)_{T_{privi}}$ and sends the obtained value $w$ to the RSU.

3. Each time that a vehicle $vi$ is authenticated by the RSU, it is added to a group $G_x$, affects it to a context $cxt$, stored in the RSU’s group table depicted in Table 2 and sent a secret key $secret_{Gx}$ to all the members of the group encrypted by their temporary public keys respectively.

Once the context is canceled, the corresponding entry in the RSU table is dropped. This is the case when at least the time or the area parameter is not valid anymore.

Moreover, a node incorporating a given context will be added to the corresponding group and the used secret will be sent using the same procedure explained previously.

Let’s note that for both cases (static and dynamic groups), when a group moves and changes the RSU with whom it communicates, the first RSU sends all needed information to the second one. This is made in order to preserve the group handover.

**4.4 V2V beaconing**

Beaconing is designed to announce the presence of vehicles in the neighbourhood: detected by the periodic sending and listening to the beacon packets. Under this pattern, messages i.e. typically, one beacon per 100 milliseconds are sent to the one hop vehicles and are not forwarded. At these rates, the security overhead will be significant. Moreover, the impact of security on safety beaconing can be harmful (Kargl et al., 2008). Hence, any security scheme has to deal with V2V beaconing security while minimizing overhead. Our proposition is then signing beacon messages by their senders. This is depicted by Figure. 4 (a).
Vehicle2’s one hop Neighbors

(a) Vehicle 2 sends a beacon message b2 signed using its temporary private key Tpriv2 to its one hop group.

Signature verification

(b) Once Vehicle 1 and Vehicle 3 received this beacon, they verify its validity by querying the RSU.

Each vehicle receiving this message, verifies the signature validity by querying the RSU such as depicted by Figure. 4 (b). Knowing, the used keys of each vehicle the RSU can statute about the signature validity.

4.5 I2V/V2I warning

These messages are sent either by the infrastructure through RSUs to the vehicles (I2V) or by a vehicle to the RSU (V2I) when a potential danger is detected such as an accident in close proximity.

Obviously, the I2V warning is not concerned by the authentication neither privacy since the RSU is the sender of the warning message.

The V2I however must be authenticated. This is made using the same process as proposed in the beginning of this section where a vehicle shares its certificate with the RSU. This latter verifies the validity of the certificate and sends a challenge to the vehicle that authenticates. The response is then the challenge encrypted with its private key. The vehicle is authenticated when the received cipher equals the calculated one.

5 Security Analysis

In this section, we analyze the security of our proposal in order to verify whether it meets the needed requirements. Two kinds of attacks can be related to authentication in VANETs: impersonation attack where the attacker pretends to be another entity and Sybil attack where the attacker uses different identities at the same time (de Fuentes et al., 2010). Concerning privacy, long-term unlinkability is generally considered (Studer et al. 2009). Finally, an authentication scheme preserving privacy has to ensure traceability in order to prevent the misbehaving vehicle from causing any further damage (Studer et al. 2009).

Vehicle impersonation. Each vehicle sending a message has to sign it. In such case, the verifier can check the validity of the used certificate and consequently a vehicle vi cannot pretend to be another vehicle since it will not be able to use the adequate private key. The same reasoning can be made for the group communication since the GH uses a temporary public key through the same procedure as explained previously.

Defense against Sybil attack and short term linkability. A malicious vehicle might try to obtain multiple temporary keys from a TA to impersonate multiple vehicles. Such behavior is depicted by Figure.5 where vehicle ‘B’ uses four false identities (C, D, E and F) to send false information to vehicle ‘A’ concerning a blocked road for example. If vehicle ‘A’ believes in the authenticity of the received messages, then it will change the road.

In our scheme, a vehicle uses the same temporary key over a short interval. A vehicle cannot perform a Sybil attack since such as explained in Section 4, a vehicle obtains new certificates only after
sending a request signed by the currently used keys whereas the TA revokes them before giving new ones. Moreover, the TA stores the correspondence between the ID and the TID.

**Figure. 5. Sybil attack**

In another hand, some VANET applications require that in the short-term, a recipient be able to link two messages sent out by the same OBU. Let’s note that the ability to track a vehicle in the short term does not hurt drivers’ privacy since the mobility patterns are constrained by roads. Hence, when a vehicle is detected at a given position $p$ at instant $t$, its position at instant $t' > t$ is mathematically predictable.

**Long-term unlinkability.** To protect drivers’ privacy, we require that messages sent by the same vehicle be unlinkable in the long-run. According to our scheme, authentication is made using temporary keys and/or group secret. These latter are changed regularly meaning that for a given vehicle $v_i$ and some messages $m_1,.., m_n$ more than $t$ time apart, an adversary should not be able to determine if $m_1,.., m_n$ originate from the same sender.

In fact, having that keys changed randomly, they provide long-term unlinkability. Moreover, the use of group secret in the case of group message does not allow tracing vehicles since this secret changes periodically i.e. each context change or due to any modification in the group composition. Hence, an eavesdropper would not be able to associate an old key with a new one.

**Traceability.** A misbehaving vehicle should be identified by authorities and even revoked. Using the proposed key management table, tracing a given vehicle would be possible for the TA. In fact, it contains the used keys/certificates for each vehicle allowing its identification while preserving its privacy for other vehicles.

### 6 Conclusion

Security plays an important role in VANET since exchanged messages are in the most related to safety applications. Due to the VANET properties, there are several challenges in designing security models from which authentication vs. privacy. In fact, while authenticating vehicles the privacy of the driver identity as well as the vehicle location must be preserved. Anonymous authentication is a basic solution needed for preserving privacy while ensuring vehicles authentication.

The main contribution of this paper concerns then, the proposition of a pattern aware authentication scheme preserving privacy. More precisely, we considered authentication scheme for V2V warning propagation, group communication as well as V2V beaconing. Our proposition is based on the use of random temporary keys as well as a challenge/response process where the update and the revocation processes are also supported. A security analysis was achieved proving the efficiency of our proposal concerning vehicle authentication the short/long term unlinkability and the Sybil attack defence.

**References**


INVESTIGATION HOW CITIZENS’ TRUST CAN INFLUENCE THE SUCCESS ADOPTION TO E-GOVERNMENT: A LITERATURE REVIEW AND A PROPOSED FRAMEWORK

Latifa Alzahrani, Brunel Business School, Brunel University London, UK. 
Latifa.Alzahrani@brunel.ac.uk

Dr. Wafi Al-Karaghouli, Brunel Business School, Brunel University London, UK. 
Wafi.Al-Karaghouli@brunel.ac.uk

Prof. Vishanth Weerakkody, Brunel Business School, Brunel University London, UK. 
Vishanth.Weerakkody@brunel.ac.uk

Abstract

This paper investigates how citizens’ trust can influence the success of e-government adoption. For this purpose, a critical and systematic review of the current literature on citizens’ trust in e-government has been provided, with a particular focus on the most critical factors influencing citizens’ trust in respect of the adoption of e-government. While there is a great deal of literature concerning e-government adoption, little has been written from the citizens’ trust perspective. The extant literature was identified through six electronic databases, from 2000 to 2014. Academic articles were reviewed if they contained a relevant discussion of the antecedents or factors influencing citizens’ trust in e-government adoption. The findings of this review reveal that there are inadequate research studies that identify the factors of trust in e-government from the multidimensional nature of trust. Most researchers focus on two dimensions of trust in e-government, which are trust in technology and trust in government, with limited consideration of other factors such as the psychology of citizens and any risk factors. Based on the findings of the review, a conceptual framework is proposed by developing the updated D&M IS Success Model to establish a framework which presents the antecedents of trust in e-government adoption.

Keywords: E-government, Adoption, Citizens’ Trust, Factors

1. INTRODUCTION

The rapid development of information and communication technology (ICT) provides opportunities for businesses and governments around the world to improve their services and information delivery to citizens, businesses and other government agencies through e-government applications (Albusaidy and Weerakkody, 2008; Abdelgaffar et al., 2010). Many scholars (e.g. Bannister and Connolly, 2011; Janssen and Shu, 2011; Weerakkody and Dhillon, 2008) argue that the successful adoption and acceptance of e-government not only ensures the availability of technical infrastructures, but also other processes, including organizational and social readiness. Thus, the successful adoption of e-government requires a deep understanding of the multi-dimensional nature of e-government by considering both technical and non-technical challenges.

The importance and potential of e-government has motivated scholars to investigate the factors influencing e-government adoption and the processes leading to its successful adoption and implementation. Many of the studies that focus on citizens’ adoption of e-government highlight the roles of trust, security and transparency as prominent factors influencing successful e-government adoption (Carter and Belanger, 2005; Layne and Lee, 2001; Moon, 2002). These studies are further supported by other research claiming that citizens’ trust is one of the significant factors in e-government adoption (Abo-Shanab and Al-Azzam, 2012; Ayyash et al., 2013; Mahmood et al., 2014; Morgeson et al., 2010; Warkentin et al., 2002; Welch et al., 2005). In the context of e-government, most studies focus on how the
technical aspects of technology and government reputation influence citizens’ trust at the adoption stage (Khasawneh et al., 2013; Rehman et al., 2012; Teo et al., 2008; Welech et al., 2005), which is believed to be the major barrier to e-government adoption. Consequently, a lack of clear understanding of the factors and issues that influence citizens’ trust in the adoption of e-government is the motivation for this research.

Trust in e-government services is a complex relationship because it includes many complicated issues that affect citizens’ trust in government services (Abu-Shanab and Al-Azzam, 2012; Alshehri et al., 2012; Bélanger and Carter, 2008; Chopra and Wallace, 2003). Trust in the context of online transactions has been discussed in e-commerce (Chang and Cheung, 2005; McKnight et al., 2000); however, few studies have analysed the role that trust plays in e-government services (Gefen et al., 2005; Horst et al., 2006; Joison, 2009; Schaupp et al., 2009). Carter and Weerakody (2008) agreed that despite cultural differences in the adoption of e-government across countries, “trust” is a universal factor influencing the adoption of e-government. Before using e-government services, it is important that citizens believe that their government will provide the effective managerial and technical resources that are required to implement and secure these online systems (Collesca, 2009). Moreover, citizens must be confident to use e-government and they must have the intention to engage in e-government services.

The aim of this study is to undertake a comprehensive and critical review of the literature relating to trust in e-government, while exploring the factors influencing citizens’ trust in e-government adoption. The paper is structured as follows: Section 2 provides a brief overview of the two concepts of this research, which are: e-government and trust. Section 3 introduces the research methodology adopted to select suitable literature for this research. Section 4 provides the critical findings of this review, followed by Section 5, which discusses and analyses the findings. In Section 6, based on the finding of the review, the D&M IS Success model is developed as a framework to the antecedents of citizens’ trust in e-government. This final section concludes, with a presentation of the limitations and contributions of the research.

2. BASIC CONCEPTS

Prior to the systematic literature review processes, the authors found the issue of trust in e-government from a citizen's perspective to be complex, as both the concept of trust and that of electronic government are multidimensional. Therefore, these concepts had to be clarified before commencing the study.

2.1 E-government

Although technology has been the driving force behind e-government, it is only one of the important components needed for successful transformation. Many researchers agree that e-government is not just using ICT to implement well-designed websites, but also requires collaboration between ICT, business processes and policy makers in order to respond to citizens’ needs (Bannister and Connolly, 2011; Hunnius and Schuppan, 2013; Janssen and Shu, 2011; Lips et al., 2013; Weerakkody and Dhillon, 2008). While the early stages of e-government development place more focus on making public services available online, e-government is more concerned with transformational changes. It is also more likely that business process reengineering (BPR) will occur, where the main focus is on customer and business processes (Bannister and Connolly, 2011; Janssen and Shu, 2011; Weerakkody and Dhillon 2008). Therefore, e-government refers to the transformational change of government services (Lips et al., 2012). In addition, it plays an adoption role by ensuring that technology systems are taken to all target areas and put into practice to facilitate efficient and effective services. According to Janssen and Shu (2008), e-government is a transformational process that integrates information systems in order to facilitate different functions aimed at creating governments that are customer oriented.

Three transformations are required for e-government to be effective: the provision of ICT-enabled services that focus on the citizen rather than the provider; the elimination of duplicated services by integrating and re-engineering back-office processes and establishing a shared services culture; and an
improvement in the delivery, management and governance of IT-enabled change (Weerakkody and Dhillon, 2008).

The main objective of e-government is to establish a digital state in which information, as well as public services, can efficiently be accessed and offered to citizens electronically (Dwivedi et al., 2011). There are many motivational attributes of e-government adoption and implementation, which can be classified into five main categories namely: managerial, technological, political, economic and social:

- **Economic Forces**: lower expenditure and fewer resources expended in revenue collection. In e-government, the dissemination of information as well as service delivery is efficient, fast and effective.
- **Technological Forces**: the advance of ICT is the major force in the adoption and implementation of e-government.
- **Political Forces**: e-government has the potential to enhance political participation in a country, such as via the discussion of important political issues online and conducting online elections.
- **Social Forces**: e-government is very instrumental in empowering citizens. The concept can be used to ensure end-to-end service delivery that enables citizens to easily access learning and education, leading to the empowerment of the population.
- **Managerial Forces**: management has a great impact on the adoption of e-government, by enhancing the transparency and accountability of governance. Consequently, e-government provides significant opportunities to increase the responsibility for making far-reaching decisions.

Although the adoption of e-government has many benefits for both citizens and government agencies, these benefits may not be appreciated if certain pertinent issues are not resolved at the early stages of an e-government initiative. Seifert and Peterson (2002) demonstrate that “despite the potential opportunities for the implementation of e-government initiatives, there are a number of challenges that could prevent the realization of these anticipated benefits” (p. 203). Table 1 provides a summary of e-government adoption challenges, as identified by many researchers (Al-Busaidy and Weerakkody, 2009; Alshehri and Drew, 2010).

**Table 1: Challenges of E-government Adoption**

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Factors causing this challenges</th>
<th>Description</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Technical Challenges</strong></td>
<td>ICT Infrastructure:</td>
<td>This refers to the physical hardware and software applications to implement business processes and ICT coverage.</td>
<td>Alshehri et al (2012); Layne and Lee (2001)</td>
</tr>
<tr>
<td></td>
<td>Privacy</td>
<td>This refers to citizens’ confidence in providing their personal information on e-government portals.</td>
<td>Al-Khoury and Bal (2004), Al-Joobri (2006); Alshehri et al. (2012)</td>
</tr>
<tr>
<td><strong>Organizational Challenges</strong></td>
<td>IT workforce capability</td>
<td>The requirements for an IT professional employee.</td>
<td>Albusaidy and Weerakkody (2009); Alshehri and Drew (2010)</td>
</tr>
<tr>
<td></td>
<td>Partnership and collaboration</td>
<td>Sharing information between public and private sectors. Also, it includes G2B.</td>
<td>Albusaidy and Weerakkody (2009)</td>
</tr>
<tr>
<td></td>
<td>Training Tools</td>
<td>Lack of computer training courses to promote e-government Advantages</td>
<td>Albusaidy and Weerakkody (2009); Alshehri and Drew, (2010)</td>
</tr>
<tr>
<td>Information Exchange</td>
<td>Resistance to change from manual to electronic processes</td>
<td>UN (2003), Charif and Ramadan (2003), Albusaidy and Weerakkody (2009)</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------------------------------------------</td>
<td>--------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Lack of policy and regulation for e-usage</td>
<td>Laws and regulations that cover applications and related services such as e-payments, e-mail usage, e-crimes, and e-commerce.</td>
<td>Albusaidy and Weerakkody (2009)</td>
<td></td>
</tr>
<tr>
<td>Strategic planning</td>
<td>Strategies and vision for e-government to align its goal and the required services (top management and leaders)</td>
<td>Palanisamy (2004), Alshehri and Drew (2010), Albusaidy and Weerakkody (2009)</td>
<td></td>
</tr>
<tr>
<td>Social Challenges</td>
<td>Culture</td>
<td>It refers to citizens’ beliefs and values regarding the services of e-government. It can influence by social structure, education, language, religion, economic philosophy</td>
<td>Moon (2002), Weerakkody et al. (2007)</td>
</tr>
<tr>
<td></td>
<td>Digital Divide</td>
<td>The differences between those with access to computers and the internet and those without because of a lack of income, necessary skills, or internet access.</td>
<td>Layne and Lee (2001); Carter and Belanger (2005)</td>
</tr>
<tr>
<td></td>
<td>Trust</td>
<td>The level of citizens’ confident regarding the e-government services.</td>
<td>Carter and Belanger (2005); Carter and Weerakkody (2008)</td>
</tr>
<tr>
<td>Individual’s Characteristic Challenges</td>
<td>Age</td>
<td>Differences in the ages of people who adopt e-government services.</td>
<td>Shareef et al. (2012); Alshehri et al. (2012)</td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>Differences between men and women in adopting e-government.</td>
<td>Shareef et al. (2012); Alshehri et al. (2012)</td>
</tr>
<tr>
<td></td>
<td>Education</td>
<td>Level of education.</td>
<td>Shareef et al. (2012); Alshehri et al. (2012)</td>
</tr>
<tr>
<td></td>
<td>Internet Experience</td>
<td>Level of user’s experience with the Internet.</td>
<td>Shareef, Jahankhani &amp; Dastbaz (2012); Alshehri et al. (2012)</td>
</tr>
<tr>
<td>Financial Challenge</td>
<td>Funding</td>
<td>The availability of the existing and expected budgetary resources to achieve the goals.</td>
<td>Moon (2002), Alshehri et al. (2012)</td>
</tr>
</tbody>
</table>

2.1 Trust
Trust refers to a willingness to depend on an unfamiliar trustee, where the trustor does not have “credible, meaningful information about, or affective bonds” for the trustee (McKnight et al., 2002). Mayer et al. (1995) also defines trust as “the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party” (p. 712). However, many researchers have difficulty defining this concept because of the multi-dimensional nature of trust. For instance, McAlister (1995) reports that “although trust’s importance has been acknowledged, the matter of how it develops and functions has received little systematic theoretical attention” (p.25). In addition, Karvonen (1999) agrees that a lack of careful analysis of the concept of trust in sociological, philosophical and technical factors is one of the fundamental causes of the confusing picture of trust.

Scholars (e.g. Chopra and Wallace, 2003; Colesca, 2009) have researched this concept in many disciplines, including psychology, sociology, economics, computer science, organizational science and
business and marketing. Each discipline focuses on different aspects of trust that cause complexity in defining trust, as per the following:

- **Psychological approach:** According to psychology, trust is the psychological state of the individual in which the trustor risks being exposed to the trustee based on positive expectations and the intention of the trustee (Rotter, 1980; Rousseau et al., 1998).
- **Sociology approach:** According to sociologists, trust is the property pivotal to both individuals and social groups (Lewis and Weigert, 1985). Consequently, culture, ethnicity and religious affiliation are important factors influencing trust (Rotter, 1971).
- **Economics approach:** In this approach, trust focuses on the reputation of the trustee and the benefits or risks that will come from the trusting relationship (Colesca, 2009). In order to earn a good profit and build the company reputation, it is important that everyone trusts each other (Dyer, 1997; Ouchi, 1984).
- **Computer science approach:** There are two components of trust in computer science – user trust and system trust. The concept of user trust is extracted from psychology and sociology, according to which it is a subjective expectation that an entity has about another’s future behaviour (Mui, 2003), while the concept of system trust is based on its reliability, effectiveness and security (Sherchan et al., 2013). Moreover, according to Srivastava and Teo (2009), there are two risks associated with trust in technology: privacy and security risks.
- **Organizational science approach:** In an organizational science approach, trust operates at multiple levels. According to Rousseau (1998), trust in an organizational context is based on a combination of institutional structures, group dynamics, and individual psychological processes. It was found by Beatty et al. (2011) that in organizations, trust exists at two levels, trust between two different organizations and the interpersonal trust that exists between two employees of two different organizations. Trust between organizations was found to be different from that of interpersonal trust between employees.
- **Business and marketing:** Trust focuses on the relationship between users and service providers. Thus, in this approach, trust is studied in relation to the influencing factors such as, in the context of e-government, the factor of ease of use, the interface design and perceived usefulness of increased trust between trustor and trustee (Gefen et al., 2005; Horst et al., 2006).

Investigating the concept of trust from different disciplines leads to many types or dimensions of trust. The different types of trust include the following:

- **Knowledge-based trust:** based on the knowledge available to the trustor about the capabilities of the trustee. According to Wang (2010), knowledge-based trust comes from familiarity and past experience with the other party, leading to the building of trust between two parties through reducing uncertainty.
- **Institution-based trust:** defined as “the belief that needed structure conditions are present (e.g. on the Internet) to enhance the probability of achieving a successful outcome” (McKnight et al., 2002). Some researchers (Abu-Shanab and Al-Azzam, 2012; Alsaghier et al., 2009; Srivastava and Teo 2009) divided institution-based trust into structural assurance, defined as “guarantees, regulations, promises, legal resources, or other procedures … in place to promote success” (p. 339), and situational normality, defined as “one’s belief that the environment is in proper order and success is likely because the situation is normal” (p. 339).
- **Calculative-based trust:** based on people’s calculations of the benefits and costs that the other party will face if engaged in an opportunistic behaviour. So citizens tend to trust when the other party has nothing to gain, or if the cost is higher than the benefit of an opportunistic behaviour ((Dashti et al., 2010; Li et al., 2008).
- **Relation-based trust:** based on the past relations between trustor and trustee. According to Rousseau (1998), trust can also derive from repeated interactions between the trustor and trustee.
Personality-based trust: based on belief in the other party, that is, the other party has specific attributes (Wang, 2010) such as competency—the skills, abilities, expertise to satisfy their needs; integrity—the belief that the online trader will behave in an honest way and will adhere to principles and standards; benevolence—whether the vendor focuses on making a profit or on customers’ interest (Lean et al., 2009).

Cognitive-based trust: refers to a situation where people build trust in the trustee based on their first impression rather than any previous interactions they have had (Wang, 2010). According to Li (2008), people resort to using cognition or first impressions if they have not had a prior experience with the trustee. Some researchers (Abu-Shanab and Al-Azzam, 2012; Karvonen, 1999, Srivastava and Teo, 2009) also believe that, when dealing with an unfamiliar trustee, cognitive-based trust is implemented instead of knowledge-based trust.

Disposition to trust: refers to a general propensity to trust others. Rotter (1971) states that a disposition to trust includes generalized expectations about the trustee, and whether the trustor has information or past experience or not.

It was important for the authors to have a deep understanding of both concepts (e-government and trust) before collecting relevant articles. A clear understanding of the basic concept of any research is required to decide whether particular articles should be included in or excluded from the review.

3. METHODOLOGY

In order to study the factors influencing trust in e-government, a desk-research was carried out, leading to the systematic review of the published literature relevant to trust in e-government. Three significant steps were adopted to identify relating articles to this study: collecting articles, selecting relevant papers and analysing data. A total of 23 articles were found to meet the requirements of the review, these are summarized in Table 2.

3.1 Collecting Articles:

In order to collect articles, a search was conducted in April 2014 for literature published in English between 2000 and 2014. Keywords were chosen in respect of the critical factors influencing citizens’ trust in e-government, which include: "trust", "e-government", "factors", “adoption”, “citizens’ perspective” and “antecedents of trust”. Published studies were identified through six electronic databases: Scopus, Summon, Proquest, ACM, ScienceDirect and Google Scholar. This resulted in the initial identification of 237 articles.

3.2 Selecting Relevant Articles

The process of selecting relevant studies involved systematic analysis for each article by reading abstracts to ensure they met specific criteria, as follows: published in English, published after 2000, and containing a relevant discussion of factors influencing citizens’ trust in e-government. Articles that focused on the implementation of e-government or had a limited discussion of factors influencing citizens’ trust in e-government were excluded.

Of the articles studied, 20 were found to meet the requirements of this study. In addition, reference lists of these articles were scanned in order to identify any further relevant articles that were not found in the initial search. Thus, two more articles were found relating to this research. Also, one unpublished study was involved as grey literature which meets the requirement of this review. A final number of 23 articles were found to be relevant to the aim of this study (see Table 2). Each article was read carefully to identify the important factors discussed as the antecedents of trust in e-government. These antecedents of trust were then grouped as common factors. The groupings were reviewed by all participating authors to validate the chosen factors. Four factors emerged from these articles: characteristic of citizens,
technology, government agencies and risk factors. The articles associated with each theme are listed in Table 2.

3.3 Data Analysis
In order to analyse the data collected from the 23 articles, a six-column table was designed to investigate the antecedents of trust in e-government in each article (see Table 2). The columns detail author and year, aim of study, antecedents of trust, dependent variable, research design & research approach, and findings of each study. The column concerning antecedents of trust classifies on four factors: characteristics of citizens, technology, governmental agencies’ factors and risk factors. Each article is grouped into the type of antecedents focused upon in their studies (note: the “×” mark represents the absence of the correspondent factor in the article analysed) (see Table 2).
## Table 2: Sample of Previous Studies on Trust on E-government

<table>
<thead>
<tr>
<th>Author/ Year</th>
<th>Aim of the Study</th>
<th>Antecedents of citizens’ trust (independent variable)</th>
<th>Dependent Variable</th>
<th>Research Approach and Design</th>
<th>Research Finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Warkentin 2002</td>
<td>This study presents how government can encourage citizen’s adoption of e-government by building trust</td>
<td>Experience, disposition to trust, characteristic-based trust</td>
<td>Institution structure assurance</td>
<td>Intention to use e-government</td>
<td>Quantitative. A survey of over 1,000 taxpaying citizens in United States, Latin America, Africa</td>
</tr>
<tr>
<td>Welch 2005</td>
<td>Studying how Internet uses, citizen satisfaction with e-government, and citizen trust in government are interrelated.</td>
<td>Internet use</td>
<td>Overall Satisfaction with government</td>
<td>Trust in e-government</td>
<td>Quantitative approach. A survey with a sample of 806 US citizens</td>
</tr>
<tr>
<td>Bélanger &amp; Carert 2008</td>
<td>Analysing impact of trust and risk perceptions on one’s willingness to use e-government</td>
<td>Disposition to trust</td>
<td>Trust in technology</td>
<td>Trust in government and user satisfaction</td>
<td>Quantitative approach. Survey for 214 citizens in USA</td>
</tr>
<tr>
<td>Teo et al., 2008</td>
<td>Examining the role of trust in e-government’s success using the updated D&amp;M IS success model</td>
<td>Trust in technology, information quality, system quality, service quality</td>
<td>Trust in e-government</td>
<td>Trust in e-government and user satisfaction</td>
<td>Mixed approach Focus group and online survey of 214 Singapore e-government website users.</td>
</tr>
<tr>
<td>Chee-Wee, Tan et al., 2008</td>
<td>This research studies the relationship between citizens’ trust and website quality</td>
<td>Quality of website, ease of use, usefulness</td>
<td>Citizens’ trust in e-government</td>
<td>Quantitative approach. Survey of 689 US citizens</td>
<td>High quality e-government websites, factors of ease of use and usefulness have a positive effect on citizens’ trust in e-government.</td>
</tr>
<tr>
<td>Study</td>
<td>Title</td>
<td>Trust Factors</td>
<td>Research Methodology</td>
<td>Sample</td>
<td>Trust Impacts</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>---------------</td>
<td>----------------------</td>
<td>--------</td>
<td>---------------</td>
</tr>
<tr>
<td>Alsaghi et al., 2009</td>
<td>Conceptualizing citizens’ trust in e-government: application of Q methodology</td>
<td>Trust in technology, perceived usefulness, ease of use, website quality</td>
<td>Mixed Approach</td>
<td>Survey for 402 citizens, interviews and focus group in Saudi Arabia</td>
<td>The factors that influence trust in e-government are disposition to trust, familiarity, trust in internet, website quality and perceived ease of use</td>
</tr>
<tr>
<td>Colesca 2009</td>
<td>Identify the determinants of trust in e-government</td>
<td>Trust in technology, perceived usefulness, perceived quality</td>
<td>Quantitative approach</td>
<td>Survey of 793 citizens in Romania</td>
<td>Technical and organizational trustworthiness, the quality and usefulness, internet experience and propensity to trust, directly enhanced trust in e-government, while age and privacy concerns have a negative influence on trust.</td>
</tr>
<tr>
<td>Srivastava &amp; Teo 2009</td>
<td>Citizen trust development for e-government adoption and usage.</td>
<td>Trust in technology</td>
<td>Qualitative approach</td>
<td>Focus group and interviews</td>
<td>High level of trust in government and in technology leads to citizens’ trust in e-government.</td>
</tr>
<tr>
<td>Abdelghaffar, et al., 2010</td>
<td>The impact of citizens’ trust on using e-government services.</td>
<td>Trust in technology</td>
<td>Quantitative approach</td>
<td>A paper-based survey to 2023 students in Egypt.</td>
<td>Internet security and credibility of e-government services are significant factors that contribute to citizens’ trust towards e-government</td>
</tr>
<tr>
<td>Liu &amp; Zhou 2010</td>
<td>Establishing an e-government trust model from the behavioural perspectives of citizen.</td>
<td>Perceived usefulness, perceived ease of use</td>
<td>Quantitative approach</td>
<td>Online Survey from 304 citizens in China</td>
<td>It is important to consider the role of trust as well as various website quality attributes in understanding e-government success.</td>
</tr>
<tr>
<td>Morgeson, et al., 2010</td>
<td>This study explores the structure of the e-government–citizen trust relationship</td>
<td>Age, internet use, education, income and gender</td>
<td>Quantitative approach</td>
<td>Telephone survey for 787 users of US federal government</td>
<td>E-government does not yet lead to greater satisfaction with an agency interaction nor does it correlate with greater generalized trust in the federal government overall.</td>
</tr>
</tbody>
</table>
4. FINDINGS

4.1 Antecedents of trust in e-government

Most of the articles addressed technical factors and government agencies factors as important antecedents of trust in e-government (see Table 3). In addition, some studies identified perceived risk as a factor influencing citizens’ trust. Perceived risk was studied along with other factors such as technical, and/or government agencies aspects. Furthermore, citizens’ characteristics were also discussed by many researchers as significant factors leading to trust in e-government. However, only two papers out of the 23 studied addressed all four factors influencing citizens’ trust. The research studies by Beldad et al. (2011) and Colesca (2009) study the factors of technology, government agencies, perceived risk and individual characteristics as major factors influencing trust.

Table 3: Findings of antecedents of trust in e-government

<table>
<thead>
<tr>
<th>Antecedents of trust</th>
<th>Number of Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychology</td>
<td>1</td>
</tr>
<tr>
<td>Government + Risk</td>
<td>1</td>
</tr>
<tr>
<td>Technology</td>
<td>1</td>
</tr>
<tr>
<td>Technology + Government</td>
<td>8</td>
</tr>
<tr>
<td>Technology + Risk</td>
<td>2</td>
</tr>
<tr>
<td>Technology + Psychology</td>
<td>2</td>
</tr>
<tr>
<td>Technology + Government + Risk</td>
<td>3</td>
</tr>
<tr>
<td>Technology + Government + Psychology</td>
<td>3</td>
</tr>
<tr>
<td>Technology + Government + Risk + Psychology</td>
<td>2</td>
</tr>
</tbody>
</table>

4.2 Theoretical Foundation

Regarding the use of a theory or model to develop the research model of trust in e-government, the majority of existing studies (11 studies) are not based on a theory or a model. Table 4 presents those studies that are based on a model or a theory in their research. Most studies that use models utilize a Technology Acceptance Model (TAM model) to explore the roles of trust in e-government, by testing the factors of usefulness and ease of use as technical factors. This study also found that only two studies use D & M IS Success Model to explore those factors influencing trust in e-government. Table 4 also shows that some studies utilized more than one model (e.g. Ayyash et al., 2013; Carter and Belanger, 2005; Morgeson et al., 2010; Wang and Lo, 2012).
Table 4: Findings of Theoretical Foundation

<table>
<thead>
<tr>
<th>Theory/Model</th>
<th>Number of Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology Acceptance Model (TAM)</td>
<td>8</td>
</tr>
<tr>
<td>Diffusions of Innovation theory (DOI)</td>
<td>3</td>
</tr>
<tr>
<td>Theory of Planned Behaviour TPB</td>
<td>1</td>
</tr>
<tr>
<td>D &amp; M IS Success Model</td>
<td>2</td>
</tr>
</tbody>
</table>

4.3 Research Approaches

Table 5 shows that quantitative methods have been the dominant methods used in most extant research studies. Qualitative methods were used by only one study, while four studies used a mixed methods approach (both quantitative and qualitative).

Table 5: Results of Research Approaches

<table>
<thead>
<tr>
<th>Research Approach</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantitative Approach</td>
<td>18</td>
</tr>
<tr>
<td>Qualitative Approach</td>
<td>1</td>
</tr>
<tr>
<td>Mixed (Quantitative + Qualitative)</td>
<td>4</td>
</tr>
</tbody>
</table>

5. DISCUSSION

The systematic review involved qualitative, quantitative and mixed methods approaches, with the major focus on quantitative approaches. The studies included in this review were conducted in a number of developed and developing countries: the USA, Romania, China, Singapore, New Zealand, Egypt, Saudi Arabia, Pakistan, Jordan and Africa (see Table 2). The findings of this review demonstrate that most researchers focus on technical and government agencies factors (Carter and Belanger, 2005; Khasawneh et al., 2013; Navarrete, 2010; Teo et al., 2008; Wang and Lo, 2012; Wang and Lu, 2010; Welch et al., 2005). This indicates a large gap in the research into the antecedents of trust leading to the successful adoption of e-government services. Therefore, the antecedents of trust in the context of e-government should be analysed with reference to the four dimensions of technology, government agencies, citizens’ characteristics and risk. Each of these four dimensions may affect e-government adoption, as shown in Figure 1. Thus, more research should be aimed at addressing the factors of risk and citizens’ characteristics as these have been somewhat ignored by the extant research on this topic. Gefen (2002) also agrees that it is important to study the issue of trust in e-government from a multi-dimensional construct in order to have a clear understanding of the concept of trust in e-government, and to address the issue of trust as a whole successfully.
5.1 Technical Factors

In order to investigate the level of the citizen’s confidence to use technology to interact with government agencies, several technical factors were identified as an important parameter. These factors also relate to the citizen’s belief that using technology for governmental services will provide effective services, accurate information and ensure safe transactions (Carter and Belanger, 2005; Pavlou, 2003). Technology has been studied by most researchers (Abo-Shanab and Azzam, 2012; Alsaghir et al., 2009; Ayyash et al., 2013; Carter and Belanger 2005; Liu and Zhou, 2010; Navarrete, 2010; Rehman et al., 2012; Teo et al., 2008; Wang and Lo 2012; Khasawneh et al., 2013; Wang and Lu, 2010; Warkentin et al., 2002; Welch et al., 2005) because it is considered to be one of the primary antecedents of trust in e-government adoption.

The literature reveals that the technical factors of an e-government service have a major impact on the citizens’ beliefs when they interact with e-government services. Some studies use the Technology Acceptance Model (TAM) to identify the technical factors influencing citizens’ trust (Alsaghir et al., 2009; Ayyash et al., 2013; Carter and Belanger, 2005; Chee-Wee et al., 2008; Liu and Zhou, 2010; Morgeson et al., 2010; Wang and Lu, 2010; Wang and Lo, 2012). In the TAM, factors of usefulness and ease of use are the major technical factors that impact individual beliefs. Few studies utilize the D&M IS Success Model to analyse the technical factors (Teo et al., 2008; Ayyash et al., 2013). The D&M IS Success Model identifies three technical factors as major contributors to the success of IS projects: information quality, service quality and system quality. The existing literature also shows that while some studies use a model or theory to identify the technical factors, most studies provide the overall aspect of trust in technology to predict citizens’ trust in e-government.

This paper will focus on utilizing the D&M IS Success Model as it highlights a clear relationship between the website quality and citizens’ intention to use and satisfaction with the service (DeLone and McLean, 2003; DeLone and McLean, 2004). Extending this argument to e-government websites, a citizen’s positive or negative perceptions of website quality (information, system and service quality) may impact his/her trust to adopt e-government services. Thus, there is a strong correlation between a citizen’s belief and the quality of the e-government website.
5.2 Government Agencies Factors
Government agencies were identified as factors to explore citizens’ trust and their willingness to expose vulnerabilities regarding the government that depend on the level of citizens’ confidence and belief in the ability of governments to provide affective services to their citizens. Papadakis (1999) demonstrates that citizens’ confidence in their governments depends on their confidence in the performance of different government institutions. Many researchers (Ganesan and Hess, 1997; Jarvenpaa et al., 1998; McKnight et al., 2002) agree that trust in the government, which refers to the citizens’ perceptions of the ability and integrity of governments to provide effective services to their citizens, is a significant dimension leading to the successful adoption of e-government services.

In e-government research, trust in the government is influenced by the reputation of governmental agencies (Belanger and Carret, 2008; Srivastava and Teo, 2009; Beldad et al., 2011). Reputation refers to the citizens’ belief that an agency is honest and concerned about its citizens (Jarvenpaa et al., 2000). Thus, agencies with a good reputation are perceived to be reluctant to jeopardize their reputational assets by acting opportunistically (Beldad et al., 2011; Srivastava and Teo, 2009). In addition, past experience with a government agency’s website is an important factor influencing trust in the government. Information that is provided by past experience affects citizens’ confidence to use e-government services (Beldad et al., 2011).

5.3 Perceived Risk
Perceived risk is defined as the “consumers’ psychological perception of risks in the process of online shopping, the subjective forecast about the likelihood and the seriousness of loss” (Wang et al., 2010, p. 342). Many researchers (e.g. Abo-Shanab and Azzam, 2012; Ayyash et al., 2013; Horsburgh et al., 2011; Liu and Zhou 2010; Rehman et al., 2012) agree that there is a strong correlation between trust and risk. A large body of research (e.g. Rotchanakitumnuai, 2007; Ruizhong et al., 2010; Wang et al., 2010) discusses the various dimensions of perceived risk, such as financial risk – the loss of money through online services; time risk – the loss of time in search or ordering a service; and finally, technology risk – which considers security and privacy as important issues (Rotchanakitumnuai, 2007; Ruizhong et al., 2010).

5.4 Citizens’ Characteristics
From the existing literature, it can be seen that the citizens’ characteristics (trustor) and their beliefs regarding another party significantly influences trust. Some researchers (e.g. Beldad et al., 2011; Colesca 2009; Warkentin et al., 2002) highlight the factors of disposition to trust and internet experience as important factors that affect citizens’ trust in adopting the services of e-government. Individuals vary significantly in their levels of trust (Mayer et al., 1995). Disposition to trust refers to “a tendency to be willing to depend on others” (McKnight et al., 1998, p. 474), where some people are more naturally inclined to trust anything and anybody, including online entities, despite having limited information about them, while others require a great deal of information on the trust target before placing their trust in them. Consequently, low levels of disposition to trust decrease the level of trust in e-government, while high levels of disposition to trust increase the level of trust in e-government. Many studies highlight disposition to trust as an important factor influencing trust in e-government (Alsaghier et al., 2009; Belanger and Carert, 2008; Colesca, 2009; Warkentin et al., 2002). In addition, some studies report that internet experience can affect one’s tendency to trust in Internet technology, which influences citizens’ trust in e-government services (Alsaghier et al., 2009; Beldad et al., 2011; Colesca, 2009).
6. THE PROPOSED CONCEPTUAL FRAMEWORK
In this section, a framework has been developed for the antecedents of citizens’ trust in e-government. This framework is based on DeLone and McLean’s IS success model (also referred to as the D&M model).

The D&M model is identified widely in the IS literature, which provides a comprehensive model related to information systems success. DeLone and McLean (1992) demonstrate that there are six factors influencing IS success: system quality, information quality, use, user satisfaction, individual impact, and organizational impact. In 2003, DeLone and McLean updated the D&M IS Success model by using “intention to use” with the factor “use” and adding the factor service quality. Thus, system quality, information quality and service quality were important technological factors in the IS success model, as shown in Figure 2.

![Figure 2. DeLone and McLean’s Updated Model (Source: DeLone & McLean, 2003).](image)

In this paper, the D&M model is developed by combining it with other factors to develop trust in an e-government framework. Thus, in the proposed framework, four factors refer to the antecedents of trust in e-government: technological factors, government agencies factors, risk factors, and characteristics of citizens. Each of these factors is explained below.

- **Antecedents of Trust**
Antecedents of trust refer to the groups of factors that influence citizens’ beliefs in using and adopting e-government services. From the systematic review, four factors were identified as influencing citizens to trust e-government: technical factors, government agencies factors, citizens’ characteristics factors and risk factors. Each of these factors is explained below.

  - **Technical factors:** the D&M model identifies three technological factors that influence citizens’ beliefs: system quality, service quality and information quality. Information quality considers the accuracy, completeness and timeliness of the information on the e-government’s website. System quality refers to the performance of the system and its reliability and accessibility. Service quality reflects the citizens’ evaluation of the service quality with respect to the service that they actually receive and what they expect (Wang et al., 2010).

  - **Government agencies’ factors:** the reputation of an agency and past experience are fundamental factors of government agencies (Beldad et al., 2011). A good reputation will develop a citizen’s
trust to adopt e-government services. In addition, the past experiences of citizens and their satisfaction with respect to online services provided by government agencies are found to influence trust in e-government.

- **Characteristic factors of citizens**: the individual characteristic is an important factor influencing citizens’ trust. In this research, disposition to trust and internet experience are considered as characteristic factors influencing trust in e-government.

- **Risk factors**: risk is a significant factor that impacts citizens’ trust in e-government. Technological risk, such as security and privacy, and performance risk are considered as important types of risk impacting trust to adopt e-government.

---

### Figure 3: Conceptual Framework for Antecedents of Trust in E-government

- **Citizens’ Belief**
  - **Trust in e-government**: Trust in e-government refers to an individuals’ beliefs and their expectation about e-government. Figure 3 shows that trust in e-government is influenced by technical, government agencies, risk and citizens’ characteristic factors. Figure 3 also shows that trust in e-government influences citizens’ satisfaction and adoption of e-government.

---
**Citizens’ satisfaction**: The extent to which an e-government service helps a citizen to achieve his/her needs. Like trust in e-government, citizens’ satisfaction is influenced by technical factors, government agencies, risk and citizens’ characteristic factors. The D&M model (2003) indicates that a citizen’s satisfaction influences his/her individual belief which is, in this framework, trust in e-government and adoption of e-government services (see Figure 3).

- **Outcome of Trust**
  The outcome of trust is the successful adoption of e-government services. In this framework, the adoption of e-government is influenced by citizens’ satisfaction and trust in e-government services.

7. **CONCLUSION**
This paper explores the success of e-government adoption from a citizens’ trust perspective. Before studying the antecedents of trust, the concepts of e-government and trust must be explored. This paper provides an overview for the both concepts to facilitate clear understanding for this research.

A systematic review was presented to allow in-depth analysis of the factors influencing citizens’ trust in the adoption of e-government services. The literature review identified three significant points. First, there are inadequate researches that identify the factors of trust in e-government from the multidimensional nature of trust. Most researchers (e.g. Khasawneh et al., 2013, Wang and Lo, 2012, Abu-Shanab and Al-Azzam, 2012, Rehman et al., 2012; Bannister and Connolly, 2011) focus on two dimensions of trust in e-government, which are trust in technology and trust in government, with limit consideration of other factors such as the psychology of citizens and any risk factors.

Second, the literature review revealed that the existing studies investigating the factors that influence citizens’ trust are limited to the theory of the Technology Acceptance Model (TAM) or Diffusion of Innovation Theory (DOI), which both focus on the technology part. In addition, most of the conceptual frameworks used in these studies were conceptualized with a general approach and limited focus on the nature of trust.

Third, in the existing studies of trust in e-government, there is a lack of research regarding the concept of trust, the multidimensional nature of trust, and the properties and types of trust. Thus, the majority of the extant researches focus on the technology or government factors, without consideration to the trust dimensions.

This paper suggests a conceptual framework for the antecedents of trust in e-government based on the literature gap and the D&M IS Success Model. The proposed framework develops the D&M IS Success Model to include not only technological factors, but also the other factors that should be considered such as government agencies factors, risk, and citizens’ psychologies as antecedents of trust in e-government.

8. **LIMITATION AND FURTHER RESEARCH DIRECTIONS**
This study’s first limitation is that the results of this research are based on secondary data analysis of the antecedents of citizens’ trust in e-government. Therefore, the results of this research cannot be considered as complete unless the proposed framework is justified with the support of primary data. The second limitation is that the review of the existing literature was identified from six electronic databases, with consideration of the keywords: “trust”, “e-government”, “adoption”, “citizens’ perspective”, and “antecedents of trust”. Future researchers should explore additional related journals and databases with the use of other combinations keywords such as “digital” and “risk”.

16
9. RESEARCH CONTRIBUTIONS

Three primary contributions are the result of this research. First, this study reviews literature that informs on trust in e-government by exploring the concept of trust and the concept of e-government. By integrating the literature on the concepts of trust and e-government, this study provides a comprehensive understanding of trust in e-government. Second, this study highlights the antecedents of citizens’ trust in e-government by a systematic review of the literature on trust in e-government. In this paper, four types of antecedents influencing citizens’ trust to adopt e-government are investigated: technology, governmental agencies, risk, and citizen’s characteristics. Finally, a conceptual framework is proposed based on the D&M IS Success Model, which explores the technical factors for IS Success and explains how these factors influence citizens’ beliefs. This study developed this conceptual framework by integrating other factors (such as organizational factors, risk, and citizen’s characteristic) that influence trust in e-government and the subsequent adoption of e-government services.

References

Navigation in Politics – Fuzzy Cognitive Map and Argumentation Approach for Policy Modelling and Analytics

Martin G. Löhe, Fraunhofer Institute for Open Communication Systems (FOKUS) Germany, email: martin.loeha@fokus.fraunhofer.de
Yuri Glikman, Fraunhofer Institute for Open Communication Systems (FOKUS) Germany, email: yury.glikman@fokus.fraunhofer.de
Thomas F. Gordon, Fraunhofer Institute for Open Communication Systems (FOKUS) Germany, email: thomas.gordon@fokus.fraunhofer.de
Habin Lee, Brunel University London UK, email: Habin.Lee@brunel.ac.uk
Youngseok Choi, Brunel University London UK, email: Youngseok.Choi@brunel.ac.uk

Abstract

This paper will illustrate with the example of the Policy Compass, an EU research project, how ICT in government can address the need for policy tools with support of AI. In sum, this paper will reveal how open data and AI can support the policy impact modelling and analysis. In the Policy Compass platform, Carneades argumentation systems is being integrated with the Adhocracy e-participation platform and Fuzzy Cognitive Maps (FCMs) plays a critical role to model and simulate the impact of policy. These function shows the potential of applicability of AI technology for a complex policy evaluation as a political process.

Keywords: open data, Fuzzy Cognitive Map, computational models of argument, Governance, policy cycle, Open Government, accountability.

1 Introduction

Information and communication technologies (ICT) are a major driver for innovation in businesses. eGovernment and eGovernance are research areas that try to employ the tools and techniques of ICT to achieve innovations in realms different from those of business: in politics or governance. Both terms refer to interactions and decision-making by a network of actors faced with a collective problem. The EU-funded Policy Compass research project aims at providing ICT-tools for the tackling of these collective decision situations.

Policy Compass project fully utilize the relevant AI technologies, Carneades argumentation and Fuzzy Cognitive Maps, to provide more enhanced tools for coping with complex decision situations for both policy makers and citizens. Based on AI technologies that can make a use of open data, Policy Compass proposes the novel approach to answer to the needs for policy impact modelling tool.

This paper will explore with the example of the Policy Compass how the use of ICT can help to improve political processes. Therefore, we will employ a model of a “policy cycle” to describe the political process and analyse its demands to ICT tools, focusing on the particular phase of “policy monitoring”. Then, we will describe the Policy Compass and its components. Afterwards, we will examine which of these components apply AI methods, how they meet the demands of the political process and how they could improve policy-making. We will conclude that the integration of already available tools in the Policy Compass concept can facilitate problem-solving in the policy cycle.

2 The Cycle of the Political Process and the Demands of Policy Monitoring

The specific needs of the political process reveal that politics is more than office communication, filling out forms and printing them. Like navigation in seafaring, steering the wheel in politics requires setting the destination, persistently monitoring the current position and adjusting the variables for the journey.

Politics are often described as a policy cycle (in eGovernance: Gordon 2005; as a general tool for policy analysis: Bridgman and Davis 2003). In this cycle its participants first set an agenda, i.e. they
identify issues and prioritize them. Second, the participants formulate policies to address issues and consider the pros and cons of these policies as instruments to achieve certain goals (policy analysis). After these deliberations, the third step is to take formal decisions, e.g. the enactment of legislation (policy adoption). Fourth, the policies are implemented: The instruments will be brought to life by converting words into actions by public administrations. Finally, in the monitoring or impact analysis phase, data is collected and analysed to evaluate whether the policy produced the intended impact. The results of this analysis can raise new issues for the agenda setting phase, readjusting the process and restarting the policy cycle.

ICT tools for the support of business map the particular needs of business processes. Just as well, eGovernance- or eGovernment-tools have to reproduce the demands of the political process. Here, we will concentrate on the phase of policy monitoring, its structures and specific demands. Policy monitoring and impact analysis evaluates the effectiveness of policies; it assesses whether the chosen policy has done its job.

Each policy proposal is based on some causal model or theory on how to achieve a certain goal through government action. For example, the participants of the policy cycle could identify crime as an important issue and set the (perceived) security of the town on the agenda. One policy that could be suggested might be to hire more policemen. Here, the causal theory could be roughly described as such: If the government hires more policemen, crime will be more easily detected, more crimes will be solved, potential criminals will be deterred from committing crimes and the people will feel safer. However, any kind of policy will have to be monitored in order to control government effectiveness. In the given case, a first step would be to monitor crime numbers and the perceived security of a town. Does the number of crimes go down and the perceptions of security go up? Policy monitoring first needs reliable data to assess the outcomes and impacts of policies. An increase of policemen on the street and an improvement of perceived security could show a correlation. Yet, in order to call the aforementioned policy a success, improved perception would have to be the causal effect of the policy. This claim would have to be backed by a causal theory that provides arguments for the asserted relation – which might be challenged by competing theories, e.g., that an efficient policy for graffiti removal or the use of CCTV surveillance during the same time period may have caused the reduction in crime. Moreover, policy monitoring would have to consider effects such as an increase in detected crime due to more police being present (less underreporting). Therefore, policy monitoring also needs tools for mapping causal relations among data and for reasoning about underlying causal theories.

This after-the-fact assessment is useful not only for holding elected governments accountable and better informing voters during elections but also for helping policy makers to take corrective action in the next loop through the policy cycle.

3 POLICY COMPASS AND ITS COMPONENTS

The Policy Compass is an online Web platform aiming to enable its users to build theories explaining impacts of policy changes on the basis of available open data. The Policy Compass consists of several web service components integrated through REST APIs (Figure 1).
Registered users can upload datasets from external data sources (open data platforms, etc.) and define their own performance metrics on the basis of them, as well as on the basis of datasets and metrics of other users already available in the Policy Compass platform. The Policy Compass works with time series datasets. The Policy Compass uses open data and openly accessible data from numerous reliable sources such as regional and national open data portals, the World Bank and Eurostat portals, national statistical portals, etc. Data stored in those portals have different structures and formats. All datasets uploaded in the platform have to be normalized to an internal PolicyCompass dataset specific representation to enable their further processing. The metrics and data manager component provides these services. Visualization manager enables users to visualize the metrics.

The events registry component is responsible for the management of historical events (policy changes events) with potential policy impacts. This component provides a user community managed catalogue of such events. The visualization component provides a key service of the Policy Compass, by enabling users to create interactive charts of time series values for different metrics, and annotate them with policy related historical events on the time axis and formulate theories about relations between policies and metric values.

The Fuzzy Cognitive Maps (FCMs) and augmentation component enables users to model multiple competing causal theories about relations between policies and metrics or generate them on the basis of available data. Finally, arguments about which theory best explains the data can be formalised and visualised in argument maps using the Carneades argumentation component. The argumentation component is also intended to support deliberations about the best performance metrics to use for some purpose, or, in the policy analysis stage of the policy cycle, to model and evaluate the pros and cons of policy alternatives, from the perspectives of diverse stakeholders.

4 USING AI TOOLS FOR POLICY IMPACT EVALUATION AND MONITORING

Artificial Intelligence (AI) is an interdisciplinary field studying the design of computer systems which can perform or support tasks typically understood as requiring intelligence, such as perception, reasoning, problem-solving, planning, learning and communicating in natural language. The “strong
AI” thesis claims the computers can potentially perform any task requiring intelligence. The competing “weak AI” thesis claims that certain tasks may be forever beyond the capabilities of computers, as a matter of principal.

As a practical matter, it does not matter much whether or not strong AI is possible. The field of AI has succeeded in producing systems which are capable of performing a broad range of tasks thought to require intelligence, such as playing chess and jeopardy, proving theorems, understanding, translating and generating both spoken and written natural language, visually recognizing objects and faces, diagnosing diseases, analyzing legal issues, flying airplanes and driving cars. Even if strong AI is unobtainable, AI has provided, and will continue to provide, tools and systems which enhance, complement and facilitate human intelligence.

4.1 Commonsense reasoning and the Carneades argumentation system

One traditional subfield of Artificial Intelligence aims to model, simulate and support common-sense reasoning. Commonsense is the ability to intelligently make and justify claims, judgments and decisions under less than ideal conditions, on the basis of incomplete, inconsistent and uncertain information and knowledge. Commonsense requires the ability to jump to conclusions from partial information, using rules of thumb (default rules), and to retract or revise conclusions when further or better information becomes available. Various approaches and methods for realizing commonsense reasoning have been developed and studied in Artificial Intelligence. Early rule-based systems, such as EMYCIN, allowed “certainty factors” to be assigned to rules to allow less than certain conclusions to be derived. Theorem provers for classical logic have been modified to support defeasible and nonmonotonic inferences, where a conclusion derivable from a set of premises may not be derivable from a superset of these premises. Work on computational models of argument in AI broadens the conception of commonsense reasoning to cover dialectical and rhetorical tasks, in addition to logic and inference. The broader scope of computational models of argument includes the development of methods and tools for deliberating about the logical language ontology, the construction of knowledge (facts and rules) from evidence, testimony, documents and other sources, the evaluation of conflicting arguments, taking into consideration various kinds of attack relations among arguments (e.g. rebuttals, undercutters and underminers), the regulation of the process of argumentation in various types of dialogues (e.g. persuasion, deliberation) using argumentation protocols (with turn-taking, termination, commitment and other procedural rules), the selection of the best arguments to put forward in an dialogue, to promote the goals and interests of a party, and the presentation of arguments in a clear and understandable way for a given audience. One strand of research aims to make complex arguments more comprehensible using graphical visualization methods, in argument maps.

In the Policy Compass project, we are using the Carneades argumentation system (Gordon 2010; Gordon 2012; Gordon 2013) to support debates about performance metrics, causal models and policies. The Carneades system is based on a computational model of argument supporting the distribution of the burden of proof among the parties in a dialogue and the evaluation of arguments using variable proof standards, inspired by legal proof standards, such as “preponderance of the evidence” and “beyond reasonable doubt” (Gordon, Prakken, and Walton 2007; Gordon and Walton 2009).

The Carneades system provides web-based, collaborative software tools for:

- (re)constructing arguments using a rule-base of argumentation schemes
- visualizing, browsing, navigating and editing argument graphs
- critically evaluating arguments, using a formal model of structured argument based on argumentation schemes and Dung abstract argumentation frameworks (Dung 1995)
- conducting structured opinion polls generated automatically from argument graphs
- evaluating the effects of policies, rules and laws, by using rule bases in interactive dialogues to construct arguments
- serializing and interchanging arguments using the Carneades Argument Format (CAF) as well as the Argument Interchange Format (Bex et al. 2012).

Carneades is a multi-user Web application with a three-tiered architecture consisting of a relational database back-end, an application logic layer and a web-based user interface.
In the Policy Compass project, Carneades is being integrated with the Adhocracy e-participation platform of the Liquid Democracy association. Adhocracy is a web-based discussion and deliberation platform with support for delegated voting but currently without support for argument reconstruction, evaluation or mapping.

4.2 FCM as an AI tool for policy impact analysis

AI can also help the participants of the policy cycle to analyze the impact of different policy options. Policy monitoring and impact analysis often requires considering large number of qualitative variables and complexity in their interconnections. In such complexly linked systems, it is difficult to judge the impacts of a change on a policy variable as the propagation route may include multiple layers of variables and deliver different level of impacts along with the route. Contemporary analytic tools have limitations in the analysis of such complex systems. Statistical tools including time series, regression analysis, or structural equation modeling tools deal with a limited number of variables and require assumptions on data (normal distribution, for example). Also, they have limitations when testing different scenarios through what-if analysis. Non-parametric tools like analytical hierarchy process and multi-objective optimization models also have limitations in the number of variables and the causal relationships among variables are in most cases hidden in a model and difficult to be analyzed. These tools also are require significant amounts of data for the robust testing of the relationships among variables. Though recent open data initiatives in the public sector increase the amount of data available for policy analysis, significant variables needed for policy monitoring and impact analysis are still, in nature, qualitative.

A fuzzy cognitive map (Kosko 1986) is an AI tool based on fuzzy logic and neural networks. They apply fuzzy logic rules and neural network learning heuristics to model and simulate complex systems. An FCM is usually represented as a labeled directed graph consisting of concepts nodes and arcs. Concept nodes can represent domain specific qualitative variables such as “perceived security of town”. Arcs are used to represent causal relationships among concept nodes. As FCMs are used to represent causal knowledge of complex systems that can have a large number of concept nodes and arcs. The quantification is based on fuzzy values. In most cases, concept nodes takes interval values in [0, 1]. For example, the perceived security of a town can take values [0, 0.3] for “low”, [0.3, 0.7] for “medium”, [0.7, 1] for “high”. On the other hand, quantified values of arcs in an FCM represent the strength of causal relationships. The strength values usually take values in [-1, 1]. A negative causal value between two concept nodes means that strengthening of the predecessor will weaken the successor. Conversely, a positive causal value means that strengthening the predecessor will strengthen the successor. FCMs can be represented as mathematical models based on the quantified values of concept nodes and arcs. The mathematical model can be used to analyze impacts of a policy change (change of value for a concept node) through a change propagation mechanism. This allows policy makers to conduct what-if analysis to see the impact of alternative policy changes on variables in the context of a large number of qualitative variables and complex interrelationships among the variables.

The most time-consuming and difficult task in FCM-based analysis is defining values for concept nodes and the strengths of causal relationships. The Policy Compass facilitates this tasks using open data available on the Internet. Open data used as performance indicators in the Policy Compass can be directly matched with concept nodes in an FCM. Also, the availability of historical data for multiple concepts allows for testing the strength of potential causal relationships among them through statistical analysis. This is expected to enable more data-driven FCM modelling and simulation for policy monitoring and impact analysis.

1 https://liqd.net/en/adhocracy/was-ist-adhocracy/
5 CONCLUSION

The political process poses very particular demands on eGovernance tools. The examination of the policy cycle, focusing on the phase of policy monitoring and impact analysis, revealed a need to access reliable data, tools for modeling causal relations among data and tools for reasoning and arguing about performance metrics, causal theories and policies. Despite of this need from government and public sectors, no relevant tools and platforms have been proposed from both research field and practical area. The Policy Compass project demonstrates that these requirements can be met via open data and artificial intelligence tools for argumentation and causal modeling.

References


Enterprise Systems

Chairs: Dr. Piotr Soja and Dr. Carsten Brockmann
Table of Contents

Track – Enterprise Systems

2. [PID-61] Luay Ahmad Anaya, Leif Skiftenes Flak and Dag Håkon Olsen, “START-UP COMPANY? GET YOUR ERP SYSTEM ASAP!”
3. [PID-104] Josephine Chong and Swati Nagar, “MARKET REACTIONS TO ENTERPRISE SYSTEMS INVESTMENTS IN AN EUROPEAN ECONOMY: AN EVENT STUDY”
DRAWING TOOLS UML TEXTUAL LANGUAGE

Mohammed Mahmoud, Applied Science University, Jordan, email: M_eshtay@asu.edu.jo

Abstract

Expressing structure, behavior, architecture and business process using Unified Modeling Language (UML) notations and UML tools is a basic operation in the process of describing the specifications of newly developing systems. UML is widely accepted as de facto in software engineering and in software development. Many tools have been developed to support UML modeling and are used by project managers, system analysts, business analysts and architects to draw UML diagrams whether by writing the instruction in a text based environment or by direct modeling using visual drawing tools. Each of these tools that use textual based instructions is using its own notation. In order to unify the way of expressing the notations of UML in textual based tools I will present a new notation to be used by or to be mapped to the notations used by the textual/graphic based tool. The new notational language is easy to write, easy to read and does not need any special editors. This language is promising to become standard for textual UML notations.

Keywords: UML modeling, Text-Based Notation, Class Diagram, Use case diagram.

1 INTRODUCTION

Modeling is very important part of the process of software development because it describes and explains the software structure and how the software operates [4]. Modeling is defined as the means that used to capture ideas, define relationships, and analyze requirements by applying well defined notations, namely modeling language. Modeling languages are used to create, display, store and exchange models [17]. UML (Unified Modeling Language) is a general purpose modeling language and is seen as the standard modeling language in software Engineering field [4][2]. UML modeling is used in all stages of software engineering process ranging from requirements and design and going to maintenance phase [1].

The UML is the most used modeling way for modeling in Software engineering [6]. UML is becoming a dominant notation for modeling in software engineering and is hardly to be replaced soon [8]. UML contains set of different types of diagrams that are used to model the structure and behavior of the system [3]. UML Diagrams are classified into two divisions Structural such as class diagram package diagram and Behavioral such as use case and activity diagrams [3]. The most commonly used diagrams are the class diagram, the use case diagram and the sequence diagram [1].

Because of the popularity of UML as modeling language we can find many graphical tools to draw and represent these diagrams, StarUML, IBM Rational Rose Enterprise Edition, SmartDraw and Microsoft Visio are examples of some common used tools to draw. Our concern in this paper is actually related to the textual based drawing tools such as ymul, MetaUML and umlgraph where we can find different tools each of which use it is own notational language to draw specific diagrams.

Due to the diversity and variety of tools used in the modeling, we found many textual based modeling tools and each of which has its own notational language, they have different languages to represent the standard diagrams of UML. Developers of certain tool need to learn another notational language if he/she decided to use another tool, moreover the interoperability between these tools is not existed. Another aspect is that many notational languages are just limited to specific domain of UML diagrams such as class diagram or sequence diagram.
In this paper we present a text-based modeling language called STUML (Standard Textual UML). STUML defines textual notations to be used by any modeling tools that want to give the ability to express any UML diagram textually using a text-based editor and then automatically draw the specified UML diagrams. In this notation I tried to apply some of the design principles mentioned in [4,5] trying to make the notation easy to read, write and deal with by the tools.

In a longer way we see that this language is going to be a standard textual notation for textual UML modelling that help in increasing the interoperability between the tools that depend on textual modelling.

This paper is organized as follows. Section 1 is the introduction. The background of text based notations and text based modeling in addition to two examples of existed text based tools is in Section 2. In Section 3, a description of the proposed notational language STUML is presented. In Section 4, the formal representation of the proposed notation is included. Section 5 is an example of how we can use the language. Finally, conclusions are included in Section 6.

2 Textual Based Modeling

In this section I will list types of tools in addition to a background about text-based modeling, advantages, and disadvantages. We can find three different types of modeling tool that are used these days: tools that use only graphical notations, tools use text based notations and tools use a combination of both [5]. When we review papers that present graphical tools we can see a dominant assumption that graphical tools are better just because they are graphical, on the other hand we see some researchers questioned this in [7]. Defining a textual language is generally easier than defining a graphical notation [9] and because of its interoperability, transparency; extensibility and ease of testing, textual notations are preferable by many developers. Essalmi, Fathi, and Leila Jemni Ben Ayed [11] used EBNF as a way to represent the grammar using class diagrams. Ibrahim, Noraini, et al. [12] provided a formal definition of use case diagram in order to check the consistency between use case diagram and activity diagram


Earl Grey is a modeling language presented in [4] but not formalized. They compared existing modeling languages with a set of features they defined. These features should be existed in general purpose textual languages.

We can summarize set of the advantages of textual notations according to developer's viewpoint as follows:

Content. The first advantage is that you need more space to display the same content graphically comparing to the same description using textual-based notations. It is easier to get an overview about the diagram using the graphical notations but developers usually looking for details and this is can be easily found in textual notations. Textual notations is simpler to print comparing with graphical model.

Speed of drawing. With inexperienced users the use of graphical tools might be simpler and faster but with experienced users the use of textual notation is going to make the creation of graphical model faster and easier specially with an editor that gives features such as autocomplete and instance checking.

Formatting. Using text-based notations the process of formatting and layout is going to be dependent on the tool by applying algorithm that is responsible of the layout. This is going to relief the developer from tasks that are going to take considerable amount of effort and time.

Independency. Because text-based notations is a textual representation; it does not need depend on the tool and does need any specific platform. You can use any text editor to create this representation.
2.1 Existing Textual Modeling Languages

In this section we will present examples of some existing textual modeling languages and we will show some limitations they have.

**yUML**

Is a free online tool for drawing specific set of UML diagrams (class diagram and use case diagram). It is used for light UML design (figure 1.) and is not appropriate for all projects. [14]

```xml
//Class Diagram
[Customer::forename:string; surname:string; salary:double; doShiz()]->orders*:[Order]
[Order]++-0.*:[Lineitem]
[Order].[note:Aggregate root{bg:wheat}]`
```

Figure 1. yUML tool.

yUml is fast and appropriate for small projects but the code is hard to read with a lot of symbols, in addition yUML cannot be used to model all diagrams of UML.

**Umlgraph**

Is a declarative language used to specify and draw UML diagrams. It gives the ability to model class diagrams and sequence diagrams, they use the java syntax to specify class diagrams figure 2. They use function-like pic macros to specify the sequence diagram. The problem is that they are using the syntax of java in class diagram to use UmlGraphdoclet for drawing and they use pic macros to draw sequence diagrams which make the notations restricted with the mentioned languages. Besides it does not represent all diagrams.[15]

```java
class Person { 
    String Name;
    
    class Employee extends Person { }

    class Client extends Person { }
```

Figure 2. Umlgraph tool.
3 STUML

In this section I will present the notation that I suggest to be a unified modeling notation for drawing tools. This notations is called Standard Textual Unified Modeling Language - STUML and I will apply features proposed in [4], the importance of defining text-based notations for UML modeling comes from the truth that textual languages are easier to formalize, independent of tools capability, we can embed this language into existing tools easily and many developers would prefer to the use of such a textual notations for the simplicity they are used to do so in programming. [4,5].

3.1 Features

In our text-based language we are not going to focus on features such as type-declaration, variable declaration or object orientation because the purpose is not to build programming language but on the other hand we will focus on properties such as readability, simplicity and unambiguity.

Wide Scope: STUML should be able to give the required structures to describe UML models whether they are static (Structural Diagrams) or dynamic (Behavioral Diagrams).

Simplicity: The language should be easy to read and understand to take the advantage of the text-based notations of giving the developer more details about the diagram. The language should be easy to write to encourage developers to use it instead of graphical modeling which is more attractive [7].

The simplicity of the language comes from the eliminating of using complicated symbols to represent relationships such as those used in yUML and metaUML and instead using simple known expressions such as implements and extends for example.

In the proposed language I tried to use structure that is close to the syntax used in many popular programming languages such as C#, C++ and Java which will make the language easy to read and write.

Unambiguous

The language should be unambiguous to avoid misinterpretation and to ease the development of the tools that will adopt STUML as a textual modeling language. The notations will not produce more than parse tree for the same expressions.

4 Notations

In this section, I will use Context Free Grammar to describe the language that might be used by any tool that is going to use textual based notation to draw UML diagrams. Context free grammar is four tuple describing a language L.[13]

Grammar consists of

- Terminals: the basic symbols from which strings are constructed.
- Non-terminal: they help in defining the language.
- Start symbol: one nonterminal is specified as start symbol.
- The productions of the grammar: each production consist of nonterminal, \( \rightarrow \), and body of zero or more terminals or non-terminals.

A context-free grammar [11] is a 4-tuple \( G = (V, \Sigma, R, S) \), where:

- \( V \): A finite set of non-terminals.
- \( \Sigma \): A finite set of terminals.
- \( R \): A finite set of production rules.
- \( S \): A start symbol.
4.1 Notational Conventions

I will use the following conventions throughout the rest of the paper to distinguish between terminals and non-terminals.

- Anything italic and start with Upper case letter is non-terminal.
- The letter S is the start of the grammar.
- Anything Bold is terminal.
- The digits and letters are terminals
- Punctuation symbols such as parentheses, comma, and so on
- | means or.

4.2 Class Diagram

Class diagram is one of the fundamental diagram types and one of most used diagrams in software development; it describes the classes and interfaces existing in the system, properties, features and variables, operations in addition to the relationships between classes such as generalization and association. The generalization relationship indicates that there a relation between super class and sub class where the sub class inherits the properties of the superclass. We can say that the subclass is a specialization of the super class. Another important relation is the association relationship which identifies a link between objects. \[16\]

4.2.1 Class Diagram Components

- Class.
- Abstract class.
- Interface.
- Properties (names, types and values).
- Operations (names, types, attributes).
- Generalization.
- Association.
  - Aggregation.
  - Composition.
- Visibility.

To define classes we will use class and to define interface we will use interface and then we will use attributes to start defining attributes. In each attribute we will specify the name and the type in addition to visibility then we will define operations the same way we did with attributes.

The inheritance relation will be defined using is-a and different styles will be used to define different kinds of associations.

4.3 Use Case Diagram

Use case diagram is a diagram used to expose the relationships between the actors and the use cases in the system.

Use case diagram has been widely used to model the functional requirements of the system. In addition it has been considered as an effective way to communicate with stakeholders.\[10\]

Use case diagram consists of set of objects. These objects are the actors, the use cases and the relationships. An actor is a person, organization, thing, or external system that interacts with the system and can initiate a use case \[16\].

Actors are usually a stick figure in use case diagram. Use cases are ovals.

Use cases represent the functionality of the system. A set of actions that the system does can be represented using use cases. The relationships might be between actor and actor, between actor and use case, and between use case and use case. \[12\]
Several types of relationships might be used in use case diagram:
- Association between actor and use case
- Association between use case and use case
  - Extend
  - Include
- Generalization between actor and actor
- Generalization between use case and use case

4.4 Grammar

Below I will specify part of the grammar by listing the production and applying the above notational conventions. Any digits, signs such as < and -, and boldface strings such as start are terminals. An italicized name is a nonterminal. [11]

\[
G = (V, \Sigma, R, S)
\]

\[
V = \{ \text{Diagram}, \text{classDiagram}, \text{usecaseDiagram}, \text{Objects}, \text{Object}, \text{Class}, \text{Interface}, \text{Relation}, \text{Modifier}, \text{id}, \text{CAttributes}, \text{COperations}, \text{CAtts}, \text{CAtt}, \text{Visibility}, \text{AttDef}, \text{Type}, \text{Ops}, \text{Op}, \text{OpType}, \text{OpParams}, \text{params}, \text{intAttributes}, \text{intOperations}, \text{intAtts}, \text{intOps}, \text{IntOp}, \text{Relation}, \text{Binary}, \text{Composition}, \text{Many}, \text{BinaryRelation}, \text{RelName}, \text{CompType}, \text{side}, \text{Role}, \text{Arity}, \text{ArityType}, \text{Number}, \text{sid}, \text{Digit}, \text{Letter}, \text{character}, \text{blocks}, \text{Entities}, \text{Entity}, \text{Actors}, \text{UseCases}, \text{URelationships}, \text{ActorRel}, \text{UseRel}, \text{ActUseRel} \}
\]

\[
\Sigma = \{ \text{start}, \text{end}, \text{ClassModel}, \text{class}, \{, \}, \text{Attributes}, -, \text{private}, \text{protected}, \text{public}, \text{Integer}, \text{Real}, \text{Boolean}, \text{String}, \text{Operations}, (, ), ;, \text{interface}, \}, \text{void}, \text{abstract}, [], \text{extends}, \text{implements}, \text{relates}, \text{Composite}, \text{Aggregate}, *, \text{connects}, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A…Z, a…z, \text{UseCaseModel}, \text{block}, \text{actor}, \text{usecase}, \text{isa}, \text{include}, \text{extend}, \text{uses} \}
\]

\[
R \rightarrow \{
\text{start} \rightarrow \text{Diagram} \rightarrow \text{end}
\}
\]

\[
\text{Diagram} \rightarrow \text{classDiagram} | \text{usecaseDiagram}
\]

\[
\text{classDiagram} \rightarrow \text{ClassModel} \text{Objects} | \epsilon \text{Objects}
\]

\[
\text{Objects} \rightarrow \text{Class} | \text{Interface} | \text{Relation} \rightarrow \text{Class} \text{Interface}
\]

\[
\text{Modifier} \rightarrow \text{class} \{ \text{id} \} \{ \text{CAttributes} \rightarrow \text{CAttributes} - \text{CAtts} - | \epsilon \text{CAtts} \}
\]

\[
\text{CAtts} \rightarrow \text{Attributes} - \text{CAtt} - | \epsilon \text{CAtt} \rightarrow \text{CAtt} \rightarrow \text{Visibility} \text{AttDef} \rightarrow \text{Visibility} \rightarrow \text{private} | \text{protected} | \text{public}
\]

\[
\text{AttDef} \rightarrow \text{Type} : \text{id}
\]

\[
\text{Type} \rightarrow \text{Integer} | \text{Real} | \text{Boolean} | \text{String} | \text{id} | \epsilon \text{COperations}
\]

\[
\text{COperations} \rightarrow \text{Ops} | \epsilon
\]

\[
\text{Ops} \rightarrow \text{Op} | \epsilon
\]

\[
\text{Op} \rightarrow \text{Visibility} \text{Modifier} \text{OpType} \text{id} (\text{OpParams}); \text{OpParams} \rightarrow \text{params} | \epsilon
\]

\[
\text{params} \rightarrow \text{params} , \text{AttDef} | \text{AttDef} \text{Interface}
\]

\[
\text{interface} [\text{id}][\]

\[
\text{intAttributes} \rightarrow \text{intAttributes} - \text{intAtts} - | \epsilon \text{intAtts}
\]

\[
\text{intAtts} \rightarrow \text{AttDef} | \text{AttDef} ; \text{intOperations}
\]

\[
\text{intOperations} - \text{intOps} - | \epsilon
\]
EXTENDING THE GRAMMAR FOR USE CASE DIAGRAM

UseCaseDiagram → UseCaseModel blocks | ε blocks
UseCaseModel → block [id] Entities blocks | Entities Entities | Entities
Entity | Entity
Entity → Entity | UseCases | URelationships Actors
actor [id]
UseCases → usecase [id]
URelationships → ActorRel | UseRel | ActUseRel ActorRel
[ids] isa [ids]
UseRel → (id|isa(id)) | (id)include(id) | (id)extend(id)
ActUseRel → [id]uses(id)

S = \{S\}

5 EXAMPLE

In this section we will present an example of how we can use the proposed notation STUML to describe class diagram for a modeling tool.

5.1 Example 1

The below code might be used by any tool to be interpreted into UML class diagram as shown in figure 3.

```plaintext
start ClassModel
class [Room] {
  Attributes -
    private Integer : x;
    private Integer : y;
    private Integer : height;
    private Integer : width;
  -
  Operations -
```
public remove();
public clone();

} interface [Drawable] {
Operations -
public remove();
public clone();

} class [Furniture] {
Attributes -
private Real : height;
private Real : width;
private Color : color;

Operations -
public remove()
public clone()

} class [Couch] {
Attributes -
private String : type;
private String : material;

Operations -
public Integer numSeats()

} Implements [Room][Drawable]
Extends [Furniture][Couch]
Composite <Room ; 1><Furniture;0..*>

5.2 Example 2
The below code might be used by any tool to be interpreted into UML usecase diagram as shown in figure 4
start UseCaseModel [Newspaper Web Site]
actor [Journalist]
actor [Reader]
usecase [Post article]
usecase [Authenticate user]
usecase [Find and read article]
(Authenticate user) includes (Post Article)
(Authenticate user) includes (Find and read article)
end

Figure 3. Possible interpretation of the code in example 1.

Figure 4. Possible interpretation of the code in example 2.
6 CONCLUSION

The textual based modeling is a powerful tool that might be preferred by many users for sake of the properties and features mentioned in this paper. Many tools are already existed and offer different notational languages for UML modeling. These already existed tools are not interoperable and each of which has its own notational language. In this paper I presented formalism for a textual language that can be used as a standard for UML textual based modeling tools.

The language is just applied on class diagram model and use case model. In future we will extend the language to cover both branches of UML modeling diagrams, the structural and behavioral models.
References

Abstract

Enterprise Resource Planning (ERP) systems are increasingly being implemented in organizations around the world. Scholars and practitioners consider the ERP system to be one of the most important technological products in an organization. Such systems have the potential to support organizations in their business operations and business growth, and can provide powerful solutions for integrating business processes. While there are a number of studies on ERP system implementation and use in general, little is known about the ERP experience of newly established companies. This exploratory study thus contributes to the growing literature on ERP implementation by studying a large, newly established company. The study applied a qualitative case-study approach to draw from the experiences of a Palestinian company that implemented an ERP system before starting business operation. Our findings suggest that new ventures can experience fewer challenges in realizing business benefits because they can more easily adopt business processes that match the software features. New ventures do not have entrenched business practices, historical business processes, or persistent culture, factors that have been found to impede realization of ERP benefits. This study recommends that new ventures, especially those who have adequate resources and expect to consistently grow in the market, should consider implementing ERP systems in the early stages, because such systems can help in establishing business operations and can support business growth. Further, many of the known barriers that obstruct benefits from ERP systems do not seem to occur in newly established firms.

Keywords: Newly-established, start-up company, new venture, enterprise systems, enterprise resource planning (ERP), benefits realization, post-implementation challenges.

1 INTRODUCTION

Technology is increasingly playing an effective role in organizations’ lives, and many studies have shown evidence of the potential of technology to create and develop business practices in many organizations. ERP systems, in particular, play an important role in business operation and business development. Scholars and practitioners consider ERP systems the most important technological product for organizations (Chen, 2009; Davenport, 1998; Hawking et al., 2004; Melin, 2010; Wagner et al., 2010). Research and practice have paid attention to enterprise systems because these systems have the potential to solve many traditional problems in a company, like scattered systems and files that are not properly integrated and do not provide comprehensive data storage, in addition to other problems such as increasing business complexity (Robey et al., 2002; Wagner et al., 2006). ERP systems are considered standard systems embedded with standard business functions that most organizations have, and these systems are designed to serve businesses in different countries, in different industries and in standard functionality, and can be implemented in different places to serve a single firm (Davenport, 1998; Williams and Pollock, 2012). Many organizations consider such systems as strategically important element for their growth. This makes many small and medium sized
enterprises adopt ERP systems (Malhotra and Temponi, 2010; Soja, 2008; Panorama Consulting, 2013), and makes newly established firms, in early times, adopt the same systems (Chen, 2009).

The literature shows many challenges that face organizations when they start using an ERP system, after it has been successfully delivered. Robey et al. (2002) found two principal challenges that influence system use and influence the reaping of the system's benefits: configuration barriers and assimilation barriers. In other studies, scholars found further challenges, such as usage resistance (Kim et al., 2005), poor change management especially in the case of extensive customization or extensive organizational change (Kim et al., 2005), poor technical competence (Rajapakse and Seddon, 2005; Kim et al., 2005), and misfits between the culture including the new processes introduced by the system compared with the existing organizational culture and the old way of working (Hawari and Heeks, 2010; Peng and Nunes, 2009; Rabaa’i, 2009; Soh et al., 2003), in addition to other challenges.

It can be argued that many of these challenges are largely addressing concerns of established firms. Investigation of the original barriers that face already-established companies shows that these companies have existing systems, entrenched working practices and staff that has historically worked in certain ways. However, little is known about ERP implementation in newly established firms. Such firms are special in that they do not have a historical organizational heritage, such as existing processes, legacy systems, or an established organizational culture. Therefore, there is a call to conduct more studies on ERP implementation in such organizations (Chen, 2009). Most importantly, one of the few studies conducted on ERP implementation in newly established firms (Chen, 2009), indicated that implementing ERP in newly established firms is critical as it can leverage organizational development and build a solid infrastructure for organizational growth. Hence, scholars (e.g. Chen, 2009) have expressed the need for more research on these organizations in order to understand many issues about ERP implementations in newly established firms and their role in business growth, survival and benefits cultivation.

Implementing ERP systems in newly established firms is assumed to be dissimilar from implementing the same system in already established firms. This is because newly established firms have contextual characteristics that differ from established firms. However, it is evident in the literature that some contextual characteristics are affecting ERP implementation and benefits gained from such systems. For example, Soja (2008) found that ERP implementations are influenced with several conditions that vary from one context to another. He argues that factors like company size, implementation scope and adopted modules, are critical criteria to evaluate results from ERP projects. To address the mentioned issues, this study examines business age. That means we aim to focus in ERP implementation in new ventures as ample of existing research does not adequately pay attention to this considerably deviated context.

Interestingly, McDougall and Oviatt (1996) argued that newly established firms that have the potential to grow or who are expecting growth, in addition to private companies seeking internationalization, are required to set up policies, processes, procedures and culture that support such potential development at a very early stage. Otherwise, when these newly established firms grow, they may experience weak performance if they did not establish a substantial and solid base to foster healthy growth in the early stages (McDougall and Oviatt, 1996). Organizations can be considered new ventures or start-up companies until they are six to eight years old (McDougall et al., 2003). Some studies considered new ventures to be those less than eight years old (Biggadike, 1979 in McDougall and Oviatt, 1996; Miller and Camp, 1985), whereas others classified new ventures as those six years old or younger (McDougall et al., 2003). In other studies, this limit was considered to be up to five years (Birley and Westhead, 1994). In this work, we use terms such as 'newly established firms', 'start-up companies' or 'new ventures' to denote organizations that implement ERP systems at a very early stage after their establishment, and particularly before the business operation. This means that such
organizations, when they start their business operation, do not have existing systems, established business processes, or a completely constructed culture. These special features of newly established firms, in addition to the lack of studies about ERP implementations in such organizations, suggest that there is a need for further investigations into newly established firms.

We conducted an exploratory case study to further develop our understanding of whether the challenges discussed in the literature are also applicable to new ventures. We put forward a set of propositions for further research.

This study investigates a Palestinian telecom company that implemented an enterprise system in the early stages of its business start-up, specifically before starting business operations. The paper continues as follows: Section 2 reviews a number of relevant studies and outlines the theoretical perspective that has been considered. Section 3 explains the methodological choices that have been applied. Results are shown in Section 4 and discussed in Section 5. Section 6 presents the conclusions and recommendations.

2 THEORETICAL BACKGROUND

2.1 Barriers to realizing the benefits of implementing ERP systems

An ERP implementation is considered an important organizational practice and its success is seen as a powerful solution for business operations and for staff working in these organizations (Robey et al., 2002). Many studies have been conducted on ERP implementation and on the improved use of the system after implementation. The previous body of research has addressed highly important aspects that influence the gaining of ERP benefits. There are studies focused on benefits classification in ERP projects (e.g. Schubert and Williams, 2011; Shang and Seddon, 2000). There are also many studies focused on the drivers or motives that could lead to more benefits from ERP systems (e.g. Anaya and Olsen, 2014; Davenport et al., 2004; Gattiker and Goodhue, 2005; Seddon et al., 2010; Staehr et al., 2012; Peng and Nunes, 2009). However, a number of studies have focused on the barriers or challenges that could obstruct benefits-realization from ERP systems (e.g. Kim et al., 2005; Markus et al., 2000; Ross and Vitale, 2000; Sedmak, 2010). Accordingly, analysing these different studies results in a set of areas that are considered the main issues that, if dealt with and managed effectively, could lead to improved benefits, and when they are neglected they could lead to a lack of benefits-realization. In this regard, Ross and Vitale (2000, p. 238) state, “It is not clear how many firms that implement ERPs will actually achieve the benefits. It is clear that there are a number of possible pitfalls that put the benefits at risk, and careful planning can reduce the risk of failure.”

There are many particular aspects related to the management of changes, whether it be organizational changes or system changes or modifications. Markus et al. (2000) emphasized the importance of change management, entailing organizational commitment and a high level of functional coordination (Anaya and Olsen, 2014; Kim et al., 2005; Markus et al., 2000; Ross and Vitale, 2000; Staehr et al., 2012). Many scholars have studied the business benefits derived when organizations implementing ERP systems change their business processes to fit the system. In fact, changes on the organizational side are not limited to changes in business processes and rules, but also include changes in the job design (Ross and Vitale, 2000; Staehr et al., 2012). On the other hand, extensive changes of the ERP product to fit the established business processes could lead to poor benefits, as the organization could lose the benefits of the best practices imbedded in the system (Markus et al., 2000). Particularly, new ventures tend to adopt the ERP system because it can equip the organization with ‘best practices’ (Chen, 2009). Most importantly, the large amount of requested changes may create conflict with the ERP structure and logic, and as a result, the staff might prefer not to use the system, leading to marginal benefits (Markus et al., 2000; Robey et al., 2002; Soh et al., 2003). This paper aims to
construct a classification for the barriers found in many studies, to be considered a theoretical base for this work. These main barriers are presented in Table 1.

Furthermore, many studies have found that ERP systems were unable to deliver the expected results because the staff did not use the system in effective ways, which can be attributed to a lack of human expertise and a lack of enthusiasm (Markus et al., 2000; Robey et al., 2002). In particular, Chen (2009) argued that newly established firms can acquire resources from related business groups to make the system work effectively and to obtain value from the ERP system. Furthermore, in many cases the organizations were disappointed with the technical features of the ERP system and its ability to deal with the historical data and the historical reporting mechanism (Markus et al., 2000; Ross and Vitale, 2000).

Table 1 provides a summary of the key barriers to benefits-realization from ERP systems.

<table>
<thead>
<tr>
<th>Key barrier</th>
<th>Literature</th>
<th>Explanations and findings from literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Organizational misfit</td>
<td>Gattiker and Goodhue, 2005; Hawari and Heeks, 2010; O’Donovan et al., 2010; Markus et al., 2000; Robey et al., 2002; Soh et al., 2003</td>
<td>Misfit between the existing systems, processes and culture from one side compared to the new ERP system, and the new processes and new ways of working from the other side.</td>
</tr>
<tr>
<td>2. Technical misfit</td>
<td>Carton and Adam, 2008; Markus et al., 2000; Ononiwu, 2013; Robey et al., 2002; Ross and Vitale, 2000</td>
<td>Dissatisfaction when the ERP system did not fulfil the needs of the business requirements, management reporting and historical data from the legacy systems.</td>
</tr>
<tr>
<td>3. People competence and availability</td>
<td>Anaya and Olsen, 2014; Boudreau and Robey, 2005; Chen, 2009; Kim et al., 2005; Markus et al., 2000; Ononiwu, 2013; Robey et al., 2002; Ross and Vitale, 2000; Saraf et al., 2013; Seddon et al., 2010; Staehr et al., 2012</td>
<td>Weaknesses in dedicated team members, who should be carefully selected, competent, well-educated, motivated and available throughout and after the implementation.</td>
</tr>
<tr>
<td>4. Managing system implementation and managing the requested changes</td>
<td>Kim et al., 2005; Markus et al., 2000; Ross and Vitale, 2000; Sedmak, 2010; Somers and Nelson, 2004; Staehr et al., 2012</td>
<td>Ineffective change management or inappropriate software modifications. Modifying the ERP system to implement the existing processes and rejecting the consideration of ERP as best practice. Lack of effective management for the consequent changes that the system entails, such as changes in roles and responsibilities.</td>
</tr>
</tbody>
</table>

Table 1.  Key barriers that influence benefits gained from ERP systems.

3 RESEARCH METHOD

3.1 Research Strategy

The objective of this exploratory study is to investigate whether implementing an ERP system in a newly established company differs from experiences reported in the existing literature. For this objective, the study employed an in-depth case study strategy. The case study strategy is known for its
ability to conduct exploratory investigations for the phenomenon under study, consequently providing compelling explanations for the findings (Eisenhardt, 1989; Yin, 2009). Furthermore, the case study helps the researchers understand the context, which in this study is a new venture implementing an ERP system before the actual business operation begins. Despite the limitation of the case study in the findings' generalization, it was seen as the most appropriate strategy to handle the data richness (Walsham, 1995). In addition to that, the case study strategy enables the researchers to develop theoretical propositions revealed from the data (Eisenhardt, 1989; Seawright and Gerring, 2008).

However, the case selection procedure was based on theoretical sampling, in which the decision to choose a case was based on a specific purpose (Eisenhardt, 1989). Multiple techniques were used to choose the case under investigation. Initially, the ‘Snowball technique’ (Patton, 2002) was used. In this regards, we sought advice from experts and consultants in Palestine to suggest candidate cases. Afterwards, we purposely selected the case as ‘Stratified purposeful sampling technique’ (Patton, 2002). This study focused in newly established firms, so we ensured the selected company was newly established, as defined earlier, when the company implemented the ERP system. Selecting cases based on environmental variations can clarify the findings domain, and make the study’s results highly pertinent to the environmental characteristics that are chosen to determine the selection (Eisenhardt, 1989). However, as the major stream of research address established firms, this study aims to fill a theoretical categorization ignored in the literature. Hence, we assume the case type is ‘deviant’ according to a classification by Seawright and Gerring (2008). Generally, the purpose of a deviant case study is to “probe for new—but as yet unspecified—explanations…there is also a second, less common reason for choosing a deviant case. If the researcher is interested in disconfirming a deterministic proposition, then any deviant case will do” (Seawright and Gerring, 2008, p. 302). In this study, we assume that many barriers suggested in existing literature, may not be applicable to start-up companies. Thus, we aim to develop a set of propositions, for further research, that can explain how the implementation of ERP system in newly established firms differ from implementing the same system in already established firms. These propositions can provide insights to what extent newly established firms face the challenges that encounter already-established ones. In doing so, we can disconfirm many assertions that argue organizations experience many challenges because of the legacy systems, existing business culture and entrenched business practices, as these aspects are absent in start-up companies.

3.2 Data Collection

This study adopted a qualitative approach, which was helpful to construct a clear and deep understanding concerning ERP implementation in a new venture. Achieving a high level of understanding required conducting several interviews with the many individuals who participated in the system implementation. Many of these individuals worked at the telecom company, but the investigation also included team members working for the consulting company that participated in the system implementation. Hence, the semi-structured interviews were the primary data source in this study. These interviews enabled the researchers to access the people-dependent knowledge by understanding the social world from the viewpoints of the people who are using the system or participated in the system’s implementation (Walsham, 1995). Furthermore, besides conducting the interviews with the main people, other techniques were also used as observation and documents analysis. In order to ensure data validity, the study’s results were discussed with external experts who are familiar with the ERP implementations in Palestine.

The researchers used the tape-recording technique to record the interviews. This technique is recommended in order to capture participants’ views and interpretations in a more effective way (Walsham, 1995). This was supplemented with note-taking to draw the most important interpretations and to record non-verbal events. Finally, transcription was used to document all interviews’ details.
Table 2 provides details about the interviewees, their business roles and the interview duration.

<table>
<thead>
<tr>
<th>Business Role</th>
<th>Duration (in minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chief Financial Officer (CFO) &amp; Project Sponsor</td>
<td>45</td>
</tr>
<tr>
<td>Head of Accounting Section &amp; Functional Consultant</td>
<td>110</td>
</tr>
<tr>
<td>Financial Accountant</td>
<td>45</td>
</tr>
<tr>
<td>Inventory &amp; Fixed Assets Accountant</td>
<td>50</td>
</tr>
<tr>
<td>Technical Consultant &amp; Application Administrator</td>
<td>60</td>
</tr>
<tr>
<td>Cash Management Accountant</td>
<td>40</td>
</tr>
<tr>
<td>Head of Human Resources Section</td>
<td>60</td>
</tr>
<tr>
<td>HR Assistant</td>
<td>40</td>
</tr>
</tbody>
</table>

Table 2. Interviewees' details.

3.3 Data Analysis

Data analysis is the cornerstone of exploratory studies that aim to develop theory from case studies (Eisenhardt, 1989). This section will briefly discuss the technique that lead the researchers outline their conclusion from the huge details they had collected. We applied the ‘within-case analysis technique’ (Eisenhardt, 1989; Yin, 2009) to comprehensively understand the case from different aspects, but, mainly, we paid attention to the specific matters related to the objective of this work. In order to develop this comprehensive understanding, we adopted hermeneutics approach (Klein and Myers, 1999) that entails understanding the whole from the parts, and the parts from the whole. In particular, we analysed every interview, as one part, by looking for interesting concepts related to this study, and we coded these concepts or themes. This process was iterated through all interviews to develop the whole understanding. Furthermore, we were iteratively revisiting the data collected from the interviews, the notes and the secondary sources, like the website and company reports, to make sense of the case, and to ensure that different sources are consistent. However, to conclude a set of themes that are highly related to this study, those reported in the results section, we analysed these themes under the light of the literature by comparing the generated themes with the theoretical constructs developed earlier. In this case, the four key barriers were considered a theoretical template (Langley, 1999) to compare the empirical data through it.

3.4 Case description

This study investigates a Palestinian company, called in this study 'Telco M'. This company provides mobile telecommunication services in Palestine, and started its business operations in 2009. Within its first three years of operation the company engaged about 600,000 subscribers in the West Bank alone. This success was despite the political and economic instability and crises that affect Palestine. Telco M has invested heavily in technology; in 2012 alone, the company invested U.S. $21.4 million for network upgrades and operational information systems. By the end of 2012, the company had 419 employees (of whom 397, representing about 95% of the company staff, had a bachelor's degree or higher), whereas it had only about 150 employees when the system's implementation was begun in early 2009. The company started the implementation of the Oracle E-Business Suite, which is classified as a tier 1 global product (Panorama Consulting, 2013). When the company launched its services to customers in November 2009, many fundamental modules (e.g. general ledger, accounts receivable, accounts payable) of this wide and global system were ready to be used. This system has been viewed as an important component of the technological infrastructure of the company, helping to introduce its business services and streamlining business processes, and leading the company towards more growth.
4 Results

Reviewing the existing literature provides a set of challenges, classified in Table 1. These principal barriers and challenges will be considered through a theoretical lens to examine the extent to which they exist in newly established firms. The findings will be presented according to the illustrated barriers.

4.1 Misfit between the existing culture or processes and the new ERP system

When Telco M implemented an Oracle ERP system in 2009, it had just been established. Therefore, the organizational culture was not yet completely formed. On the contrary, the ERP system was seen as an organizational initiative to effectively contribute to a modern culture for the company. The key functional consultant, who also served as the accounting section head, commented, “When we face a new business requirement or any business issue, the first thing we think about is how this emerged matter could be assimilated into the system, and what functionality and what features are in the system that could help us deal with this issue.” Echoing this, the finance director noted that “in many cases and through the meetings with the company board, we stated some terms used in the ERP system as if they are business terms known to everybody, before being reminded by the board members that these terms were unclear to them.” Therefore, the ERP system, including its terms and concepts, was used by the company staff, and the system brought in new terms to become part of the new corporate culture of the company. The human resources section head commented that “in order to recruit a new staff member and through the job interviews, we ask the applicants if they have worked with the Oracle ERP business suite, as this competence is desirable, and we mention to them that they will work on this ERP system and that this enterprise system will be the main information system that they will work on for most of the day.” However, it is clear from the data that the system brought a new and acceptable culture to the company. The company considered the ERP system a regulation mechanism or a working platform that provided a base for the regular business procedures and rules. In fact, the informants did not see any significant conflict between their business procedures and rules and the ERP system, with some exceptions related to country rules, as the system was seen as a driving force to do work.

This conceptualization differs from the existing literature, which has found that ERP implementations in organizations, particularly in established companies, create a cultural dialectic between the existing working practices and culture on one side, and the practices implied by the ERP system on the other.

4.2 Technical barriers from the software

Since Telco M is newly established and the system has been adapted to use ‘best practices’, there was no obvious misfit between the business requirements and the ERP system. Although the ERP system provided a standard functionality for the company, many informants noted that the system lacked many important reports. Therefore, the company staff, in collaboration with external consultants and implementers, developed a wide range of reports necessary for the company. Most importantly, since the company was a new venture, there was no data migration. In contrast, the literature shows that inconsistencies between old data and new system functionality can create significant problems (Markus et al., 2000) and impede benefits-realization.

4.3 People competence and availability

The company began the implementation with a limited number of staff (120 employees), including only three in the IT department. In addition, the company hired external staff as technical and functional consultants, who were available throughout the implementation and in the first period after the implementation. Telco M also hired staff members, many of them were seniors, in different business functions that would use the system, and hired additional staff when required. By the time the ERP system was in operation, the number of staff was increasing, with approximately 450 employees
by the end of 2012. Staff members underwent training sessions when they joined the company, and the ERP system functionality and procedures were the main topical areas. This meant the staff were learning about the system without unlearning previous procedures. The HR assistant noted, “I was very motivated to work on the system, and I was asking my colleagues, and looking at the system help files and on professional blogs, to learn some features and to learn how to use the system efficiently.” It is worth mentioning that many of the staff who were hired after the implementation were competent in and had previous experience with ERP systems, and were motivated to use the system and to develop their experience working with such systems.

The literature shows that one of the main barriers that obstructs ERP system use and benefits-achievement is staff competence. This is because the existing staff's skills are related to the previous system and they most likely lack the skills related to large-scale ERP systems.

4.4 Managing the implied changes

Telco M did not start its business operations until it had an ERP system. The company started implementation of its ERP system at the beginning of 2009 and continued for nine months, until the system was ready for use in September 2009, the same period during which the company launched its services. The financial director remarked, "We adopted the ERP system to bring best practices for the company and to start our business operation according to these international standards, so we were keen to implement the system without extensive modifications." It can be inferred from the interviewees that the system was seen as a cornerstone of the infrastructural system to drive business operations. Furthermore, because there were no pre-existing work practices, there were no system changes or extensive customizations. The technical consultant and the application administrator from the consulting company responsible for the system implementation said, "The customization was limited and it was just to address country needs, like the currency treatment.” It is also important to mention that there were no new roles or changed job definitions, so there were no wide organizational changes. The job structures had been recently designed and the company had not yet finished the development of these designs, so there were no changes to job practices to be managed, as there would be in an established firm.

5 DISCUSSION

Telco M began its business operations after the ERP system had been put into use. When the company started using the system, it did not have any historical background or existing culture that could resist the new culture (including organized processes for decision-making and a profound reliance on technology and digital means) embedded in the implemented system. In this case, the organizational cultural conflict revealed in many studies (e.g. Markus et al., 2000; Soh et al., 2003) did not seem to exist. This means that Telco M was not attracted to traditional working practices; rather, it was a newly established company that needed an enterprise system as a basis for introducing a modern way of doing business, based on international standards.

The issue of ERP implementation in newly established firms has received little attention in the literature. This study accentuates the need to address this issue in information system research. We, therefore, posit a number of propositions to be addressed in further research.

The literature shows that many organizations prefer to adapt the system to the organization. These adaptations create tension between what is called the 'commodity standard product' and the organizational processes, which sometimes causes serious conflicts with business strategies or with the success of the whole system, especially in the assimilation stage, when people start using the system (Robey et al., 2002; Chen, 2009; Markus et al., 2000; Melin, 2010; Wagner and Newell, 2004; Wagner...
et al., 2010). However, the gap between the new system's functionality and the existing business practices necessitates changes on one of the sides. Organizations either change their business processes to embrace the system's functionality, with a low level of customization, or they change the system to fit the existing business processes (Markus et al., 2000; Melin, 2010; Staehr et al., 2012). This study found that it was beneficial for the case company to create organizational processes that were consistent with the system's functionality; they needed to delay defining the organizational processes until the ERP system was implemented. Therefore, the conflict between existing processes and the new system's functionality, which is apparent in many studies, was not present in this study.

We therefore propose:

P1: Newly established firms experience less organizational misfit between the ERP system and organizational processes than do established firms.

P2: Newly established firms experience fewer challenges in change management related to ERP implementations than do established firms.

However, as illustrated in previous studies (e.g. Markus et al., 2000; Ross and Vitale, 2000), some organizations become quite disappointed with ERP systems because their business requirements and reporting needs were not deliberately met or because they were undermined by the implementers.

There are challenges related to incorporating existing data with the ERP system, especially when some organizations need to retain legacy data for many years (e.g. for regulatory compliance or because their products remain in service for many years) (Markus et al., 2000). However, Telco M, as a new venture, did not encounter such challenges as the company had no historical data. Therefore, the absence of entrenched business practices and existing historical data helped this new venture to implement the new ERP system in a fresh environment without many obstructions, as cited in previous studies.

Hence, we propose:

P3: Newly established firms experience less technical misfit related to ERP implementations than do established firms.

Furthermore, many studies (e.g. Boudreau and Robey, 2005; Kim et al., 2005; Robey et al., 2002; Saraf et al., 2013) have highlighted the importance of a dedicated and motivated team with expertise and enthusiasm to use the system. These studies found that many barriers that obstruct ERP benefits-realization are attributed to the lack of staff members' competence. Such competence would enable them to understand the ERP system and its potential, and would significantly improve their ability to manage such wide-scale systems (Kim et al., 2005). Robey et al. (2002) found that a dedicated core team that is carefully selected, motivated with incentives and empowered to act, along with effectively managed consulting relationships, are critical for responding to configuration challenges; the absence of such staff or their resistance can negatively influence the system's use and the benefits gained from it. However, Telco M only had 120 staff, including three employees in the IT department, when the ERP system was implemented. One challenge faced the company, which is the need for experienced people able to play effective role in the early stages. The company decided to hire a number of key persons who had worked in ERP system in telecom business to be served as advisors to other staff members, and to communicate with the consulting company implemented the system. This is aligned with Chen (2009) who suggested hiring staff in new ventures from relevant business groups. ‘Telco M’ was thus able to hire staff with adequate skills. The interviews showed that the new staff members were motivated to use the system and there was no significant user resistance. Interestingly, the findings revealed that more than 95% of the company's staff had a bachelor's or higher degree. We
conjecture that when new ventures are able to hire competent and educated people, they are less likely to have problems dealing with wide-scale technological systems. This is related to cultural and technical competence, which is one barrier attributed to the lack of benefits-realization from enterprise systems in previous studies (Rajapakse and Seddon, 2005; Robey et al., 2002).

Accordingly, we propose:

P4: Newly established firms are more agile in terms of acquiring the required ERP competences than are established firms.

P5: Staff in newly established firms will demonstrate less resistance to using ERP systems than will staff in established firms.

Therefore, ERP implementation in new ventures seems to be less challenging because it is not laden with the many obstacles found in previous studies. Consequently, this study suggests that new ventures should give priority to implementing an ERP system early in the organizational life cycle and should allocate appropriate funds for it. The implementation will thus be less risky and challenges can be managed more easily. In addition, this can provide more benefits for newly established companies, such as providing a healthy base for business growth and a working environment based on best practices in the field.

We conjecture that it was essential for the case company to implement the ERP system to develop the organizational infrastructure. The company operates in the telecom industry, which is based on the technology, and the management believes in the role of technology to build strong capabilities. Businesses in other industries not have the same belief, and not pay attention to technology products early. Even if this company was newly established, the case revealed that the company hired 150 employees in the first year. The company also allocated a great deal of investment to implement technology products. This issue is most likely absent in many small and medium enterprises that lack the resources and usually do not give priority for investment in technology products (Malhotra and Temponi, 2010). Furthermore, this company was planning to grow, and one reason for their adoption to the system was to help them build a solid base to leverage their future growth. The number of employees was 150 in 2009, and jumped to 419 at the end of 2012. Other businesses, particularly small organizations, do not usually have similar growth rate. We argue that implementing an ERP system in a start-up company can lay an important foundation for growth, but the implementation should take under consideration other factors. We advocate the factors suggested by Chen (2009), as they influence the implementation and even the decision to implement ERP systems in newly established firms. For example, factors like company’s growth stages, the unique industrial characteristics, and having information technology capabilities found as critical contingencies revealed in Chen’s work (2009) and supported in this study. Furthermore, research found that organizations that define their business objectives and align their business strategies with the ERP system implementations are able to create more value and override conflict possibilities (Chen, 2009; Soja, 2008). Thus, new ventures that define their objectives and pay attention to their strategic plans become more able to utilize their ERP systems to leverage growth.

6 CONCLUSION

Based on an exploratory case study, this research suggests that well-known barriers to achieving benefits from ERP systems can be far less problematic for newly established companies than for established companies. Furthermore, we provide interesting suggestions that can assist new ventures in their implementation of ERP systems. Early adoption of an ERP system provided, in our case, a solid foundation for business growth and was seen as a key enabler of business development. At the same time, many barriers that usually affect implementation and benefits from ERP systems in established companies are not clearly present in new companies. For example, dialectic tensions between the
processes implied by the system and the existing organizational processes, and even staff resistance, were not considered very challenging in our case. The findings provide deep insights into the ERP implementation in a newly established company and a set of propositions for further research. It is recognized that employing a single case study has many limitations, especially the generalizability. Despite such limitation, developing the aforementioned propositions reveals the experience of the company under investigation, but certainly these propositions are suggested within many contextual or environmental characteristics discussed like industry environment, the potential to growth, and the capabilities and resources available or can be acquired in early times. Thus, we invite further research, preferably with large samples, to confirm whether organizations that rapidly adopt ERP systems before they have broad functional needs are likely to be more successful than organizations that delay adopting ERP systems until later stages of the organization's life. Finally, there is a need for further studies to investigate the challenges that face newly established firms in particular.

References


MARKET REACTIONS TO ENTERPRISE SYSTEMS INVESTMENTS IN EUROPEAN ECONOMIES: AN EVENT STUDY

Josephine L.L. Chong, Faculty of Business and Law, Auckland University of Technology, Auckland, New Zealand, josephine.chong@aut.ac.nz
Swati Nagar, Faculty of Business and Law, Auckland University of Technology, Auckland, New Zealand, swati.nagar@aut.ac.nz
Xiaowei Wang, Faculty of Business and Law, Auckland University of Technology, Auckland, New Zealand, xiaowei.jason.wang@aut.ac.nz

Abstract

This study explores the impact of enterprise systems investments on the market value of the firm. Using event study methodology, this study identified 31 announcements traded at the London Stock Exchange between 2009 and 2014. This is the one of the few studies to examine the extent to which announcements about enterprise systems would add market value of the firm in the European economies. Our findings confirm the results from previous empirical studies. We find that enterprise systems investments do not generate abnormal positive stock market returns. Also, investors and shareholders do not react positively to the enterprise systems investments during bear markets. This study contributes to the existing information systems research on the value of enterprise systems investment in a European context.

Keywords: Enterprise systems investments, Event study, European economies.

1 INTRODUCTION

In today’s increasingly competitive business environment, information technology (IT) has become pervasive within contemporary organisations as an essential driver of business performance. In particular, enterprise systems play an important role in allowing an organisation to gain or sustain competitive advantages. Increasingly organisations are adopting enterprise systems in order to enhance their efficiency and productivity.

Although important to a firm, it is argued that enterprise systems may not always help an organisation in enhancing its performance to generate an economic payoff (Kohli and Sherer, 2002). Therefore, it is crucial to measure the economic payoff of IT investments. Studies have looked at the market reaction to IT investment announcements in evaluating firm performance. However, there is an ambiguity concerning the mixed results of IT investment announcements on the market value of the firm. Most of these studies found that there is a positive relationship between IT investments and firm performance (Chatterjee et al., 2004; Dardan et al., 2006; Dehning et al., 2003; Ngam and Kautz, 2008; Ranganathan and Brown, 2006). While studies conducted by Dos Santos and his colleagues (1993) and Hunter (2003) have shown that IT investments have a negative influence on firm performance. This phenomenon is known as the IT productivity paradox (Hitt and Brynjolfsson, 1996). As a result, organisations have to evaluate their IT strategies if the benefits from IT investments do not enhance firm performance. In light of an inconclusive consensus on the impact of IT investments on firm performance, it is important to further examine the IT productivity paradox.
Many of these studies were based in the United States (Dehning et al., 2003; Dos Santos et al., 1993; Im et al., 2001) while one was conducted in both Australia (Ngam and Kautz, 2008) and China (Meng and Lee, 2006). Dehning and his colleagues found that transformational IT investments have a positive influence on the performance of firms in the United States. While Ngam and Kautz (2008) found that larger Australia firms tend to reap bigger payoffs from their IT investments. Within the United States, Im et al. (2001) found that the size of IT investments is negatively related to organisational performance; while Dos Santos et al. (1993) showed that non-innovative IT investments have a negative influence on firm performance. More importantly, Meng and Lee (2006) found that the impact of IT investments on the market value of the firm is significant positive in China but not in the United States. This finding reflects the importance of conducting cross country analyses to determine if the effects tended to vary across different countries.

This study has two research objectives. First, this study is to revisit the IT productivity paradox, in particular extending the existing body of knowledge by understanding the effect of enterprise systems announcements on the market value of a firm. Second, this study looks at the effect of IT investment announcements on market value of firms in a European context, which remains relatively understudied.

2 THEORETICAL PERSPECTIVE AND RESEARCH HYPOTHESES

2.1 Adoption of IT in the European economies

The European Union (EU) comprises of 30 countries: Austria; Belgium; Bulgaria; Croatia; Cyprus; Czech Republic; Denmark; Estonia; Finland; France; Germany; Greece; Hungary; Iceland; Ireland; Italy; Latvia; Lithuania; Luxembourg; Malta; Netherlands; Norway; Poland; Portugal; Romania; Slovakia; Slovenia; Spain; Sweden and United Kingdom. During the period of 2001-2008, most of the countries in the EU experienced positive gross domestic product (GDP) growth. However this economic growth was coupled with an increasing sovereign debt during the same period (please refer to Appendix A). The global financial crisis in 2007 further aggravated the European debt crisis. Amongst the European nations, Cyprus, France, Germany, Greece, Ireland, Italy, Portugal and United Kingdom have had a high debt-to-GDP ratio over the last seven years. For example, Cyprus, Greece and Portugal have a debt-to-GDP ratio above 100 percent; while France and Spain have a debt-to-GDP ratio above 90 percent; and Germany has a debt-to-GDP ratio above 80 percent. ICT investment has a significant impact on economic growth not only as traditional investment, but also as a boost to efficiency in growth. A higher level of ICT capital stock per capita allows an economy to achieve a higher growth rate for given levels of growth in labour and capital inputs (Vu, 2005). For this reason, it is important to note that lack of economic growth in the European region has perhaps influenced the level of IT investments made by the firms in the region.

This point is supported by a recent report published by Gartner Incorporation (2014). According to the firm, worldwide spending on IT is estimated to be over US$ 3.8 trillion in 2014, that is a 3.2 percent increase as compared to the worldwide IT spending in 2013 which stood at US$3.6 trillion. Table 1 compares IT spending estimates across four major markets: enterprise software, hardware, IT services and telecom services. Amongst the four markets, the enterprise software market has the largest market capitalisation. For example, the enterprise software spending for 2014 was estimated to be at US$ 320 billion, with an estimated growth rate of 6.9%. In this competitive business environment, contemporary organisations have to invest in enterprise soft-ware systems such as customer relationship management (CRM) and supply chain management (SCM), that are important organisational drivers for positive performance of an organisation.

<table>
<thead>
<tr>
<th></th>
<th>2013 Spending</th>
<th>2313 Growth (%)</th>
<th>2014 Spending</th>
<th>2014 Growth (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device</td>
<td>660</td>
<td>-1.4</td>
<td>689</td>
<td>4.4</td>
</tr>
<tr>
<td>Data Center Systems</td>
<td>140</td>
<td>-0.2</td>
<td>143</td>
<td>2.3</td>
</tr>
<tr>
<td>Enterprise Software</td>
<td>299</td>
<td>4.9</td>
<td>320</td>
<td>6.9</td>
</tr>
</tbody>
</table>
Despite the economic impact, a result of global financial crisis and sovereign debt, organisations in Europe have continued to invest in hardware, enterprise software, IT services and telecommunications equipment and services over the last seven years. Table 2 for example exhibits the percentage of investments in Western and Central Europe for the period 2007-2014. As seen in the table, the European IT market was a steady growth during 2007-2008. However, the spending saw a down turn during 2009-2010. This was largely attributed to the slow global economic growth and the Eurozone debt crisis that engendered uncertainty for both businesses and consumers in the region. As the economies made a gradual recovery, so did the investments in the IT market. For example, as seen in the table, the level of IT spending was at its highest during 2011-2012. This growth was short lived. By the end of 2012, most of the European nations had excessive high debt-to-GDP ratios that plunged the European economies into its worst recession. This in turn engendered a negative impact on IT spending in the region. As seen in the table, IT spending grew by only 0.8% in 2013. There was a gradual recovery in 2014.

2.2 Impact of enterprise systems investments on firms

In contemporary organizations, enterprise systems have become vital for information integration and processing. Enterprise systems are large-scale, real time, integrated application software packages that use computational data storage, and data transmission power of modern information technology to support processes, information flows, reporting, and business analytics within and between complex organisations (Seddon et al., 2010, p. 305). Amongst other enterprise systems this study considers enterprise resource planning (ERP), customer relationship management (CRM), supply chain management (SCM), and e-procurement systems (Shang and Seddon, 2002).

Organisations are pervasively adopting enterprise systems despite their high implementation failure rate. The global ERP market was valued at $47.7 billion in 2011 and is projected to be over $67.7 billion in 2017 (Jacobson et al., 2007; Lucintel, 2012). Over the years, the European nations have been extensively investing in enterprise systems. For instance, the total spending for ERP applications was projected to be US$25.4 billion in 2013 (Columbus, 2013). While the CRM software market is estimated to reach $5.5 billion by the end of 2014.

The primary reason for organisations to adopt enterprise systems is to enhance business performances. ERP are complex business software applications that enable an organisation to integrate its business processes and functions, access real time information and share data among across the whole organisation (Shang and Seddon, 2002). CRM software applications allow an organisation to establish and build a long term relationship with its customers by collecting, analysing and managing their related information so as to maximise the value of the relationship portfolio (Hung et al., 2010; Reinartz et al., 2004). SCM software applications enable an organisation to react more agilely to demand and supply changes, with the use of finite capacity planning algorithms to better manage and integrate organisational logistical and inventory processes (Davenport and Brooks, 2004; Hendricks et al., 2007). Table 3 shows the benefits that organisations can reap from their investments in ERP, CRM, and SCM.

The current literature does not present many event studies that analyse the impact of enterprise systems investment announcements on the market value of a firm in the European context. More
recently, there are only two studies that evaluate the impact of IT investments on the European stock market. Dobija et al. (2012) for example, examines the stock market reactions to IT investment announcements in Poland. Extending the study conducted by Dobija et al. (2012), Janke and Packova (2013) analysed at the impact of ICT investments on performance of firms in Czech Republic, Hungary and Slovakia. Both studies found that the market reacted positively to the investment announcements.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operational</td>
<td>Cost reduction</td>
</tr>
<tr>
<td></td>
<td>Cycle time reduction</td>
</tr>
<tr>
<td></td>
<td>Productivity improvement</td>
</tr>
<tr>
<td></td>
<td>Customer service improvement</td>
</tr>
<tr>
<td>Managerial</td>
<td>Better resource management</td>
</tr>
<tr>
<td></td>
<td>Improved decision making and planning</td>
</tr>
<tr>
<td></td>
<td>Performance improvement</td>
</tr>
<tr>
<td>Organisational</td>
<td>Changing work patterns</td>
</tr>
<tr>
<td></td>
<td>Facilitating organisational learning</td>
</tr>
<tr>
<td></td>
<td>Empowerment</td>
</tr>
<tr>
<td></td>
<td>Building common vision</td>
</tr>
</tbody>
</table>

*Table 3. Benefits of ERP, CRM and SCM (Source: Shang and Seddon 2002)*

Previous event studies investigating the impact of IT investment announcements on the market value of the firm suggest mixed results (refer to Table 4). The primary rationale for the market to react positively to investments is that IT investment is seen as favourable by investors and stakeholders (Jeong and Stylianou, 2010). The investment to a specified technology provides a means for the organisation to compete effectively in a given market. Therefore it can be assumed that organisational IT investments have a positive impact on the market value of a firm (Konchitchki and O’Leary, 2011). Several studies provide empirical findings in supporting that IT investment announcements will produce positive abnormal returns for a firm (Dardan et al., 2006; Filbeck et al., 2005; Hayes et al., 2001; Ranganathan and Brown, 2006).

<table>
<thead>
<tr>
<th>Study</th>
<th>Announcement Type</th>
<th>Key findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chatterjee et al. (2004)</td>
<td>Infrastructure</td>
<td>Positive abnormal returns</td>
</tr>
<tr>
<td>Dardan et al. (2006)</td>
<td>CRM</td>
<td>Positive abnormal returns</td>
</tr>
<tr>
<td>Dehning et al. (2003)</td>
<td>IT investments</td>
<td>Positive abnormal returns</td>
</tr>
<tr>
<td>Dos Santos et al. (1993)</td>
<td>IT investments</td>
<td>Negative abnormal returns</td>
</tr>
<tr>
<td>Filbeck et al. (2005)</td>
<td>SCM</td>
<td>Positive abnormal returns</td>
</tr>
<tr>
<td>Hayes et al (2001)</td>
<td>ERP</td>
<td>Positive abnormal returns</td>
</tr>
<tr>
<td>Im et al. (2001)</td>
<td>IT investments</td>
<td>Negative abnormal returns</td>
</tr>
<tr>
<td>Hunter (2003)</td>
<td>IT investments</td>
<td>Negative abnormal returns</td>
</tr>
<tr>
<td>Ngam and Kautz (2008)</td>
<td>IT investments</td>
<td>Positive abnormal returns</td>
</tr>
<tr>
<td>Ranganathan and Brown (2006)</td>
<td>ERP</td>
<td>Positive abnormal returns</td>
</tr>
</tbody>
</table>

*Table 4. Previous event studies that investigate IT investments and business value*

Over the past four decades, enterprise systems such as ERP, CRM and SCM have evolved to become advanced and are seen as vital technologies that allow organisations to achieve better organisational performance. Currently, it is a norm for most firms to adopt enterprise systems as standard software applications to support their cross-functional organisational processes. When the adoption of an enterprise system does not allow a firm to gain first mover advantage, investors and stakeholders might react negatively to the investment announcement (Konchitchki and O’Leary, 2011). For example, in January 2002, Philadelphia-based CIGNA HealthCare migrated 3.5 million of its members to new claims processing and customer service processes and systems. This broad-based $1 billion initiative included CRM and an overhaul of its legacy technology infrastructure. In the rush to go live, the system’s ability to handle claims and service from front to back and in large volumes was
not adequately tested. Problems in one area cascaded into others; staffing levels were inadequate, and staff was improperly prepared. Rather than realise that benefits would come over time as the company became used to new processes and systems, they expected them the day the switches were flipped. Benefits did not materialise as planned and resulting impacts on customer service caused the nation’s fourth largest insurer to lose 6 percent of its health-care membership in 2002 (Tech Target, 2004). Given the above example, it can be assumed that organisations that are slow in adopting enterprise systems are likely to lag behind those which are quick in adopting thus gaining a first mover advantage (Konchitchki and O’Leary, 2011). This is also supported by two prior event studies that suggest that stock markets react negatively to late adopters (see Dos Santos, 2001 and Im et al., 2001).

Therefore, we hypothesise:

**H1:** Enterprise systems adoption announcements of investments are negatively associated with abnormal return for the firms in Europe.

Further, prior event studies also provide empirical evidence that stock market conditions can have an impact on the stock market reaction (Dobija et al., 2012; Roztocki and Weistroffer, 2009). Typically, there are two types of market conditions: bull markets (favourable markets) and bear markets (unfavourable markets). When an organisation adopts a technology in a bull market, a positive market reaction is expected. Conversely, in a bull market, investors and stakeholders tend to be more optimistic about an organisation’s future economic prospects thus encouraging more IT investments (Roztocki and Weistroffer, 2009). Given the above discussion, it can be assumed that the global financial crisis has resulted into a bear market in Europe since 2008.

Therefore, we hypothesise:

**H2:** The market reaction to the announcement of enterprise systems adoption will be negative for the firms in Europe.

### 3 METHODOLOGY

#### 3.1 Event study methodology

This study adopts an event study methodology with the aim to investigate the impact of an event as assessed by reactions in stock prices (Ranganathan and Brown, 2006). It draws upon the theory of efficient markets, asserting that stock prices fully reflect all available information (Fama, 1970). If the market reacts positively to the released information as announcements, the stock price of the firm will increase. In other words, the IT investment announcement has a positive impact on a firm’s market value (Im et al., 2001).

Event research methodology has been used widely in disciplines such as accounting and finance. For example, accounting researchers tend to use event research methodology to investigate how accounting information disclosure can impact a firm’s future earnings; while finance researchers deploy it to study how mergers and acquisitions can impact a firm’s market value (Im et al., 2001). In the IS discipline, researchers use this methodology to examine the effect of adoption, implementation, purchase and deployment of information systems technology (Konchitchki and O’Leary, 2011).

#### 3.2 Date collection

We obtained enterprise systems adoption announcements of investments in Europe from the period 2009-2014 using Factiva as the main database. Our data includes announcements of IT investments in ERP, SAP, CRM and SCM. The data presented in this paper reflects announcements from 31 public listed firms based in Europe (please refer to Appendix B).
3.3 Data analysis

We used the concept of abnormal returns, or excess returns to evaluate the impact of IT investment announcements on the market value of a firm (Peterson, 1989). The market model is used to calculate the estimated abnormal returns.

\[ R_{it} = \alpha_i + \beta_i R_{mt} + \epsilon_{i,t} \]

where \( R_{it} \) is the return on the share price of firm \( i \) on day \( t \), \( R_{mt} \) is the market return (estimated by FTSEurofirst 300 index for the London Stock Exchange) on day \( t \), \( \alpha_i \) and \( \beta_i \) are the slope parameters for firm \( i \), \( \epsilon_{i,t} \) is an error term for firm \( i \) on day \( t \). \( \alpha_i \) and \( \beta_i \) are estimated by regression analysis of daily returns of stock \( I \) against daily returns of the stock market; over \( T \) days, where \( T = 120 \) begins 123 days before the announcement day (event day) and ending 2 days before the announcement.

We then calculate the abnormal return with the equation:

\[ AR_{it} = R_{it} - (\alpha_i + \beta_i R_{mt}) \]

To reflect the statistical error in the determination of abnormal returns, we standardised AR by dividing it by standard deviation (Peterson, 1989):

\[ SAR_{it} = \frac{AR_{it}}{SD_{it}} \]

\[ SD_{it} = \left( \sum_{t} (R_{mt} - R_{m})^2 / \sum_{t} (R_{mt} - R_{m}) \right)^{1/2} \]

where \( S_i \) is the standard deviation for residual returns for company \( I \) in period \( t \).

To measure the stock price reaction to the enterprise systems adoption investments, two different event windows were used. The first event window starts one day before the announcement day and is denoted as \((-1, 0)\). The second event window ends one day after the announcement day and is denoted as \((-1, 1)\). For each particular window, cumulative abnormal returns (CAR) were calculated by summing the AR for the days of the event window.

To assess the stock price reaction for significance, cumulative standardised abnormal returns (CSAR) for each company were calculated for each of the event windows \((-1, 0)\) and \((-1, 1)\) using:

\[ CSAR_i = \sum_{t=1}^{t_2} SAR_{it} / \left( t_2 - t_1 + 1 \right)^{1/2} \]

\[ Z = \sum_{t=1}^{N} CSAR_i / \left( N \right)^{1/2} \]

Where \( N \) is the number of investment announcements included in the sample. If \( Z \) value is close to zero, it indicates that the observed results are not significant, i.e. that the observed “abnormal” returns may be due to chance variation.

4 DISCUSSION

Table 5 reports the cumulative standardised abnormal returns (CSAR) and \( Z \) values for the 31 IT investment announcements. The test for significance of the effect included the effects on the first event window which is \([-1, 0]\) and the second event window which is \([-1, 1]\).

<table>
<thead>
<tr>
<th>Event window</th>
<th>CAR</th>
<th>Z-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day [-1.0]</td>
<td>-0.354</td>
<td>-1.974*</td>
</tr>
<tr>
<td>Day [-1.1]</td>
<td>-0.714</td>
<td>-3.977*</td>
</tr>
</tbody>
</table>

Table 5. Cumulative standardised abnormal returns

*Significant at 5% level

Hypothesis 1 is supported. The overall effect of enterprise systems adoption announcements was found to be statistically negatively significant. Importantly, our study finds that IT investment announcements were associated with negative abnormal returns on the day of the announcement. This indicated that investors and shareholders reacted negatively to the adoption of enterprise systems. The
result confirmed the results of previous studies whereby investments in enterprise systems do not lead to positive abnormal returns.

Hypothesis 2 is supported. For bear markets, the investors and stakeholders react significantly more negatively to the enterprise systems adoption announcements. This finding reflects the importance of stock market conditions which can influence the behaviour of investors and shareholders in reacting to the stock market. During the crisis period (2009-2014), investors and shareholders reacted negatively to the market.

5 CONCLUSION

This study contributes to the literature on the impacts of IT investments and the market value of the firm. In this study, we have investigated the role enterprise information systems and their impact on firm performance. In particular, our study seeks to re-examine the productivity paradox of information systems. Our findings confirm the results from previous studies which contend that firms do not necessary reap economic payoffs from their costly IT investments. The results confirm that enterprise systems investments do not generate abnormal positive stock market returns. Second, investors and shareholders react negatively to enterprise systems investments during bear markets. We believe that our study should contribute to the information systems literature, whereby it is one of the few studies that focused on enterprise systems investments and their impact on firms in the European economies.

Like every research, this study has one limitation. The main limitation is that the sample size considered in this study is small. Our aim is to expand the number of firms in our future research on this topic. We also hope to draw comparisons of the effect of IT investment announcements on the market value of firms in the United States and compare that to firms in Europe during the 2009-2014 period. This may provide us with a better understanding of the impact of IT investments on stock market reaction.

APPENDIX A

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>1.69</td>
<td>1.69</td>
<td>2.69</td>
<td>2.69</td>
<td>3.67</td>
<td>3.67</td>
<td>3.71</td>
<td>3.71</td>
<td>-3.82</td>
<td>1.77</td>
<td>2.83</td>
<td>0.87</td>
<td>0.39</td>
</tr>
<tr>
<td>Belgium</td>
<td>1.36</td>
<td>0.81</td>
<td>1.83</td>
<td>1.83</td>
<td>2.83</td>
<td>2.83</td>
<td>2.83</td>
<td>2.83</td>
<td>-2.79</td>
<td>2.42</td>
<td>1.84</td>
<td>-0.28</td>
<td>0.09</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>1.84</td>
<td>2.65</td>
<td>5.51</td>
<td>6.75</td>
<td>6.36</td>
<td>6.51</td>
<td>6.49</td>
<td>6.19</td>
<td>-5.48</td>
<td>0.39</td>
<td>1.84</td>
<td>0.78</td>
<td>0.50</td>
</tr>
<tr>
<td>Croatia</td>
<td>4.88</td>
<td>5.37</td>
<td>4.13</td>
<td>4.28</td>
<td>4.94</td>
<td>5.06</td>
<td>5.06</td>
<td>-6.95</td>
<td>-2.28</td>
<td>-0.05</td>
<td>-1.98</td>
<td>-0.60</td>
<td>0.09</td>
</tr>
<tr>
<td>Cyprus</td>
<td>1.03</td>
<td>1.21</td>
<td>1.97</td>
<td>1.83</td>
<td>1.38</td>
<td>1.13</td>
<td>1.13</td>
<td>-1.86</td>
<td>1.31</td>
<td>0.53</td>
<td>-2.43</td>
<td>-8.71</td>
<td>0.09</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>3.10</td>
<td>3.15</td>
<td>0.77</td>
<td>4.74</td>
<td>6.05</td>
<td>7.02</td>
<td>7.02</td>
<td>-4.51</td>
<td>2.47</td>
<td>1.82</td>
<td>-1.24</td>
<td>-0.37</td>
<td>0.09</td>
</tr>
<tr>
<td>Denmark</td>
<td>0.71</td>
<td>0.47</td>
<td>0.38</td>
<td>2.30</td>
<td>2.44</td>
<td>3.40</td>
<td>3.40</td>
<td>-0.78</td>
<td>-5.67</td>
<td>1.58</td>
<td>1.10</td>
<td>-0.38</td>
<td>0.08</td>
</tr>
<tr>
<td>Estonia</td>
<td>7.52</td>
<td>7.94</td>
<td>7.77</td>
<td>6.34</td>
<td>8.85</td>
<td>10.1</td>
<td>7.49</td>
<td>-4.15</td>
<td>-14.1</td>
<td>2.57</td>
<td>9.56</td>
<td>3.94</td>
<td>1.54</td>
</tr>
<tr>
<td>Finland</td>
<td>2.28</td>
<td>1.83</td>
<td>2.01</td>
<td>4.13</td>
<td>4.92</td>
<td>4.41</td>
<td>5.34</td>
<td>0.30</td>
<td>-8.54</td>
<td>3.36</td>
<td>1.73</td>
<td>-0.83</td>
<td>0.65</td>
</tr>
<tr>
<td>France</td>
<td>1.84</td>
<td>0.93</td>
<td>0.90</td>
<td>2.55</td>
<td>1.83</td>
<td>2.47</td>
<td>2.29</td>
<td>-0.08</td>
<td>-3.15</td>
<td>1.73</td>
<td>2.03</td>
<td>0.01</td>
<td>0.19</td>
</tr>
<tr>
<td>Germany</td>
<td>1.64</td>
<td>0.03</td>
<td>0.39</td>
<td>0.59</td>
<td>3.68</td>
<td>3.39</td>
<td>3.10</td>
<td>-5.09</td>
<td>3.86</td>
<td>1.50</td>
<td>0.90</td>
<td>0.50</td>
<td>0.09</td>
</tr>
<tr>
<td>Greece</td>
<td>2.20</td>
<td>3.44</td>
<td>5.94</td>
<td>4.37</td>
<td>2.28</td>
<td>5.51</td>
<td>3.54</td>
<td>-0.21</td>
<td>-3.14</td>
<td>-4.94</td>
<td>-0.71</td>
<td>-6.39</td>
<td>-4.20</td>
</tr>
<tr>
<td>Hungary</td>
<td>3.71</td>
<td>4.51</td>
<td>3.85</td>
<td>8.00</td>
<td>3.96</td>
<td>3.89</td>
<td>0.11</td>
<td>0.89</td>
<td>-6.77</td>
<td>1.31</td>
<td>1.65</td>
<td>-1.73</td>
<td>0.16</td>
</tr>
<tr>
<td>Iceland</td>
<td>3.92</td>
<td>0.14</td>
<td>4.23</td>
<td>7.84</td>
<td>7.23</td>
<td>4.71</td>
<td>5.99</td>
<td>1.19</td>
<td>-6.57</td>
<td>-4.10</td>
<td>2.89</td>
<td>1.64</td>
<td>1.91</td>
</tr>
<tr>
<td>Ireland</td>
<td>4.99</td>
<td>5.42</td>
<td>3.73</td>
<td>4.20</td>
<td>6.08</td>
<td>5.51</td>
<td>4.97</td>
<td>-2.16</td>
<td>-6.38</td>
<td>-1.06</td>
<td>2.17</td>
<td>0.16</td>
<td>0.59</td>
</tr>
<tr>
<td>Italy</td>
<td>1.86</td>
<td>0.45</td>
<td>0.05</td>
<td>1.73</td>
<td>0.93</td>
<td>2.20</td>
<td>1.68</td>
<td>-1.16</td>
<td>-5.49</td>
<td>1.72</td>
<td>0.37</td>
<td>-2.37</td>
<td>-1.78</td>
</tr>
<tr>
<td>Latvia</td>
<td>7.35</td>
<td>7.22</td>
<td>7.60</td>
<td>8.86</td>
<td>10.1</td>
<td>11.2</td>
<td>9.60</td>
<td>-3.28</td>
<td>-17.7</td>
<td>-0.94</td>
<td>5.48</td>
<td>5.58</td>
<td>4.00</td>
</tr>
<tr>
<td>Lithuania</td>
<td>6.70</td>
<td>6.84</td>
<td>10.3</td>
<td>7.37</td>
<td>7.79</td>
<td>7.81</td>
<td>9.80</td>
<td>2.91</td>
<td>-14.8</td>
<td>1.52</td>
<td>5.87</td>
<td>3.62</td>
<td>3.43</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>2.52</td>
<td>4.09</td>
<td>6.17</td>
<td>3.55</td>
<td>4.38</td>
<td>6.59</td>
<td>5.86</td>
<td>-0.74</td>
<td>-3.07</td>
<td>2.89</td>
<td>1.66</td>
<td>0.34</td>
<td>0.55</td>
</tr>
<tr>
<td>Malta</td>
<td>-0.00</td>
<td>2.43</td>
<td>0.72</td>
<td>-0.29</td>
<td>3.59</td>
<td>2.58</td>
<td>4.07</td>
<td>3.88</td>
<td>-2.81</td>
<td>-3.19</td>
<td>1.82</td>
<td>1.04</td>
<td>1.05</td>
</tr>
<tr>
<td>Netherlands</td>
<td>1.93</td>
<td>0.08</td>
<td>0.34</td>
<td>2.24</td>
<td>2.05</td>
<td>3.39</td>
<td>3.92</td>
<td>1.80</td>
<td>-3.67</td>
<td>1.53</td>
<td>0.95</td>
<td>-1.25</td>
<td>1.27</td>
</tr>
<tr>
<td>Norway</td>
<td>1.99</td>
<td>1.50</td>
<td>0.99</td>
<td>3.95</td>
<td>2.60</td>
<td>2.29</td>
<td>2.65</td>
<td>0.03</td>
<td>-1.39</td>
<td>0.24</td>
<td>1.31</td>
<td>3.02</td>
<td>1.57</td>
</tr>
<tr>
<td>Poland</td>
<td>1.21</td>
<td>1.44</td>
<td>3.87</td>
<td>5.35</td>
<td>6.32</td>
<td>6.23</td>
<td>6.79</td>
<td>5.13</td>
<td>1.63</td>
<td>3.88</td>
<td>4.52</td>
<td>1.87</td>
<td>3.32</td>
</tr>
<tr>
<td>Portugal</td>
<td>1.98</td>
<td>0.76</td>
<td>0.91</td>
<td>1.56</td>
<td>0.78</td>
<td>1.49</td>
<td>2.37</td>
<td>-0.10</td>
<td>-2.91</td>
<td>1.94</td>
<td>-1.29</td>
<td>-3.24</td>
<td>-1.76</td>
</tr>
</tbody>
</table>
European, Mediterranean & Middle Eastern Conference on Information Systems 2015 (EMES'15)
June 1st – 2nd 2015, Athens, Greece

<table>
<thead>
<tr>
<th>No.</th>
<th>Company name</th>
<th>Date</th>
<th>No.</th>
<th>Company name</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RheinEnergie</td>
<td>October 14, 2009</td>
<td>17</td>
<td>Aeroflot Company</td>
<td>October 4, 2012</td>
</tr>
<tr>
<td>2</td>
<td>Severn Trent Water</td>
<td>February 11, 2010</td>
<td>18</td>
<td>TATNEFT</td>
<td>June 19, 2013</td>
</tr>
<tr>
<td>3</td>
<td>CEMEX</td>
<td>April 20, 2010</td>
<td>19</td>
<td>Intershop Communications</td>
<td>September 13, 2013</td>
</tr>
<tr>
<td>4</td>
<td>Deutsche Post</td>
<td>July 14, 2010</td>
<td>20</td>
<td>CTAC N.V</td>
<td>September 18, 2013</td>
</tr>
<tr>
<td>5</td>
<td>Bayer AG</td>
<td>August 3, 2010</td>
<td>21</td>
<td>Zurich Insurance Group</td>
<td>November 4, 2013</td>
</tr>
<tr>
<td>6</td>
<td>URENCO Limited</td>
<td>December 13, 2010</td>
<td>22</td>
<td>Royal Dutch Shell</td>
<td>November 13, 2013</td>
</tr>
<tr>
<td>7</td>
<td>HSBC</td>
<td>February 24, 2011</td>
<td>23</td>
<td>ConTech Conveyor Belt Group</td>
<td>November 21, 2013</td>
</tr>
<tr>
<td>8</td>
<td>KAMAZ Group</td>
<td>March 21, 2011</td>
<td>24</td>
<td>PIK Group</td>
<td>November 21, 2013</td>
</tr>
<tr>
<td>10</td>
<td>DTEK</td>
<td>July 18, 2011</td>
<td>26</td>
<td>Schindler Group</td>
<td>February 11, 2014</td>
</tr>
<tr>
<td>12</td>
<td>Nyrstar N.V</td>
<td>December 9, 2011</td>
<td>28</td>
<td>Amadeus IT Group</td>
<td>June 18, 2014</td>
</tr>
</tbody>
</table>

**APPENDIX B**

**Government Debt (Percent of GDP)**

<table>
<thead>
<tr>
<th>Country</th>
<th>Government Debt (Percent of GDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>66.8</td>
</tr>
<tr>
<td>Belgium</td>
<td>68.6</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>68.6</td>
</tr>
<tr>
<td>Croatia</td>
<td>34.7</td>
</tr>
<tr>
<td>Cyprus</td>
<td>62.3</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>23.9</td>
</tr>
<tr>
<td>Denmark</td>
<td>49.6</td>
</tr>
<tr>
<td>Estonia</td>
<td>4.78</td>
</tr>
<tr>
<td>Finland</td>
<td>42.5</td>
</tr>
<tr>
<td>France</td>
<td>56.9</td>
</tr>
<tr>
<td>Germany</td>
<td>59.1</td>
</tr>
<tr>
<td>Greece</td>
<td>103</td>
</tr>
<tr>
<td>Hungary</td>
<td>52.5</td>
</tr>
<tr>
<td>Iceland</td>
<td>45.9</td>
</tr>
<tr>
<td>Ireland</td>
<td>34.4</td>
</tr>
<tr>
<td>Italy</td>
<td>108</td>
</tr>
<tr>
<td>Latvia</td>
<td>14.1</td>
</tr>
<tr>
<td>Lithuania</td>
<td>22.7</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>6.31</td>
</tr>
<tr>
<td>Malta</td>
<td>58.9</td>
</tr>
<tr>
<td>Netherlands</td>
<td>50.7</td>
</tr>
<tr>
<td>Norway</td>
<td>31.9</td>
</tr>
<tr>
<td>Poland</td>
<td>37.6</td>
</tr>
<tr>
<td>Portugal</td>
<td>51.1</td>
</tr>
<tr>
<td>Romania</td>
<td>27.3</td>
</tr>
<tr>
<td>Slovakia</td>
<td>48.9</td>
</tr>
<tr>
<td>Slovenia</td>
<td>29.1</td>
</tr>
<tr>
<td>Spain</td>
<td>55.6</td>
</tr>
<tr>
<td>Sweden</td>
<td>54.7</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>37.3</td>
</tr>
</tbody>
</table>

Note: The government debt for each country is shown as a percentage of GDP.
<table>
<thead>
<tr>
<th></th>
<th>Company Name</th>
<th>Date</th>
<th></th>
<th>Company Name</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>Kerry Group</td>
<td>May 3, 2012</td>
<td>29</td>
<td>Neste Oil</td>
<td>August 13, 2014</td>
</tr>
<tr>
<td>14</td>
<td>TNK BP</td>
<td>May 4, 2012</td>
<td>30</td>
<td>Odjfell Drilling</td>
<td>August 26, 2014</td>
</tr>
<tr>
<td>15</td>
<td>Aareal Bank</td>
<td>June 19, 2012</td>
<td>31</td>
<td>Sympatex Holding GmbH</td>
<td>September 18, 2014</td>
</tr>
<tr>
<td>16</td>
<td>Banco Bilbao Vizcaya</td>
<td>September 12, 2012</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

References


Green Information Systems

Chairs: Habin Lee, Afshin Mansouri, Ahmet Bilici and Levent Demiriz
Table of Contents

**Track – Green Information Systems**


2. [PID-31] Nursen Aydin, Afshin Mansouri and Habin Lee, “INCREASING ENVIRONMENTAL SUSTAINABILITY BY OPTIMIZING VESSEL SPEED IN LINER SHIPPING”

3. [PID-63] Hojeong Yoon RELATIONAL BENEFITS AND CUSTOMER ENGAGEMENT IN SOCIAL MEDIA AND BRAND COMMUNITIES”

4. [PID-64] Saowanit Lekhavat, Nursen Aydin and Habin Lee, “DECISION SUPPORT SYSTEM FOR MULTI-OBJECTIVE SUSTAINABLE MARINE SHIPPING”

5. [PID-69] Oluwakayode Aluko, “FACTORS INFLUENCING ADOPTION OF GREEN SHIPPING PRACTICES AND INFLUENCE ON ORGANISATIONAL PERFORMANCE”
THE USE OF SOCIAL MEDIA FOR IMPROVING ENERGY CONSUMPTION AWARENESS AND EFFICIENCY: AN OVERVIEW OF EXISTING TOOLS

Uthayasankar Sivarajah, Brunel Business School, Brunel University London, UK
sankar.sivarajah@brunel.ac.uk

Gary Fragidis, Cleopa GmbH, Germany
gfragidis@cleopa.de

Michele Lombardi, Alma Mater Studiorum - Universita Di Bologna, Italy
michele.lombardi2@unibo.it

Habin Lee, Brunel Business School, Brunel University London, UK
habin.lee@brunel.ac.uk

Zahir Irani, College of Business, Arts and Social Sciences, Brunel University London, UK
zahir.irani@brunel.ac.uk

Abstract

Raising consumers’ awareness of energy consumption is one of the first steps in encouraging the adoption of energy saving behaviours that result in energy efficiency. Green information systems are becoming recognised as a solution to many environmental problems although information technology (e.g. disposal of IT devices) has also been associated with causing detrimental effects on the environment. Researchers and practitioners have begun to focus on Green ICT but there is little scholarly research on the use of ICT tools such as social media from an energy efficiency context to raise consumer awareness and improve their engagement in tackling environmental issues. Therefore, the aim of this paper is to explore the use of social media and existing tools for the interaction of people on energy saving discussions and for generating awareness and engagement (which thereby leads to energy efficiency behaviour). In this paper the authors provide a state of the art review around the concept of energy awareness, models of consumer engagement, and more importantly the use of social media in the energy efficiency context. This research is based on a desk-based normative review and seeks to provide a better understanding to both scholars and practitioners involved in the use of ICT for driving energy consumer awareness and engagement for energy efficiency.

Keywords: Information and Communications Technology (ICT), Energy Efficiency, Awareness, Consumption, Social media, Green ICT

1 INTRODUCTION

Building sustainable cities that are better connected and well managed are among the most important challenges of our time. In particular, managing the ever-growing energy consumption through energy efficiency processes and technology and the smart re-use of waste and energy is a mammoth task for any government (Hill, 2015). In the energy efficiency context, efficiency gain has to be achieved at all stages of the energy chain, from generation to final consumption. Energy consumers’ awareness is a first step in encouraging the adoption of energy saving behaviours, leading to efficient use of energy by consumers (Piccolo et al., 2014). Energy consumers need to know, first of all, how much energy
they use in their everyday life practices, as well as learn the results of their saving efforts. However, energy consumption is very difficult to observe for the consumer, for various reasons (Jain et al., 2012). For instance, consumers have practically very few possibilities to effectively monitor their energy usage level, especially at the time of consumption. One of their basic options is to review their billing account, which can only take place after the consumption. Nowadays the availability of technologies for monitoring energy consumption, especially ‘smart meters’ can provide real time monitoring and can drill down to the consumption of individual appliances. However, they have only recently been developed and they are still a rather expensive novelty for most people. The lack of real time information on energy consumption for consumers in terms of how much energy they consume in their daily activities seems to dampen their motivation to increase energy savings. Hence, they may be careless, indifferent to energy efficiency solutions and insensitive to environmental friendly policies.

It seems rational to assume that increasing consumers’ awareness about their energy consumption may influence (i.e. reduce) their consumption behaviour. However, a change of the energy consumption behaviour will not come automatically through awareness (Abrahamse et al., 2005). Awareness is often necessary, but not sufficient for energy saving. Behavioural change depends on the motivation of the consumers and their engagement; that is the conscious adoption of energy saving practices, which is developed as a result of understanding the importance of energy efficiency and being motivated to begin and continue saving energy (Jain et al., 2012). Understanding what energy consumers want to know, how they perceive and realize energy efficiency and how they can be motivated to improve their energy saving behaviour is very important for the development of energy efficiency behaviours (Sivarajah et al., 2014a). It is in this context that the use of ICT has a role to play.

Although the disposal of information technology (IT) equipment (e.g. old personal computers, laptops, etc.) and recent surge in data centres are a major cause of environmental concerns, information systems (IS)/ICT have been known to be a key contributor to productivity growth in many countries over the last century (Watson et al., 2008). The use of ICT/IS to achieve environmental objectives is viewed as Green IS (Dedrick, 2010; Watson, 2008). This view includes improving efficiency in industries that are major sources of greenhouse gas (GHG) emissions, such as the transportation, manufacturing, and energy sectors. The Green IS view sees information systems as a possible solution to many environmental problems while Green IT emphasises reducing the environmental impacts of IT production and use (Dedrick, 2010). The use of ICT tools such as social media for improved awareness and engagement of the energy consumers in energy efficiency and energy saving practices is an emerging phenomenon (Opower, 2015; Petkov et al., 2011). Nowadays, people spend a lot of time in social networking sites every day to communicate with friends, get informed, interact with others, for entertainment, etc. (Sivarajah et al., 2014b).

Even though academics and practitioners have begun to focus on Green ICT, there is little scholarly research focusing on the use of ICT tools such as social media from a Green ICT context in tackling environmental issues. Therefore, the aim of this paper is to explore the use of social media for the interaction of people on energy saving discussions and for generating awareness and engagement, which thereby leads to energy efficiency behaviour. This paper provides a state-of-the-art review of the concepts of energy awareness and energy engagement, types of feedback and most importantly the use of ICT (especially social networking and existing social media applications) for improving energy consumption. This research is based on a desk-based normative review and seeks to provide a better understanding to both scholars and practitioners interested in leveraging ICT tools such as social media for driving energy consumer awareness and engagement in order to improve energy efficiency.

The remainder of the paper is organised as follows. Section 2 and 3 presents the research approach and a literature review of the basic concepts of energy awareness and the key aspects of feedback and its effectiveness respectively. This is followed by section 4, in which the concept and the models of consumer engagement are presented. Section 5 highlights the role of social media for improving energy efficiency and the existing applications that are available are reviewed. Finally, section 6 presents the conclusions wherein the summary of the paper is reported along with the contributions and the future directions of this study.
2 RESEARCH APPROACH

This research is based on a desk-based review that has involved identifying, gathering and categorizing literature in the area of energy awareness, consumer consumption and engagement and ICT from an energy efficiency perspective. In this respect, the research approach combined the review and synthesis of literature (Vom Brocke et al., 2009) with secondary analysis of existing social media applications used in the energy efficiency context. Due to the emerging nature of the field of research, a broad literature review was needed to investigate the phenomenon of Green IS and in particular the use of social media in this context. This literature review enabled to scope the defined area of research and identify the void in literature and issues surrounding the role of social media as a Green ICT tool. It is here that this paper seeks to add to the body of literature. The selection of the social media applications was based on the following criteria: a) if and whether the social media application is used from an energy efficiency context, b) the diverse beneficiaries (e.g. citizens, building managers, etc.) and c) different functionalities between the applications. These criteria were used to: i) ensure diversity of the cases and hence the broader applicability of the research outcomes derived from the study, and ii) determine the impact of the different stakeholders and the availability of different functionalities in order to help with raising awareness and tackling energy consumption levels.

3 CONCEPT OF ENERGY AWARENESS

Awareness is related to the knowledge about how much energy is consumed. Without energy awareness, consumers do not know if they need to reduce their energy consumption and how to do it. In the literature there is abundance of evidence for the importance and effectiveness (or lack thereof) of specific energy awareness approaches (Karjalainen, 2011; Horst, 2011). In general, consumers are unaware of their consumption and there is a need for more information, especially concerning the proportional consumption of appliances to be able to make the right choices about their use of energy (Karjalainen, 2011). In particular, consumers need to know how much electricity they use, when they use it, and how much it costs to appropriately impact their monthly bill (Horst, 2011). As a general rule, the more knowledgeable they are, the more interested they are in adopting energy saving practices and in participating in energy saving policies and programmes (IBM, 2011). A number of studies reveal significant reductions in energy consumption as a result of improved energy awareness (Piccolo et al., 2014; Jain et al., 2012).

Energy awareness is the knowledge that users acquire about how and why to save energy by operating devices more efficiently. Awareness is a pre-condition for energy saving and energy efficiency: it cannot ensure that an actual engagement in energy-efficiency behaviours takes place, but it does manage to “make energy visible” (Piccolo et al., 2014). Energy awareness systems provide feedback to the consumer for the energy consumption with the objective to encourage energy efficient behaviour, reduce energy consumption and bring economic and environmental benefits to the consumers and the societies, respectively. Energy awareness systems are frequently referred in the literature as ‘eco-feedback systems’ (Jain et al., 2012).

Energy awareness can be developed by the delivery of information and feedback to the energy consumers. The terms ‘information’ and ‘feedback’ are often used interchangeably, as synonyms in the literature. Typically, ‘information’ is a general term that is used to describe the transmission or communication of some kind of knowledge concerning energy use. On the other hand ‘feedback’ refers to the return of information about the result of an activity that is usually employed for evaluation purposes. In energy efficiency programmes, information has a general awareness and learning purpose, while feedback refers to learning by associating actions/behaviours with the resulting energy consumption. Information and feedback helps consumers use what they learn into practice and eventually develop a routine that leads to lower energy use. So, it seems significant to provide consumers with feedback and the information about the consequences of household actions that involve energy consumption and advice tips for further or smarter energy savings. Accordingly, the interest in this section is mostly focused in feedback from energy consumption, which is discussed below.
3.1 Types of Energy Consumption Feedback

Existing literature distinguishes between two types of energy consumption feedback: direct feedback and indirect feedback (Darby, 2001; Ehrhardt-Martinez et al., 2010).

- **Direct feedback** mechanisms provide energy use information at the time of consumption (or shortly after consumption) and include: real-time feedback, appliance-specific real-time feedback, and simple automation.

- **Indirect feedback** refers basically to energy consumption bills. Bills may provide some kind of analysis on the consumption, such as comparative or historical reports, and possibly advice for energy saving; they can use statistical modelling techniques to estimate (and potentially disaggregate) total household energy usage based on a variety of parameters, such as the household type, appliance information and billing data. Indirect feedback can be given monthly or weekly (or even daily) and is based on real energy use measures gathered by a utility or third party. Delivery to the customer can take place with mailed reports (e.g. monthly bills) or via the web or email (especially for shorter period feedback).

Furthermore, Froehlich et al., (2010) distinguished between low-level feedback, which can provide explicit information about how to change or improve a specific behaviour, and high-level feedback, which is comprehensive and can help improve performance towards a goal or in comparison to others. Some of the main aspects and parameters of energy consumption feedback are presented in Table 1.

<table>
<thead>
<tr>
<th>Aspects of Energy Consumption Feedback</th>
<th>Description</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison</td>
<td>Comparisons provide a frame of reference for energy consumption levels. They can take place with historic standards or normative standards. The assumption is that consumers will be motivated to save energy if they can compare their consumption to their own prior consumption or to the consumption of others.</td>
<td>(Desley et al., 2013; Petkov et al., 2011)</td>
</tr>
<tr>
<td>Measurement units</td>
<td>Feedback on energy use and savings can be provided by the use of different measurement units, such as energy consumption (e.g. in kWh and seldom in kJ), cost (in monetary units, such as euro) and environmental impact (e.g. in carbon dioxide emissions in kg, equivalence of trees, etc.).</td>
<td>(Karjalainen, 2011)</td>
</tr>
<tr>
<td>Frequency of feedback</td>
<td>Energy consumption feedback can be provided yearly, monthly, weekly, daily, hourly basis or in real-time. Preferably, the consumer should be able to change and choose the display of energy consumption for the time period of interest.</td>
<td>(Fischer, 2008)</td>
</tr>
<tr>
<td>Representation of feedback</td>
<td>The method used to demonstrate household energy consumption feedback affects the way that this feedback is understood. The presentation method is important both for the paper-based energy bills and for the in-home displays of energy management systems. These are usually graphical, numeric and textual representations.</td>
<td>(Froehlich et al., 2011)</td>
</tr>
<tr>
<td>Disaggregation of feedback</td>
<td>Disaggregation of energy consumption allows analysing the total/aggregate household energy consumption into appliance or space specific consumption data. The main methods for doing it are the use of smart meters and sensors or the use of statistical methods.</td>
<td>(Darby, 2006; Karjalainen, 2011)</td>
</tr>
</tbody>
</table>

Table 1. Aspects of Energy Consumption Feedback

In Europe, the Directive 2012/27/EU for Energy Efficiency suggests metering systems must provide final customers with information on actual time of use. In addition, final customers must have the
possibility of easy access to complementary information on their own historical consumption. This must include at least the following information:

- Cumulative consumption data corresponding to the intervals for which frequent billing information based on actual consumption has been produced. Such data should be made available for at least the three previous years or the period since the start of the supply contract, if this is shorter.

- Detailed consumption data according to the time of use for any day, week, month and year. Such data should be made available to the final customer for the period of at least 24 months or the period since the start of the supply contract if this is shorter.

The main functions of providing feedback as per Wood and Newborough (2003) are: (a) learning function - energy consumers can learn about the connection between the amount of energy they use and their energy consuming behaviour; (b) habit formation - energy consumers put the information they have learnt into practice and may develop a change in a routine habit; and finally (c) internalisation of behaviour - energy consumers develop new habits and after a while they change their attitudes to suit that new behaviour.

3.2 The Effectiveness of Energy Consumption Feedback

Research has shown that the right feedback can indeed have an influence on energy savings, and that different types of feedback can have different results (Darby, 2006; Fischer, 2008). The research activity in this field has been enormous: some of the research results are highlighted hereafter to reveal the general trends on the effectiveness of energy consumption feedback.

Several meta-analytical studies concluded that consumer awareness systems are an effective tool for reducing energy consumption. A review analysis report developed by Darby (2006) found that direct feedback, alone or in combination with other factors, is the most promising single type, with almost all of the projects involving direct feedback producing savings from 5% to 20%. In addition, direct feedback in conjunction with some form of advice or information gave savings in the region of 10% in average. Indirect feedback does not produce comparably high results in energy savings, although there was also agreement between most of the studies that interest and awareness levels of consumers were raised as a result of supplying informative bills.

Fischer (2008) concluded that typical energy savings range between 5 and 12%; the most effective feedback methods include multiple feedback options (e.g., consumption over various time periods, comparisons, additional information like energy saving tips, etc.), frequently updated feedback, personalization of feedback and interactivity with the consumer (e.g. the consumer should “drill-down” into data), and disaggregation of feedback (e.g. detailed, appliance specific breakdown of the energy usage). The results in other studies favoured also the effectiveness of mixed and comprehensive methodologies. Abrahamse et al., (2005) and Chiang et al., (2012) concluded that direct feedback tends to bring better outcomes, but a greater effect might be achieved when both direct and indirect are combined and when they are provided continuously and more frequently.

Regarding the effectiveness of historic and normative comparisons, most studies favour the historic approach. Historic feedback appears to be readily understandable, relevant, and useful for consumers (Fischer, 2008). The effectiveness of normative comparative feedback, on the contrary, is quite unclear (Desley et al., 2013). Still, several studies that deployed normative comparison resulted in energy savings from the consumers, as consumers would like to compare and discuss their energy performance with others (family, friends, neighbours, etc.). Normative comparisons produces better results when it is used in conjunction with historical comparison, by contextualizing both current and historical consumption in relation to a user’s peers (Desley et al., 2013). However, comparative feedback can lead to counter effects. In ten studies reviewed by Fischer (2008), there was no savings benefit with comparative standard feedback, while in certain studies the results were opposite to what was intended, as those who received the comparative standard feedback consumed more than those who did not. This phenomenon has been referred to as the “boomerang” effect (Schultz et al., 2007),
and it is typically due to the fact that people with lower-than-average consumption feel justified to consume more. The “boomerang” effect can be countered by providing not only descriptive norms, but also including injunctive norms that somehow indicate what is commonly socially acceptable (or unacceptable) within a certain culture (Schultz et al., 2007).

Concerning the effectiveness of real-time feedback, some studies have shown electricity savings in the range of 9–12% (Chiang et al., 2012) and 3–13% (Ehrhardt-Martinez et al., 2010). Buchanan et al., (2015) comment also that only few trials have assessed the contribution of real-time feedback to energy reductions, despite the fact that such trials have the highest relevance for identifying energy savings. On the contrary, an Electric Power Research Institute report (EPRI, 2012) reviewed studies about the use of in-house displays (IHD) and it found no statistically significant impact on energy savings. Among the web-based portals that provide real-time consumption feedback, the Google PowerMeter Application was examined; it was found to produce initial and average savings, but the savings diminished to zero over time. The reason for this is probably that participants begin eagerly, but soon they lose their enthusiasm or even their motivation, because probably they do not find it ‘innovative’ or ‘rewarding’ any more. Further research is needed to understand this behaviour of participants in smart energy programs. This topic is further analysed later, in the section for consumer engagement.

Many research efforts to analyse the effectiveness of energy feedback contain problems and suffer from weaknesses and limitations. First of all, there are big differences in the results between different studies, which come from differences in study design and methodologies (Buchanan et al., 2015). A recent meta-analysis demonstrated that from a methodological perspective, less robust studies without controls yielded higher energy savings, whereas more robust studies that used either a control group and/or also took into consideration either household demographics and/or weather, yielded lower energy savings (Delmas et al., 2013). In addition, most research works consider a short-term experiment period (3 months or less), which makes unclear and difficult to ascertain to what extent these effects are temporary and consumers return to their old practices or they persist in the long term (Verbong et al., 2013). The use of multiple feedback strategies within a single intervention study obscures the results and makes it difficult to pinpoint exactly which aspect of feedback was most effective or if it only worked because a combination of strategies were simultaneously utilized. Buchanan et al., (2015) raises also concerns about self-selection bias and ‘Hawthorne effects’, whereby participants change their behaviour as a result of being involved in an experiment or study.

According to a report of the Electric Power Research Institute (EPRI, 2012), the challenges of energy awareness studies that remain to be answered by future research include the following topics:

- The impact of various demographics on the effect that feedback has on consumers.
- The impact of feedback on consumption reduction.
- The specific actions people are taking as a result of the feedback.
- The relative effectiveness of alternative feedback mechanisms.
- The interaction between feedback, price structures and control technology.

The key assumption underpinning the majority of the studies for energy feedback is that feedback will raise awareness and thereby encourage people to change their behaviour and engage them in energy saving practices. However, besides awareness, consumers need also motivation in order to be engaged in energy efficient behaviours. Therefore, the concept of consumer engagement is discussed in the next section.

4 CONSUMER ENGAGEMENT

Consumer engagement is important for energy efficiency, since energy awareness by itself may be ineffective, or lead to behaviour that cancels the saving in energy usage or even causes higher energy consumption. For instance, the ‘rebound effect’ (Hertwich, 2005; Gavankar and Geyer, 2010) refers to using the new appliance much more than the older one, due to its higher efficiency, and the
‘boomerang effect’ (Schultz et al., 2007) refers to feeling free to increase energy consumption to reach the average levels (which may happen to consumers with the lowest energy usage).

4.1 The concept of energy consumer engagement

Energy consumer engagement is a broad term that involves consumer interest, motivation, comprehension, evaluation, reflection and personal characteristics. It refers to the environmentally responsible behaviour of the energy consumers and the motivation for such behaviour (Buchanan et al., 2015). According to Abrahamse et al., (2005) behaviours related to energy conservation can be divided into two categories: efficiency and curtailment behaviours.

- Efficiency behaviours are typically performed by replacing obsolete appliances with new ones that are more energy efficient or by investing in more energy efficient methods and technologies (e.g. insulation).
- Curtailment refers to reducing energy consumption, for example by turning off lights or appliances when nobody uses them or by lowering thermostat settings. Even though efficiency behaviours are considered to have a higher energy-saving potential compared to curtailment behaviours, the former may suffer from the ‘rebound effect’ (Hertwich, 2005).

Froehlich et al., (2010) distinguishes between ‘rational choice models’, and ‘norm-activation models’ for consumer engagement. Rational choice models are based on a three part linear progression leading from knowledge to concern to environmentally responsible behaviour. The basic assumption here is that human behaviour is regulated by a systematic evaluation process of the expected utility. According to this, people basically act to maximize rewards and minimize costs. On the other hand, norm-activation models are based on the premise that moral norms determine the environmentally responsible behaviour of the consumers, i.e. consumers behave responsibly because this is the right thing to do. In the following section, the existing models of energy consumer engagement are discussed.

4.2 Models of energy consumer engagement

Energy consumer engagement is studied in environmental psychology. Numerous theoretical models have been developed that provide insight for consumers’ motivation and for the design of tools and services for energy efficiency (Pierce and Paulos, 2012). For instance, Fischer (2008) suggests that people must realize, first of all, that there is a problem; then they must realize that their behaviour is relevant to the problem, and they must become conscious about the possibilities to influence their behaviour and its outcomes. Only then will they reflect upon changing their behaviour in order to solve the problem. Horst (2011) explains the steps of the consumer engagement process, from influence and awareness to an understanding and motivation to take action (see Figure 1). Influence refers to developing sensitivity for energy saving, especially with minimal impact to the consumer’s lifestyle, and it is central to developing the interest of the people. Awareness provides feedback on energy consumption and it is the pre-condition for understanding what we consume and how we can reduce it. Only when people develop deep understanding of their energy consumption do they become motivated for taking action to change their current behaviour pattern and adopt energy efficiency practices.
Honebein et al., (2011) agree with Horst (2011) that engagement is driven by action; it requires action by both the energy provider/administrator and by the customer. The energy provider/administrator can only provide the opportunity and the incentives for customers’ action. Lanier and Hampton (2009) suggest the aim of engagement is to hold customer attention, and this is accomplished through experiences that unfold over time; thus, engagement goes beyond awareness and education/understanding alone.

There are three key factors in consumer engagement: habit, technology, and friendship (Honebein et al., 2011). Habit refers to easing customers into engagement through small actions that pave the way down to larger actions. Human beings are wired to develop habits beginning by little, tentative steps, and then playing with something or ‘trying it out’, and then finally by taking up a particular action. People often ‘get used’ to doing things when they repeat them and make them part of their lives. Technology helps individuals make a routine of many complex actions, avoiding trial and error, documenting progress, and easing learning. Friendship leads to customer action and customer engagement down the social road because it is a relationship between people, and because the relationship is valuable and enjoyable for its own sake. This will be better discussed in the following section, where we focus on social comparisons between people using social media and its role in increasing energy consumer awareness and engagement for improving energy savings.

5 DISCUSSION: SOCIAL MEDIA FOR ENERGY EFFICIENCY

Comparisons between individuals or groups can be used to provide motivation to energy consumers. According to Froelich et al., (2010) the effectiveness of social comparisons in the field of environmental psychology provides mixed reports. On the one hand, knowing the behaviour and performance of others may provide motivation to improve yourself. On the other hand, people do not necessarily change their behaviour as a result of comparisons. Very often, when a threshold is reached, further improvement over past performance or the performance of others may not be effective or regarded as critical.

In this section, social comparisons through social networks and media is discussed and in general the role and the importance of social media in increasing energy awareness and engagement and improving energy savings. This is an emerging topic in the literature, which still remains a rather unexplored terrain (Froelich et al., 2010). The use of social media for energy awareness and engagement requires further research in order to better understand the opportunities it provides. After a brief review of the literature, the subsequent section presents and discusses some of the existing social media applications for energy efficiency available in the market.
5.1 The use of social media in energy efficiency context

Social comparisons refer to the concept of a group of households comparing to some other similar group of households (Petkov et al., 2011). Comparing the consumption of one household to that of others is said to elicit accountability and social pressure to understand why consumption levels differ and to stimulate competition and mutual improvement. Social media tools, such as Facebook or Twitter, support social comparison and sharing, as well as public commitments and competition. For instance, a Facebook application named “I am green” (launched in 2007), allowed users to display their environmental attitude as a badge on their own page which provided the opportunity for like-minded people to chat and share their opinions and provided tips and ideas for growing and improving environmental friendly behaviour (Langley and Van den Broek, 2010).

As noted by Langley and Van den Broek (2010), social media can be used to overcome two important barriers to environmentally sustainable behaviour: fatalism and busyness. Fatalism refers to the lack of belief that a sustainability initiative will have a significant impact, which tends to prevent people from joining energy efficiency initiatives. Social media can stimulate energy efficiency efforts by presenting evidence of the goals and achievements and by helping participants to share this information with other participants in their own social networks. The presentation of evidence on energy efficiency goals and achievements can stimulate efforts via a number of mechanisms; (1) feedback from social media can work as a form of persuasion that what people are doing is approved by their social network; (2) people are more likely to participate in initiatives which have already attracted a large number of participants; (3) social media applications make the power of small actions visible, and hence they can augment the impact; and finally (5) sharing information and experiences in social networks can exercise social-psychological pressure and stimulate others to become engaged in similar actions. On the other hand, busyness is the typical state of the majority of the population whereby other priorities and preoccupations do not allow them to act on energy efficiency. Social media can stimulate ‘busy people’ to take action by helping them sharing experiences and by reducing the effort required to act. Social media applications are used as an infrastructure to diffuse individual attitudes and life-styles. Sharing of experiences gives active members recognition of their activities, which can be an incentive to other members to become active.

Langley and Van den Broek (2010) concluded after an empirical study that there is a strong positive correlation between presentation of achievements and sharing with non-participants (i.e. potential participants), as well as between sharing with participants and behavioural change. Another main result is the strong negative correlation between the number of participants and required behavioural change. This shows that initiatives that have many participants tend to require smaller behavioural changes, while initiatives that require large behavioural changes tend to attract a limited number of participants.

5.2 Social media applications for energy efficiency

In this section, we present some of the existing social media applications for energy efficiency, trying to draw out parallels and highlight differences between them.

5.2.1 Opower Facebook Application

Opower, Facebook and the Natural Resources Defense Council (NRDC) launched a social energy app (Opower, 2015). The application combines Opower’s expertise in the use of normative comparisons and games for energy savings with Facebook’s global platform for connecting, sharing and interacting with others. In this way, the application aims to drive behaviour change and enable socially driven engagement. The key functionalities of the application are designed to encourage reductions based on meaningful comparisons with others. In particular, consumers can compare their energy use to similar homes, share and discover energy-saving tips and advice, compete with friends, and participate in team-related energy-reduction challenges. The application's features include a "Friend Rank," which compares the energy usage of users to that of their friends. Users can also invite friends to a "Group" with a specific energy goal. In a group, members compare energy use and communicate in a friendly manner creative energy-saving tactics. Users receive also tips and special offers, which they can share with others.
5.2.2 EnergyWiz

EnergyWiz is a web-based and mobile application that operates through Facebook. It enables users to compare their energy use with their past performance, the usage level of their neighbours, of contacts from social networking sites and of other EnergyWiz users. This application was developed by Petkov et al., (2011) and includes the following main features:

- **The Live Data feature** presents the current energy consumption in the household. The design allows the user to switch between different units of energy consumption (kWh, kg of CO2 and money). To connect to the material impacts of the consumption, they employed an explanatory comparison to assess the amount of consumed energy (i.e., number of trees needed to compensate the generated CO2 emissions).

- **The History feature** allows users to analyse and evaluate their energy consumption over time.

- **The Neighbours feature** allows users to compare themselves with two groups of neighbours - efficient and inefficient neighbours.

- **The Challenge feature** allows the users ‘challenge’ a Facebook friend of theirs on a weeklong energy saving competition. This feature is based on the premise that people are willing to compete in online social networks and compare with real and known people. During the challenge, users are able to post the current score to their Facebook wall. Such public posts can boost the commitment of both parties, leading to discussion among the people who comment on the posts or even make them save energy following the example of their friends.

- **The Ranking feature** enables users compare themselves to similar EnergyWiz users (in terms of household and residence type) according to their consumption in the last seven days. Here, the comparison targets are similar EnergyWiz users, mainly because similarity between Facebook friends is not always given. In addition, users know they are not alone in energy saving and they can enter in discussions about energy saving with people with similar energy attitudes.
5.2.3 Wattsup

Foster et al., (2010a, 2010b) describe ‘Wattsup’ as an application that displays live data from the Wattson energy monitor, allowing users to compare domestic energy consumption on Facebook. The application includes three core interfaces: “My Energy”, “Friends” and “Rankings”.

- The **My Energy** screen shows the energy consumption with a dial visualisation and a seven-day history bar chart.
- The **Friends** screen displays personal energy consumption against selected friends.
- The **Rankings** screen shows a table of highest and lowest energy users of the application.

The rankings table allow users to visualise what their standing is against others, but it does not provide tools for allowing friends to comment on their energy consumption or that of others. A “Comments Board” was added to the user interface later. The results of Wattsup suggest that social networking sites make energy monitoring more enjoyable.
5.2.4 **Green Network**

Lehre and Vasudev (2011) examined how social media networks may facilitate the communication between workplace occupants and building managers. They developed a proprietary social media application (“Green Network”), which contains storyboards and various scenarios.

The user interface includes a newsfeed, profile and an inbox. The newsfeed highlights relevant activities on the network. The profile includes personal information, a billboard, groups that the user belongs to and the apps that users have added to the profile. A user’s persona comprises basic demographic information, likes and interests, and social components such as badges earned on energy-related activities. The billboard displays recent posts, including problems, questions, tips, links, events and surveys. Users can send a post and comment on posts in their newsfeed, ‘like’ posts and comments and answer questions and problems. Each user can create and belong to groups. In general, groups provide a forum for participants with common interests to participate in related discussions and events. A tag cloud visually depicts all the existing groups to give the users an idea of current trends and popular topics.

In addition, the site uses a recommendation system to suggest groups that might be of interest to users. Users can also use various energy-related features and apps, such as “my energy goals” app, which allows users to create, search and sign up for various energy-related goals. Users can rate their performance on their goal on a weekly basis, get “green points,” and perform normative comparisons of their goals with others in the network. Other apps are named “my energy use” and “group energy use” and they track individual and group energy use by pulling data from external devices, such as meters and building energy management systems, and render the results using interactive visualizations. Users have the option of customizing energy graphs by choosing the type of visualization, parameters to be graphed (e.g. cost, energy per person, comparisons to averages and other people, etc.) and the temporal granularity of the display.

5.2.5 **StepGreen.org**

Mankoff et al., (2010) describe StepGreen.org, a site intended to motivate people to make energy-reducing changes to their behaviours. StepGreen.org combines features such as committing to and reporting on actions and can serve information to a person’s social network profile page. The site emphasizes financial savings and CO2 savings.

Users are invited by a friend or after seeing the applet on a friend’s profile page. After the initial account creation, the user may install the StepGreen.org social network plugin on a preferred site. Any friend who can view the user’s profile page can see the visualization and suggested actions. Thus, the visualization serves as a combination of public commitment and reminder. The reminder function is facilitated by the fact that many people visit their favourite social networking sites once a day or more. To encourage participation, suggested actions include many things that participants would like to do. StepGreen.org combines committing to actions and reporting on whether commitments are fulfilled in the “real world.”

StepGreen.org uses a variety of mechanisms to advertise actions, including suggestions sent directly to a person’s social network page, a table of actions, and a tag cloud. Users can explore actions on the site in one of three ways. An action browser allows users to leaf through all available actions on the site. For each action, the browser displays the action’s name, projected per-year dollar savings, projected per-year CO2 savings, category, and overall popularity. Users can sort actions by any of the properties being displayed. The site prompts users to explore actions via suggested actions, as well as popular actions, which lists the names of the most popular actions in a tag cloud.
6 CONCLUDING COMMENTS AND CONTRIBUTIONS OF THE STUDY

Green ICT emerges as a new research domain that emphasises on the cross-functional role of ICT in addressing the significant issue of improving energy efficiency and environmental sustainability. This paper focuses on the opportunities of Green ICT for encouraging energy efficiency practices at the household level, by the energy consumer, as without their engagement the results of any energy efficiency programme will be ineffective. In particular, this research has examined the potential role of social media as a Green ICT tool in the effort for energy awareness and engagement of the consumer.

This paper highlighted that energy awareness is one of the initial steps towards energy efficiency behaviours, since energy consumers need to know how much energy they use in their everyday life practices, as well as to learn the results of their saving efforts. Without precise and detailed feedback on their energy consumption, consumers are not able to better understand in the first place how much energy they consume in their daily activities and subsequently they do not receive motivation for energy savings. Furthermore, this paper reveals that feedback can indeed have an influence on energy savings, and that different types of feedback can have different results. However, many challenges remain to be answered by future research, especially with regard to the motivation and the long-term engagement of the consumers in energy saving behaviours. This paper discussed that creating awareness of energy consumption is essential but not enough for achieving energy savings. This is because behavioural change depends on the motivation of the consumers and their engagement; by engagement, we refer to the conscious adoption of energy saving practices, which is developed as a result of understanding the importance of energy efficiency and being motivated to begin and continue saving energy.

This research concludes that ICT tools such as social media can potentially support both energy consumer awareness and engagement through social interaction for sharing information and learning from others consumers and their good practices. Social interaction is notably beneficial for energy awareness and engagement and it can support the change of behaviours through the adoption of more sustainable life practices. Social media plays a decisive role in understanding what energy consumers want to know, how they perceive and realise energy efficiency and how they can be motivated to improve their energy saving behaviour. Social media applications are able to help advance the social discourse and sensitisation whilst proliferating good practices amongst consumers. These conclusions have been reached by reviewing the main techniques employed in existing social media applications for improving energy awareness and promote engagement. The authors of this paper believe that the availability of such a concise and yet effective overview in itself is a valuable contribution to both practitioners and scholars.

In terms of contribution to knowledge, this study adds to the Green Information System and ICT literature on how ICT tools such as social media can be used to raise energy consumer awareness and engagement to foster energy efficiency practices. From a practical perspective, the review conducted in this paper seeks to provide a better understanding of the energy consumer awareness and engagement methods, and the use of ICT tools in methods of developing energy consumer awareness and engagement. This paper is part of research done in the context of the EU DAREED project. The project aims at the development of a web platform to improve the energy efficiency of urban districts, by providing tools to support the activities of the main stakeholders: municipalities, energy providers, and citizens. This review will serve as a guide for the design of platform components for increasing the awareness and promote the engagement of the district citizens: a clear overview of the state-of-the-art is crucial for directing future efforts and obtaining the best results.

Acknowledgements

This work evolved in the context of the project DAREED (Decision support Advisor for innovative business models and useR engagement for smart Energy Efficient Districts), www.dareed.eu, a project co-funded by the EC within FP7, Grant agreement no: 609082. The authors express their gratitude and acknowledgement to the contributions of the DAREED project partners. The content of this article
represents the view of the authors, respectively. The European Commission cannot be made liable for any content.

References


INCREASING ENVIRONMENTAL SUSTAINABILITY BY OPTIMIZING VESSEL SPEED IN LINER SHIPPING

Nurşen Aydın, Brunel Business School, Brunel University London, nursen.aydin@brunel.ac.uk
Afshin Mansouri, Brunel Business School, Brunel University London, afshin.mansouri@brunel.ac.uk
Habin Lee, Brunel Business School, Brunel University London, habin.lee@brunel.ac.uk

Abstract

Due to the increasing concern on the environmental impact of the Maritime operations, shipping companies have focused on regulating their fuel consumption. Sailing speed is directly related to the fuel consumption and green house gas emissions. We consider the speed optimization problem in the liner shipping. In this problem, vessel tracks a fixed route to transport cargo and the handling time at ports follows a stochastic process. The objective is to find an optimal speed policy between ports so that the total voyage cost is minimized. We approximate the port times and formulate the problem as nonlinear programming model. The proposed model is tested against an existing shipping schedule. Computational experiments reveal that the proposed method performs well under a variety of conditions.

Keywords: Speed optimization, Fuel emission, Liner Shipping.

1 INTRODUCTION

In recent years, economic and environmental concerns have brought a new perspective to maritime operations. According to the International Maritime Organization (IMO), the international shipping industry is responsible for approximately 2.2% of global greenhouse gas emissions in 2012 (Third IMO GHG Study, 2014). CO₂ emissions from maritime transport represent a significant part of total global greenhouse gas emissions. It is known that CO₂ emission is directly proportional to fuel consumption which significantly depends on the speed of the vessels (Third IMO GHG Study, 2014). Therefore, many operational strategies focus on reducing vessel speeds. Although sailing with the slowest speed is favourable with respect to the fuel cost and CO₂ emissions, it may not be always feasible due to the time constraints imposed by the ports. Mansouri et al. (2015) provide an extensive review on the current studies in environmental sustainability in Maritime shipping.

The recent studies in maritime literature have focused on environmental impacts of the ships. Christiansen et al. (2013) examine the ship routing and scheduling problems in liner, tramp and industrial shipping. In liner shipping, port rotation is fixed and vessels follow the planned schedule with weekly frequency. On the other hand, vessels do not have to follow a fixed route in tramp and industrial shipping, and the ship routing and scheduling decisions are similar. Although the overall objective function is different for each shipping area, the common objective is to decrease the operational costs. Ronen (2011) points out the importance of reducing vessel speed on operating cost. He works on the speed optimization problem by considering the service frequency and the required number of vessels. Fagerholt et al. (2010) and Hvattum et al. (2013) work on the speed decisions in fixed shipping routes with port time windows. They assume that vessel always arrives within the time window of each port. Fagerholt et al. (2010) discretize the arrival times and solve the problem by using shortest path algorithm. Hvattum et al. (2013) develop an exact solution algorithm for the deterministic problem. Wang and Meng (2012) work on the speed optimization problem with transhipment and container routing. They formulate the problem as mixed-integer nonlinear model and propose outer-approximation algorithm to obtain approximate solution. Norstad et al. (2011) incorporate speed decision in the tramp ship routing and scheduling problem and propose a local
search method. They first develop a solution algorithm for speed optimization problem with fixed route. Then, they utilize this algorithm to generate initial solution for the proposed local search method. Zhang et al. (2014) extend the work of Fagerholt et al. (2010) and Norstad et al. (2011), and study on the optimality properties. We refer reader to Psaraftis and Kontovas (2013) for a comprehensive review of the studies concentrated on speed decisions.

In this paper, we study speed optimization problem in liner shipping. Our study is closely related to the work of Fagerholt et al. (2010). We focus on minimizing fuel consumption by optimizing the speed of the vessel along a fixed route. Different than the model of Fagerholt et al. (2010), we do not restrict the vessel to arrive within the time window. In real-life applications, vessels may arrive outside of the time window. Therefore, in our problem formulation, we consider early arrivals and delays while constructing the optimal schedule. We also focus on minimizing the time spent at ports. To sum up, our objective is to find an optimal speed policy for a vessel along a fixed route by considering time window violation at each port.

The rest of the paper is organized as follows. In Section 2, we formulate the speed optimization problem in liner shipping route. We discuss our solution approach in the same section. In Section 3, we present computational experiments on a real liner shipping route. Section 4 provides the concluding remarks and future research directions.

2 Model Formulation

We consider a liner shipping company which provides shipping services over a number of ports denoted by set $N = \{0, 1, ..., n\}$. Port 0 shows the stating node of the network. A vessel can visit port $i$ within its time window. If it arrives earlier than the available slot, it has to wait until port service is open. If it arrives later than the time slot, it will lose a fraction of its overall service level. Service time of a vessel in each port is a stochastic variable. We assume that service time follows uniform distribution. We have to decide the speed between ports in order to minimize total fuel consumption and maximize service level. The general parameters used in the problem formulation are summarized in Table 1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>Total number of ports</td>
</tr>
<tr>
<td>$S_i$</td>
<td>Random service time in port $i$ such that $S_i \in U[l_i, u_i]$ and $S_0 = 0$</td>
</tr>
<tr>
<td>$[\alpha_i, \beta_i]$</td>
<td>earliest and latest planned arrival times at port $i$,</td>
</tr>
<tr>
<td>$v_i$</td>
<td>Average speed between ports $(i - 1)$ and $i$, and it is limited by $[v_{\min}, v_{\max}]$</td>
</tr>
<tr>
<td>$d_i$</td>
<td>distance between port $(i - 1)$-th and $i$-th port-of-call</td>
</tr>
<tr>
<td>$t_i^a$</td>
<td>Arrival time of vessel at port $i$, $(t_0^a = 0)$</td>
</tr>
<tr>
<td>$t_i^d$</td>
<td>Departure time of vessel at port $i$, $(t_0^d = 0)$</td>
</tr>
<tr>
<td>$\varphi$</td>
<td>Fuel consumption per hour during waiting and service at each port</td>
</tr>
<tr>
<td>$\theta$</td>
<td>Penalty for delay per hour</td>
</tr>
<tr>
<td>$f_p$</td>
<td>Price of fuel per ton consumed at ports</td>
</tr>
<tr>
<td>$f_s$</td>
<td>Price of fuel per ton consumed during sailing</td>
</tr>
</tbody>
</table>

Table 1. The general parameters used for the problem formulation

In the literature, quadratic function of sailing speed is generally used to compute fuel consumption of a vessel. The vessel can sail between the upper and lower speed limits and we assume that the speed of a vessel is constant in each leg, i.e. between two consecutive ports. We use the empirical formula of Fagerholt et al. (2010) to calculate fuel consumption rate per nautical mile at sailing speed $v_i$. The fuel consumption function is as follows:

$$g(v_i) = 0.0036v_i^2 - 0.1015v_i + 0.8848$$  \hspace{1cm} (1)

Then, the fuel consumption between ports $(i - 1)$ and $i$ is given by $d_i g(v_i)$. We assume that the vessel consumes a fixed amount of fuel per hour during waiting and service time at each port. Given
the prices of fuel consumed during sailing and berthing, the total fuel consumption cost is computed as,

$$\sum_{i=1}^{n} \left( f_i d_i g(v_i) + f_p \varphi(t_i^d - t_i^a) \right)$$  \hspace{1cm} (2)

Arriving later than the given time window will result in missing the available time slot. Finding a new slot for berthing would be difficult and costly. To maximize the service level and avoid delays, we penalize the vessel for each hour of being late. Then, the overall cost function is given by

$$\sum_{i=1}^{n} \left( f_i d_i g(v_i) + f_p \varphi(t_i^d - t_i^a) \right) + \sum_{i=1}^{n} \theta[t_i^a - \beta_i]^+$$  \hspace{1cm} (3)

where $[t_i^a - \beta_i]^+ = \max\{t_i^a - \beta_i, 0\}$. Given the speed decision $v_i$ and service time $S_i$ at port $i$, the states of the system at the following ports are defined by the following system dynamics equations;

$$t_i^n = t_{i-1}^n + \frac{d_i}{v_i},$$  \hspace{1cm} \text{subject to} \quad t_i^n = t_{i-1}^n + \frac{d_i}{v_i}, \quad i = 1, \ldots, n$$  \hspace{1cm} (5)

$$t_i^n = t_i + E[S_i],$$  \hspace{1cm} \text{subject to} \quad t_i^n = t_i + E[S_i], \quad i = 1, \ldots, n$$  \hspace{1cm} (6)

$$t_i \geq \alpha_i,$$  \hspace{1cm} \text{subject to} \quad t_i \geq \alpha_i, \quad i = 1, \ldots, n$$  \hspace{1cm} (7)

$$t_i \geq t_i^a,$$  \hspace{1cm} \text{subject to} \quad t_i \geq t_i^a, \quad i = 1, \ldots, n$$  \hspace{1cm} (8)

$$v_{\min} \leq v_i \leq v_{\max},$$  \hspace{1cm} \text{subject to} \quad v_{\min} \leq v_i \leq v_{\max}, \quad i = 1, \ldots, n$$  \hspace{1cm} (9)

where $t_0^a = t_0^d = 0$. Constraints (5) and (6) correspond to the system dynamics equations. Constraints (7) and (8) ensure that the vessel starts service after it arrives and the time window starts. Constraints (9) guarantee that the speed of the vessel is within the lower and upper limits. This model provides a scheduling and speed policy to minimize total fuel consumption. It does not capture the temporal dynamics of the problem due to the approximation of stochastic variables. However, we can improve its performance by refining the speed decision variable during the practical application. Specifically, given the state variable $t_i^a$ at port $i$ (after vessel arrives to port $i$), we update the constraints (5) and (8) with realized $t_i^a$ and solve the deterministic problem accordingly.

### 3 Preliminary Results

In this section, we present our preliminary results. We conduct numerical experiments to evaluate the effects of speed decisions on the total fuel consumption. We solve the model (5)-(10) with a non-linear programming solver in MATLAB using the interior point algorithm.

We conduct experiments by using data from a real liner shipping route of a major European shipping company. This data includes the distances between ports, vessel arrival and berthing times, average port service time and the average vessel speed between ports. Table 2 presents an existing schedule of this shipping company. Lacking the actual service time distribution, we assume that service time at port $i \in N$ follows uniform distribution with mean $\lambda_i$. For a fair comparison, we set $\lambda_i$ to the service times given in Table 2. Since we use the fuel consumption function given in Fagerholt et al. (2010), we assume that sailing speed ranges from $v_{\min} = 14$ knots to $v_{\max} = 20$ knots. Our cost function
includes three main parts; fuel consumption cost, port time cost (waiting and service) and delay penalty. The first two cost types are directly related to the fuel cost. Vessels consume two types of fuels during route trip. According to the data obtained from the shipping company, the fuel consumed at ports is more expensive than the one consumed during sailing. In our experiments, we set the unit price of fuel consumed in port twice the price of the one used during sailing. We fix the fuel cost to $250/ton and $500/ton during sailing and berthing, respectively. The delay penalty, on the other hand, is related to the service level and it is difficult to estimate. Another important parameter affects the speed policy is time window. In real time applications, the schedule of the vessel is planned with respect to the available time slots of the ports. According to the data of the shipping company, each port specifies an available time slot \((t_i^e)\) and provides a time window around that slot. We define \(\omega\) to compute the time window in our numerical example. In other words, the earliest and latest possible service times of port \(i\) are given by \(\alpha_i = t_i^e\) and \(\beta_i = t_i^e + \omega\).

Although the route schedule (arrival times and vessel speed) is determined before the vessel departs from the first port, it can be changed during the journey. Weather conditions, disruptions and congestion at ports are the main reasons of the deviation from the planned schedule. According to the existing schedule in Table 2, the vessel generally arrived much earlier than the berthing time. This behaviour can be attributed to the impact of uncertainty in ports and weather conditions. Being earlier than the planned schedule resulted in long waiting times at ports. As it is seen in Table 2, the vessel waited more than two days in port P.6 and it sailed at 15.2 knots on average between P.5 and P.6 instead of sailing at the minimum speed. Since sailing with high speed and arriving much earlier than the berthing time only increases the costs, we conjecture that there is a miscommunication between the vessel and the port during this journey.

To reveal the effects of insufficient information sharing, we test our proposed model with two scenarios. In the first scenario, we set the starting time of the time window \((t_i^e)\) to the berthing times given in Table 2. The arrival and berthing times of the existing schedule are given by the following vectors.

\[
t^e = \{0, 7, 66, 105, 239, 385.5, 456, 574, 661, 735, 813.5, 1062.5\},
\]

\[
t^* = \{0, 8, 88, 114, 256, 392, 511.5, 574, 700, 762, 817, 1062.5\}.
\]

<table>
<thead>
<tr>
<th>Port</th>
<th>Distance</th>
<th>Arrival</th>
<th>Berthing</th>
<th>Departure</th>
<th>Service Time (h)</th>
<th>Avg. Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>P.0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>12/04/14 22:00</td>
<td>-</td>
<td>16.6</td>
</tr>
<tr>
<td>P.1</td>
<td>116</td>
<td>12/05/14 05:00</td>
<td>12/05/14 06:00</td>
<td>12/06/14 12:15</td>
<td>30</td>
<td>15.4</td>
</tr>
<tr>
<td>P.2</td>
<td>409</td>
<td>12/07/14 16:00</td>
<td>12/08/14 14:00</td>
<td>12/09/14 06:00</td>
<td>15</td>
<td>14.8</td>
</tr>
<tr>
<td>P.3</td>
<td>15</td>
<td>12/09/14 07:00</td>
<td>12/09/14 16:00</td>
<td>12/10/14 14:30</td>
<td>22</td>
<td>15.0</td>
</tr>
<tr>
<td>P.4</td>
<td>1546</td>
<td>12/14/14 22:00</td>
<td>12/15/14 14:00</td>
<td>12/16/14 07:30</td>
<td>18</td>
<td>15.1</td>
</tr>
<tr>
<td>P.5</td>
<td>1681</td>
<td>12/20/14 23:30</td>
<td>12/21/14 06:00</td>
<td>12/22/14 10:30</td>
<td>27</td>
<td>15.2</td>
</tr>
<tr>
<td>P.6</td>
<td>537</td>
<td>12/23/14 22:00</td>
<td>12/26/14 05:30</td>
<td>12/28/14 10:20</td>
<td>48</td>
<td>14.9</td>
</tr>
<tr>
<td>P.7</td>
<td>142</td>
<td>12/28/14 20:00</td>
<td>12/28/14 20:00</td>
<td>12/30/14 17:45</td>
<td>29</td>
<td>15.1</td>
</tr>
<tr>
<td>P.8</td>
<td>617</td>
<td>01/01/15 11:00</td>
<td>01/03/15 02:00</td>
<td>01/04/15 03:30</td>
<td>31</td>
<td>15.1</td>
</tr>
<tr>
<td>P.9</td>
<td>143</td>
<td>01/04/15 13:00</td>
<td>01/05/15 16:00</td>
<td>01/07/15 07:30</td>
<td>35</td>
<td>15.0</td>
</tr>
<tr>
<td>P.10</td>
<td>180</td>
<td>01/07/15 19:30</td>
<td>01/07/15 23:00</td>
<td>01/08/15 20:30</td>
<td>19</td>
<td>14.2</td>
</tr>
<tr>
<td>P.11</td>
<td>3200</td>
<td>01/18/15 04:30</td>
<td>01/18/15 04:30</td>
<td>01/18/15 13:30</td>
<td>8</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2. An existing schedule of the shipping company
We compare the performance of the proposed model against the existing schedule of the shipping company. In this experiment, we set the width of the time window to 2 hours ($\alpha_i = t^*_i$ and $\beta_i = \alpha_i + 2$, $i \in N$). According to the data of the shipping company, a vessel consumes approximately 0.1 ton fuel per hour. Therefore, we set the waiting costs to $50/hour. Since delays may result in poor service, we set the delay penalty $1000/hour to prevent lateness. Table 3 and 4 summarize our findings. First row of Table 3 presents the berthing times at all ports. The next two rows present the arrival times of the shipping company ($t^e$) and the optimal arrival time of the optimization model ($t^*_i$). The remaining rows show the resulting waiting times of these schedules. The first observation we have is that the waiting time in P.6 is extremely high although it is lower than the one resulted from existing schedule. Early arrivals like this case occur due to the given time windows. The distance between some consecutive ports can be short. If the consecutive time windows are not tight, the vessel always arrives early to the next port even if it sails at minimum speed. In this scenario, time windows are loose. Therefore, the optimum speed of deterministic model is always around 14.0 knots. Although the vessel sails with its minimum speed, it arrives ports 2, 6, and 9 much earlier than the available berthing time. When we compare it with the resulting waiting times of existing schedule, we observe that deterministic policy performs better. Table 4 presents the fuel consumption costs during waiting and sailing. When we look into the total costs, we obtain that the percentage gap for waiting cost is 20.74%. This shows that, deterministic policy provides a more fuel efficient schedule.

<table>
<thead>
<tr>
<th>Ports</th>
<th>P.1</th>
<th>P.2</th>
<th>P.3</th>
<th>P.4</th>
<th>P.5</th>
<th>P.6</th>
<th>P.7</th>
<th>P.8</th>
<th>P.9</th>
<th>P.10</th>
<th>P.11</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t^*$</td>
<td>8</td>
<td>88</td>
<td>114</td>
<td>256</td>
<td>392</td>
<td>511.5</td>
<td>574</td>
<td>700</td>
<td>762</td>
<td>817</td>
<td>1062.5</td>
</tr>
<tr>
<td>$t^e$</td>
<td>7</td>
<td>66</td>
<td>105</td>
<td>239</td>
<td>385.5</td>
<td>456</td>
<td>574</td>
<td>661</td>
<td>735</td>
<td>813.5</td>
<td>1062.5</td>
</tr>
<tr>
<td>$t^*_i$</td>
<td>8.3</td>
<td>70.3</td>
<td>105.2</td>
<td>247.4</td>
<td>393.6</td>
<td>461.5</td>
<td>574.6</td>
<td>665.5</td>
<td>736.4</td>
<td>814.5</td>
<td>1064.5</td>
</tr>
<tr>
<td>Waiting time ($t^e_i$)</td>
<td>1</td>
<td>22</td>
<td>9</td>
<td>17</td>
<td>6.5</td>
<td>55.5</td>
<td>-</td>
<td>39</td>
<td>27</td>
<td>3.5</td>
<td>-</td>
</tr>
<tr>
<td>Waiting time ($t^*_i$)</td>
<td>0.3</td>
<td>17.7</td>
<td>8.8</td>
<td>8.6</td>
<td>-</td>
<td>49.5</td>
<td>-</td>
<td>34.5</td>
<td>25.6</td>
<td>2.5</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 3. The optimal arrival times for a given port time

<table>
<thead>
<tr>
<th>Policies</th>
<th>Fuel Consumption (Waiting)</th>
<th>Fuel Consumption (Sailing)</th>
<th>Total Fuel Consumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Existing Schedule</td>
<td>$9,025.0$</td>
<td>$369,003.35$</td>
<td>$378,028.35$</td>
</tr>
<tr>
<td>Optimization Model</td>
<td>$7,475.0$</td>
<td>$363,592.61$</td>
<td>$371,067.61$</td>
</tr>
<tr>
<td>Potential Saving</td>
<td>20.74%</td>
<td>1.49%</td>
<td>1.88%</td>
</tr>
</tbody>
</table>

Table 4. The fuel consumption cost

In the second scenario, we set the starting time of the time window ($t^*_i$) to the arrival times given in Table 2 and we assume that handling service also starts at the arrival time. In other words, the existing schedule does not incur any waiting cost. In this case, the arrival and berthing times of the existing schedule are given by the following vectors.

$$t^e = t^* = \{0, 7, 66, 105, 239, 385.5, 456, 574, 661, 735, 813.5, 1062.5\}$$

For a fair comparison with the previous scenario, we set the width of the time window to $\omega = 2$. Table 5 presents the schedule for this scenario. When we look into the arrival times, we observe that the optimal speed policy of optimization model always prevents early arrivals and delays. Moreover, comparing optimal speed values in Table 5 with those in Table 2, we note that the sailing speeds of the proposed model are always lower than the ones in the existing schedule. Table 6 presents the resulting fuel consumption costs. In this case, the potential saving is around 0.85%. Since the starting time of
the service is earlier in this experiment, the optimum speed values are higher than 14.0 knots (optimum speed policy of the previous scenario). As a result, total fuel consumption increases.

<table>
<thead>
<tr>
<th>Port Index</th>
<th>P.1</th>
<th>P.2</th>
<th>P.3</th>
<th>P.4</th>
<th>P.5</th>
<th>P.6</th>
<th>P.7</th>
<th>P.8</th>
<th>P.9</th>
<th>P.10</th>
<th>P.11</th>
</tr>
</thead>
<tbody>
<tr>
<td>(t^*)</td>
<td>7</td>
<td>66</td>
<td>105</td>
<td>239</td>
<td>385.5</td>
<td>456</td>
<td>574</td>
<td>661</td>
<td>735</td>
<td>813.5</td>
<td>1062.5</td>
</tr>
<tr>
<td>(t_i^p)</td>
<td>7.8</td>
<td>66.4</td>
<td>105.4</td>
<td>240.5</td>
<td>386.8</td>
<td>458</td>
<td>575.7</td>
<td>663</td>
<td>736.9</td>
<td>815.5</td>
<td>1064.5</td>
</tr>
</tbody>
</table>

Table 5. The optimal arrival times for scenario 2

<table>
<thead>
<tr>
<th>Policies</th>
<th>Fuel Consumption (Waiting)</th>
<th>Fuel Consumption (Sailing)</th>
<th>Total Fuel Consumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Existing Schedule</td>
<td>-</td>
<td>$369,003.35</td>
<td>$369,003.35</td>
</tr>
<tr>
<td>Optimization Model</td>
<td>-</td>
<td>$365,924.34</td>
<td>$365,924.34</td>
</tr>
<tr>
<td>Potential Saving</td>
<td>-</td>
<td>0.85%</td>
<td>0.85%</td>
</tr>
</tbody>
</table>

Table 6. The fuel consumption cost for scenario 2

4 Conclusion and Future Research

In this paper, we addressed speed optimization problem in liner shipping with uncertain port times. By considering the waiting and the delay costs in ports, we formulated the problem as a constrained nonlinear model which can be solved by nonlinear programming solvers. We tested the performance of our model by implementing on a real-life case from a liner shipping company. We observed that the existing schedule results in long waiting times at ports. These long waiting times can be resulted from insufficient information sharing between the vessel and the ports. Even if the time windows are not tight, the vessel preferred to sail at higher speeds to avoid delays due to the uncertainties along the journey. On the other hand, the optimal policy of the proposed model decreases the waiting cost by around 20%. This result reveals the importance of communication during route trip. Many shipping companies use decision support systems (DSS) to manage shipment operations. The proposed speed optimization model can be integrated to DSS to decide and update sailing speed dynamically. For instance, in case of a deviation from planned schedule, the sailing speed can be updated by using the proposed model.

This is an ongoing research. An extension of this work could look into tackling the uncertainty in the system. The proposed model assumes that the stochastic port times take on their expected values. This may be unreasonable in situations where the service times at ports are highly variable. A promising direction to pursue for future research would be to develop a dynamic model for speed decisions. Dynamic model considers all possible combinations of service time and hence, it can cope with uncertainty better than the deterministic models. As a future work, we would work on the dynamic programming formulation.

Acknowledgement

This research was supported in part by EU FP7 project MINI-CHIP (Minimising Carbon Footprint in Maritime Shipping) under grant number PIAP-GA-2013-611693.
References


RELATIONAL BENEFITS AND CUSTOMER ENGAGEMENT IN SOCIAL MEDIA AND BRAND COMMUNITIES

Hojeong Yoon, Korea Local Information Research & Development Institute, Korea, email: hojeong512@gmail.com

Abstract

Following the explosive growth of social media users, companies have used social media as a business activity platform. For effective and efficient social media management, companies need to understand user intention in using social media. This study focused on relational benefits as user intention in using social media brand communities and analysed the relationship between relational benefits and customer engagement considering industry characteristics. A social media monitoring tool and a web-based survey platform were used to collect related data. The expected benefit types of potential brand community users and the average engagement rate in brand communities varied by industry. Finally, the study identified the effective relational benefits for increasing customer engagement. Based on the results, the study highlighted the need for the appropriate management of social media brand communities by identifying user intentions and understanding industry characteristics.

Keywords: Social Media, Brand Community, Engagement, Relational Benefit.

1 INTRODUCTION

Social media has become an everyday thing for modern people, and information diffusion via social media is not limited to regions but influences the entire world. With the development of mobile devices, social media users can now generate and share information anytime and anywhere on a real-time basis, thereby satisfying their desire for ‘building relationships’ (Hong & Han, 2011). The increasing number of social media users has led firms to set up brand communities where they can meet with their customers on social media, such as Facebook. Firms may convey information to customers via brand communities on social media while also obtaining valuable data on how customers react to such information, what they say about the organization, who support the brands, and when there is a crisis. To this end, firms must understand and measure the reactions of people within the community. A typical indicator that can measure the above is ‘engagement’. Engagement represents how much users are engaged in the community and how well a firm makes use of social media. Since this is related to customer loyalty or financial performance as well, this indicator must be carefully examined by firms. To efficiently operate communities to increase customer engagement, firms must first be aware of users’ intentions in using brand communities. A relationship between users and a brand forms in a brand community because both users and the firm gain certain benefits (Gwinner et al., 1998). This study opines that users’ intention in using brand communities on social media is to obtain relational benefits, and determines what relational benefits are expected by potential users of brand communities and what relational benefits affect customer engagement in brand communities by industry. The purpose of this study is to develop baseline data and strategies that facilitate the effective and efficient management of firm-hosted brand communities, by empirically analysing relational benefits and engagement.
2 THEORETICAL BACKGROUND

2.1 Brand Communities and Customer Engagement

Brand communities act as an important platform used by firms to induce customer engagement behaviours (Brodie et al., 2011; Dholakia et al., 2004; Kane et al., 2009; McAlexander et al., 2002). With the increasing number of online users, more and more firms are implementing online communities for commercial purposes to build customer relationships and strengthen their brands (Wiertz and de Ruyter, 2007). Brand communities in virtual space have been formed mostly by consumers, but firm-hosted communities have also emerged as a form of brand management strategy (Arnone et al., 2010). Community users can be engaged in various activities within the community (Nambisan and Baron, 2009). Shang et al. (2006) found that even without active participation in the community, engagement, such as merely reading other people’s comments or lurking, may increase customer loyalty. Since more and more people spend time in online communities along with the development of technology, it is necessary to investigate customer engagement in brand communities (Kaplan and Haenlein, 2010).

Engagement is an indicator that measures how well the fans or followers of a certain brand interact with social media content, and it is one of the most important elements that must be measured in the social sciences. Customer engagement can be defined as “a customer’s behavioural manifestations that have a brand or firm focus, beyond purchase” (van Doorn et al., 2010). Engagement generates interaction between customers and the brand and enhances the brand experience value of consumers (Brodie et al., 2011). Engagement is also referred to as the highest state of loyalty (Bowden, 2009), and such behavioural manifestations include all behaviours beyond a high level of loyalty (Libai, 2011). Van Doorn et al. (2010) argued that customer engagement, among other customer behaviours, is effective in reinforcing customers’ relationship with the brand. Customer engagement shows a direct and positive relationship with brand relations, such as satisfaction, reliability, emotional commitment, and loyalty (Brodie et al., 2011). Loyalty is considered a fundamental reason for engagement in brand communities (McAlexander et al., 2002). Muniz and O’Guinn (2001) suggested that customers’ engagement in a brand community increases loyalty and strengthens relationships with the firm, and Brodie et al. (2011) confirmed that greater engagement indicates greater customer satisfaction and loyalty. Social media is a place where customers can reveal their engagement behaviours to firms, and this platform has a feature that is suitable for the development of customer relationships (Kane et al., 2009).

There may be various items to define the indicators that measure social media engagement, depending on the evaluation criteria or perspective of those who interpret the results. In previous studies, methods to measure engagement vary among researchers. Lee et al. (2009) measured engagement in online community activities by taking into account the number of times the users had used the community bulletin boards, whether they had posted their own photographs, the number of times they had attended offline gatherings, the number of calls they had made for inquiries, and the period in which they had been engaged in community activities. Lee et al. (2011) studied Twitter, measuring the desire of its users to follow a corporate Twitter account, intimacy, their desire to share information, and their desire for engagement. Then, they measured the degree of user engagement based on the four stages of customer engagement by Kellogg et al. (1997). Recent studies on engagement that are focused on social media have calculated engagement based on data generated on social media along with the responses of users. Donna L. (2010) divided the indicators that can be measured on social media based on social media type and outcome of activity — brand awareness, brand engagement, word of mouth — and presented Facebook likes, comments, and user posts as indicators related to engagement. Brodie, RJ, et al. (2011) measured engagement based on the length of posts or comments as well as engagement intervals, whereas Gummerus et al. (2012) considered the frequency of community visits, likes, and comments in calculating engagement. Facebook Insight, a data analysis tool officially provided by Facebook, takes into account the actions of clicking on content, likes, and comments and asking questions when distinguishing engaged users. Facebook
calculates engagement by considering the number of likes, comments, shares, and the total number of fans.

2.2  **Relational Benefits and Impacts of Industrial Characteristics**

The uses and gratifications theory originates from the interest in what people do with media and the process of clarifying the above (Kim, 2011). This theory approaches media choice and use in terms of social psychology under the premise that users have certain intentions or goals in using media. Since people use certain media by synchronizing not only external factors but also internal factors, such as personal goals or intentions (Stafford & Stafford, 2001), it is necessary to find out what the user motivators are in using social media. Blumler (1979) suggested that media, which is to be explained by the uses and gratifications theory, must have the elements of choice, engagement, and use. According to Rafaeli and LaRose (1993), members who are interested in virtual communities are engaged in using communities by showing the active communication behaviours of writing and posting messages themselves. Therefore, virtual communities may become the object of media that can be approached through the uses and gratifications theory. Motivators for certain media reflect the characteristics of that media, and motivators for using virtual communities emphasize characteristics such as seeking information, forming homogenous relationships, confirming identity, pursuing personal interests, and having an exchange with others (Kim, 2005). Other studies identify interaction (Kwon, Woo, 2005), exchange with others (Nardi et al., 2004), and forming and maintaining relationships (Kim, 2005) as motivators for using virtual communities. These motivators reflect users’ expectations for relationships.

Based on the fact that relationships are formed between the brand and users and among users in the brand community on social media, it can be inferred that there are relational benefits (Gwinner et al., 1998) in users’ intentions in using brand communities. Gwinner et al. (1998) defined relational benefits as those including not only the fundamental benefits of core services provided by firms to form and build relationships with customers but also all other benefits provided to customers. These relational benefits are classified into four types in terms of how much customers perceive and value the benefits: social, psychological, economic, and customization benefits. Social benefits include intimacy between the firm and customers, personal recognition, and friendship, while psychological benefits comprise anxiety reduction, trust, and conviction. Economic benefits include discounts and saving time, and customization benefits comprise preferential treatment, additional services, and determination and management of customer needs. To these, Dholakia et al. (2004) added entertainment benefits, which include convenience and fun, as they reflect the characteristics of online communities, and found that such benefits may promote community engagement. Based on Dholakia et al. (2004), this study examines the relational benefits of social media brand communities by focusing on social, entertainment, economic, and customization benefits. Based on the uses and gratifications theory, obtaining relational benefits is the motivator for using brand communities on social media reflecting the characteristics of social media. This may have influence on engagement in and loyalty to brand communities, which is empirically analysed by Gummerus et al. (2012). Furthermore, regardless of the intention of use, the degree of perception may also directly influence engagement and loyalty. According to Dholakia et al. (2004), the value perceived in virtual communities ultimately has influence on engagement behaviours. Gummerus et al. (2012) confirmed that perceived benefits focused on relational benefits have a direct influence on loyalty.

Studies have been made on the relationship between business performance and the implementation of new technology (Barua et al., 1995). Firms operating new media called brand communities on social media also do not have the same industrial structure or characteristics and may thus have differences in the operation and performance of brand communities. Ortega et al. (2006) found that corporate characteristics show a moderating effect on online behaviour, and stated that firms with more experience in technology accept IT more quickly than other firms. Moreover, Gronholdt et al. (2000) investigated how customer satisfaction and loyalty varied among industries, and determined that six industries had different values perceived by consumers, which affected their
satisfaction and loyalty. Furthermore, satisfaction and loyalty had a positive correlation but had relative differences among industries. Elowitz and Li (2009) examined the correlation between the use of social media and financial performance of the top 100 global brands. The results of the study showed differences among industries in the application level of social media by brand; a higher application level indicated the higher financial performance of the relevant firm.

3 RESEARCH QUESTIONS

Based on theoretical discussions and previous studies, this study suggests that the intention in using brand communities is to obtain relational benefits, based on the uses and gratifications theory. Moreover, the engagement of users is affected by their satisfaction in relational benefits, and this engagement leads to corporate performance, such as customer loyalty. Based on the above, the following three research questions were formulated.

Research Question 1: What are the relational benefits expected by potential users in brand communities by industry?

Previous studies have confirmed that potential users may start using brand pages as they expect relational benefits from them. The benefits perceived by potential users may also vary among industrial characteristics. However, no study has argued that users may expect different benefits from each brand page by taking this observation into account. This research question discusses which relational benefits are expected by potential users in brand communities by industry as well as the level of such expectations.

Research Question 2: Does customer engagement in brand communities show different characteristics by industry?

Previous studies have found that there are impacts on user behaviours according to industrial characteristics and this phenomenon is likely to occur in brand communities on social media as well. Previous studies did not compare engagement by industry. This research question examines whether customer engagement in brand communities shows different characteristics by industry, in comparison with the average user engagement in brand communities.

Research Question 3: What relational benefits increase user engagement in brand communities?

Previous studies have confirmed that satisfaction in the benefits obtained by users from brand communities affects engagement and loyalty. The benefits perceived by users in brand communities may also vary among industries and affect loyalty behaviours. Therefore, benefits that affect engagement are likely to occur differently among communities. This research question examines the difference in the satisfaction of users in relational benefits among different industries and distinguishes the relational benefits that increase user engagement in brand communities by industry.

4 RESEARCH METHOD

Measurement items for selection attributes to measure the expectation level and satisfaction in relational benefits were extracted from previous studies. To measure the expectation level in each type of relational benefits, items related to the content of this study were selected among the survey items used in Dholakia et al. (2004), Hong and Han (2011), and Kim (2005), which were then modified to reflect the characteristics of using brand communities on social media and to be explained by a scale ranging from “have no expectations at all” to “have high expectations”. To measure satisfaction by type of relational benefits, this study changed the form of responses in the survey used in the measurement of expectation level, using a scale ranging from “not satisfied at all” to “highly satisfied”. Variables measuring the expectation level and satisfaction were all rated on a five-point Likert interval scale, and general details, such as the demographic characteristics of respondents, were rated on a nominal scale. The contents of the questionnaire are shown in <Table 1>.
Table 1. | Questionnaires

<table>
<thead>
<tr>
<th>Category</th>
<th>Item</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social benefits</td>
<td>Feeling intimacy toward this brand</td>
<td>Hong &amp; Han (2011)</td>
</tr>
<tr>
<td></td>
<td>Sharing thoughts and opinions about this brand with others</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Feeling intimacy toward other people who like this brand</td>
<td></td>
</tr>
<tr>
<td>Entertainment benefits</td>
<td>Having fun and feeling pleasure</td>
<td>Kim (2005)</td>
</tr>
<tr>
<td></td>
<td>Being offered an unusual experience</td>
<td>Dholakia et al. (2004)</td>
</tr>
<tr>
<td></td>
<td>Spending spare time</td>
<td></td>
</tr>
<tr>
<td>Economic benefits</td>
<td>Receiving economic benefits, such as discounts, on the page</td>
<td>Hong &amp; Han (2011)</td>
</tr>
<tr>
<td></td>
<td>Receiving information about economic benefits on the page</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Receiving information about other events on the page</td>
<td></td>
</tr>
<tr>
<td>Customization benefits</td>
<td>Receiving abundant information about the brand</td>
<td>Hong &amp; Han (2011)</td>
</tr>
<tr>
<td></td>
<td>Feeling as if receiving information more quickly than other customers via the page</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Receiving appropriate and specific responses from the page administrator</td>
<td></td>
</tr>
</tbody>
</table>

Variables to measure engagement were used in light of the typical engagement behaviours of Facebook users. Engagement is calculated by the newly added likes, comments, and shares on a given day on each page, divided by the total number of fans on the given day. For analysis, the moving average during the data collection period was used as the average engagement value. To conduct research, the top 74 Facebook pages of Korean brands with at least 35,000 fans were selected. Engagement data per page was collected using Analytics Pro, the social media monitoring tool provided by Socialbakers. The brands selected as the samples were classified into six industries based on the Korean Standard Industrial Classification, and industries that did not include enough samples were excluded, leaving a total of 64 pages to be classified into the manufacturing, fast-moving consumer goods (FMCG), and IT and Media industries. The brand pages included in each industry are attached in the appendix. The scope of page data collection was limited to one month, from November 5 to December 5, 2012. Data collection on user expectation and satisfaction in each type of relational benefits was conducted using the online survey platform, targeting users that were fans of at least one brand page on Facebook. As a result, 356 valid respondents were secured and data from a total of 1360 pages was collected. Among these, 1078 responses were on the expectation level of potential users, and 282 responses were on the satisfaction of actual users.

The collected data were analysed using SPSS 20.0, and reliability and factor analyses were conducted on each relational benefit variable to be used in the analysis. Moreover, a multinomial logistic regression analysis was conducted to determine the difference in expectation for relational benefit types by industry and the difference in engagement. Furthermore, a binomial logistic regression analysis was conducted to determine the relationship between satisfaction and engagement in relational benefits by industry. The logistic regression analysis was used so that it can determine the independent variables affecting the categorical dependent variables and obtain more accurate results than a t-test or analysis of variance (ANOVA) since it considers the correlations among independent variables (Oh & Han, 2009).

5 Research Findings

5.1 Analysis of Findings for Research Question 1: Types of Relational Benefits Expected by Potential Users

A factorial analysis was conducted to examine whether it was valid to classify the investigated items that formed the expectation for each type of relational benefits into four factors.
For the rotation method, the varimax rotation of principal component analysis was used as it minimizes the loss of variable information and effectively increases the total explanatory power of variance. The cumulative loading of the four factors that formed the expectation for each type of relational benefits was 76.964%, confirming that the survey structure was appropriate.

<table>
<thead>
<tr>
<th>Factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Eigenvalue</th>
<th>Explanatory power of variance</th>
<th>Cronbach’s alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economic_2</td>
<td>.937</td>
<td>-.118</td>
<td>-.012</td>
<td>.155</td>
<td>4.205</td>
<td>20.635</td>
<td>.893</td>
</tr>
<tr>
<td>Economic_1</td>
<td>.935</td>
<td>-.106</td>
<td>-.018</td>
<td>.107</td>
<td>3.487</td>
<td>19.252</td>
<td>.870</td>
</tr>
<tr>
<td>Economic_3</td>
<td>.767</td>
<td>.031</td>
<td>.119</td>
<td>.339</td>
<td>1.962</td>
<td>18.478</td>
<td>.835</td>
</tr>
<tr>
<td>Entertainment_2</td>
<td>-.024</td>
<td>.858</td>
<td>.237</td>
<td>.161</td>
<td>2.988</td>
<td>19.966</td>
<td>.870</td>
</tr>
<tr>
<td>Entertainment_1</td>
<td>.095</td>
<td>.838</td>
<td>.296</td>
<td>.121</td>
<td>2.444</td>
<td>19.432</td>
<td>.835</td>
</tr>
<tr>
<td>Entertainment_3</td>
<td>-.091</td>
<td>.833</td>
<td>.223</td>
<td>.058</td>
<td>2.001</td>
<td>18.432</td>
<td>.835</td>
</tr>
<tr>
<td>Social_2</td>
<td>-.059</td>
<td>.230</td>
<td>.832</td>
<td>.179</td>
<td>1.766</td>
<td>18.132</td>
<td>.835</td>
</tr>
<tr>
<td>Social_3</td>
<td>.021</td>
<td>.278</td>
<td>.823</td>
<td>.096</td>
<td>1.569</td>
<td>17.822</td>
<td>.835</td>
</tr>
<tr>
<td>Social_1</td>
<td>.127</td>
<td>.243</td>
<td>.789</td>
<td>.172</td>
<td>1.364</td>
<td>17.522</td>
<td>.835</td>
</tr>
<tr>
<td>Customer_2</td>
<td>.130</td>
<td>.087</td>
<td>.121</td>
<td>.859</td>
<td>1.442</td>
<td>17.222</td>
<td>.835</td>
</tr>
<tr>
<td>Customer_1</td>
<td>.232</td>
<td>.030</td>
<td>.225</td>
<td>.768</td>
<td>1.972</td>
<td>17.429</td>
<td>.748</td>
</tr>
<tr>
<td>Customer_3</td>
<td>.168</td>
<td>.198</td>
<td>.082</td>
<td>.706</td>
<td>1.321</td>
<td>17.122</td>
<td>.835</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>76.964</td>
<td></td>
<td>.811</td>
</tr>
</tbody>
</table>

Factor extraction method: Principal component analysis; Rotation method: Varimax with Kaiser normalization

Table 2. Results of Factor Analysis and Reliability Analysis on Potential Users’ Expectation for Relational Benefits

The Cronbach’s alpha (a confidence coefficient that represents the internal consistency of each factor, and thus verifies the reliability of a study) of each factor was 0.7 or higher, showing adequate reliability. Moreover, the overall reliability coefficient was analysed to be 0.811, confirming that reliability was assured at the overall level. To examine which relational benefits were expected by users in each brand community by industry, the factor scores deduced by categorising the 12 items into four factors were used as independent variables according to the procedure of factor analysis.

Examination of the benefits expected by potential users in each industry based on the factor scores of each factor extracted from the factor analysis revealed evident differences among the industries in the expectation for entertainment benefits and economic benefits. A multinomial logistic regression analysis was conducted to test this statistically.
Figure 1. Relational Benefits Expected by Potential Users by Industry

<Table 3> shows the results of the likelihood ratio test of multinomial logistic regression analysis. A complete factor model was used for analysis, and all four factors were significant at the 1% level. The variable with considerable influence on distinguishing industries was the expectation for economic and entertainment benefits, whereas the expectation for social and customization benefits was relatively low.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Model customization standard</th>
<th>Likelihood ratio test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-2 Log likeness of scale model</td>
<td>Chi-square</td>
</tr>
<tr>
<td>Intercept</td>
<td>1715.381</td>
<td>75.997</td>
</tr>
<tr>
<td>Social benefits</td>
<td>1649.289</td>
<td>9.905</td>
</tr>
<tr>
<td>Entertainment benefits</td>
<td>1844.591</td>
<td>205.207</td>
</tr>
<tr>
<td>Economic benefits</td>
<td>2022.564</td>
<td>383.179</td>
</tr>
<tr>
<td>Customization benefits</td>
<td>1641.926</td>
<td>2.542</td>
</tr>
</tbody>
</table>

Table 3. Likelihood Ratio Test on Potential Users’ Expectation for Relational Benefits

To analyse the research question, this study selected the IT and media industry as the reference group. The result of multinomial logistic regression analysis on the expectation for relational benefits by industry showed that the analysis model was significant at the 1% level. The logistic regression analysis used “There is no difference between the actual data and predicted value of the model” as the null hypothesis to test the goodness of fit of the model, and the significance level of the goodness of fit must be p>0.05 for the model to be fit. Pearson’s chi-square, which shows the consistency between prediction of the model and actual data, was p>0.05 at 1686.378 (degree of freedom=1874), thereby supporting the goodness of fit of the model. The prediction accuracy of the model was 57.6%.

The results can be analysed based on the code of the predicted beta value and odds ratio (Exp(B)), showing that the relevant variable had a positive impact if the odds ratio was above 1.0 and a negative impact if the odds ratio was below 1.0. The odds ratio is an estimated value of the
probability that the variable will be classified from the reference group to the relevant group when the value of the explanatory variable increases by 1. Based on the above, the relative influence of each benefit on the dependent variables can be identified. The analysis results of the research question can be interpreted as follows: As the value of the relevant benefit increases by 1 in each industry, the probability that it will be classified into the relevant industry from the IT and media industry, which is the reference group, also increases by Exp(B).

<table>
<thead>
<tr>
<th>Industry</th>
<th>B</th>
<th>Standard error</th>
<th>Wald</th>
<th>Significance probability</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturing</td>
<td>Intercept</td>
<td>.382</td>
<td>.087</td>
<td>19.105</td>
<td>.000</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>Social benefits</td>
<td>-.069</td>
<td>.084</td>
<td>6.89</td>
<td>406</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>Entertainment benefits</td>
<td>-.753</td>
<td>.095</td>
<td>63.114</td>
<td>.000**</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>Economic benefits</td>
<td>7.95</td>
<td>.089</td>
<td>80.319</td>
<td>.000**</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>Customization benefits</td>
<td>.089</td>
<td>.081</td>
<td>1.207</td>
<td>272</td>
</tr>
<tr>
<td>FMCG</td>
<td>Intercept</td>
<td>-.503</td>
<td>.121</td>
<td>17.388</td>
<td>.000</td>
</tr>
<tr>
<td>FMCG</td>
<td>Social benefits</td>
<td>-.306</td>
<td>.104</td>
<td>8.744</td>
<td>.003**</td>
</tr>
<tr>
<td>FMCG</td>
<td>Entertainment benefits</td>
<td>1.497</td>
<td>.120</td>
<td>156.766</td>
<td>.000**</td>
</tr>
<tr>
<td>FMCG</td>
<td>Economic benefits</td>
<td>2.179</td>
<td>.144</td>
<td>227.596</td>
<td>.000**</td>
</tr>
<tr>
<td>FMCG</td>
<td>Customization benefits</td>
<td>-.036</td>
<td>.103</td>
<td>121</td>
<td>728</td>
</tr>
</tbody>
</table>

a. The reference category is IT and media.
Final-model adequate-information model standard: -2log-likelihood 1639.384
Final-model adequate-information likelihood ratio test: chi-square 536.807; degrees of freedom 8; significance probability .000**
Goodness of fit: Pearson’s chi-square 1686.378; degrees of freedom 1874; significance probability .999
Prediction accuracy: 57.6% (manufacturing 41.6%; FMCG 64.3%; IT and media 68.2%)

Table 4. Results of Multinomial Logistic Regression Analysis of Potential Users’ Expectation for Relational Benefits

The results in <Table 4> showed that expectations for entertainment and economic benefits were significant factors between IT and media and manufacturing. Based on the Exp(B) value, there was statistical significance – the expectations for economic benefits were high and those for entertainment benefits were low in the pages of manufacturing compared with the pages of IT and media. Expectations for social, entertainment, and economic benefits were significant factors between IT and media and FMCG. There was statistical significance – the expectations for social and entertainment benefits were low and those for economic benefits were extremely high in FMCG compared with those in IT and media. In sum, social benefits are most expected in IT and media, followed by manufacturing and FMCG; entertainment benefits are most expected in IT and media, followed by manufacturing and FMCG; economic benefits are most expected in FMCG, followed by manufacturing and IT and media; and customization benefits are expected at a similar level in all industries. In particular, FMCG showed an extremely high level of expectations for economic benefits and low expectations for entertainment benefits, while IT and media showed an extremely high level of expectations for entertainment benefits and low expectations for economic benefits.
5.2 **Analysis of Findings for Research Question 2: Average Engagement of Brand Communities by Industry**

To determine whether engagement varies among brand communities by industry, this study collected the engagement indices of 74 brand pages in Korea and used the engagement data of 64 pages in the analysis: 15 pages in manufacturing, 27 pages in FMCG, and 22 pages in IT and media. Verification of the average engagement in each industry based on descriptive statistics revealed IT and media to have the highest engagement at 1.0353%, followed by manufacturing at 0.5820% and FMCG at 0.4167%. The average engagement in the 74 collected pages was 0.7184%. Multinomial logistic regression analysis was conducted to test this statistically. The multinomial logistic regression analysis was used so that it can test whether the average engagement actually has influence on distinguishing industries and, unlike ANOVA and t-test, it simultaneously presents the analysis results of industrial differences for a comparison of the average among industries.

![Average Engagement of Users in Brand Pages by Industry](image)

<Table 5> shows the results of the likelihood ratio test of multinomial logistic regression analysis. A complete factor model was used for the analysis, and the average engagement was a significant variable at the 5% level. In other words, the average engagement had influence on distinguishing the industries.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Effect</th>
<th>Chi-square</th>
<th>Degree of freedom</th>
<th>Significance probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2 Log likeness of scale model</td>
<td>103.989</td>
<td>2</td>
<td>.035</td>
</tr>
<tr>
<td>Average engagement</td>
<td>105.047</td>
<td>7.781</td>
<td>2</td>
<td>.020*</td>
</tr>
</tbody>
</table>

As in 4.1, this study selected the IT and media industry as the reference group to analyse the research question. <Table 6> shows the results of the multinomial logistic regression analysis on the expectations for relational benefits by industry, and the analysis model was significant at the 5% level. Pearson’s chi-square was $p>0.05$ at 90.478 (degree of freedom=84), thereby supporting the goodness of fit of the model. The prediction accuracy of the model was 56.2%.
Table 6. Results of Multinomial Logistic Regression Analysis of Average Engagement of Brand Community Users

<table>
<thead>
<tr>
<th>Industrya</th>
<th>B</th>
<th>Standard error</th>
<th>Wald</th>
<th>Significance probability</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>.109</td>
<td>.475</td>
<td>.053</td>
<td>818</td>
<td></td>
</tr>
<tr>
<td>Average ER</td>
<td>-.642</td>
<td>.483</td>
<td>1.762</td>
<td>184</td>
<td>.526</td>
</tr>
<tr>
<td>FMCG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>.978</td>
<td>.433</td>
<td>5.094</td>
<td>.024*</td>
<td></td>
</tr>
<tr>
<td>Average ER</td>
<td>-1.214</td>
<td>.542</td>
<td>5.016</td>
<td>.025*</td>
<td>.297</td>
</tr>
</tbody>
</table>

a. The reference category is IT and media.

Final-model adequate-information model standard: -2log-likelihood 97.267

Final-model adequate-information likelihood ratio test: chi-square 7.781; degree of freedom 2; significance probability .020*

Goodness of fit: Pearson’s chi-square 90.478; degree of freedom 84; significance probability .295 Prediction accuracy: 56.2% (manufacturing 0.0%; FMCG 85.2%; media 59.1%)

The significance probability was p=0.184> 0.05 between IT and media and manufacturing, and the difference in engagement was thus not statistically significant. However, the significance probability was p=0.025<0.05 between IT and media and FMCG, confirming that the difference in engagement was statistically significant. In other words, the average engagement of the pages in IT and media was higher than that of the pages in FMCG.

5.3 Analysis of Findings for Research Question 3: Types of Relational Benefits That Increase User Engagement

A factorial analysis was conducted to examine whether it was valid to classify the investigated items that formed satisfaction in each type of relational benefits into four factors. For the rotation method, the varimax rotation of principal component analysis was used as it minimizes the loss of variable information and effectively increases the total explanatory power of variance. The cumulative loading of the four factors was 76.964%, confirming that the survey structure was appropriate. The Cronbach’s alpha of each factor was 0.7 or higher, showing adequate reliability. Moreover, the overall reliability coefficient was analysed to be 0.835, confirming that reliability was assured at the overall level.
To examine brand community users’ satisfaction in each type of relational benefits, the factor scores deduced by categorizing the 12 items into four factors were used as independent variables according to the procedure of factor analysis.

Examination of the average satisfaction of users in each benefit type by industry based on descriptive statistics of each independent variable revealed evident differences among industries in satisfaction in entertainment and economic benefits. Logistic regression analysis was conducted to test this statistically.

### Table 7. Results of Factor Analysis and Reliability Analysis of Users’ Satisfaction in Relational Benefits

<table>
<thead>
<tr>
<th>Factor</th>
<th>Factor loadings</th>
<th>Eigenvalue</th>
<th>Explanatory power of variance</th>
<th>Cronbach’s alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Entertainment satisfaction_1</td>
<td>.896</td>
<td>-.051</td>
<td>259</td>
<td>.099</td>
</tr>
<tr>
<td>Entertainment satisfaction_2</td>
<td>.858</td>
<td>.014</td>
<td>305</td>
<td>.157</td>
</tr>
<tr>
<td>Entertainment satisfaction_3</td>
<td>.784</td>
<td>-.045</td>
<td>323</td>
<td>.122</td>
</tr>
<tr>
<td>Economic satisfaction_2</td>
<td>-.030</td>
<td>.937</td>
<td>-.012</td>
<td>.121</td>
</tr>
<tr>
<td>Economic satisfaction_1</td>
<td>-.008</td>
<td>.934</td>
<td>-.041</td>
<td>.068</td>
</tr>
<tr>
<td>Economic satisfaction_3</td>
<td>-.040</td>
<td>.771</td>
<td>149</td>
<td>.335</td>
</tr>
<tr>
<td>Social satisfaction_3</td>
<td>.308</td>
<td>-.016</td>
<td>843</td>
<td>.147</td>
</tr>
<tr>
<td>Social satisfaction_2</td>
<td>.324</td>
<td>-.018</td>
<td>824</td>
<td>.083</td>
</tr>
<tr>
<td>Social satisfaction_1</td>
<td>.276</td>
<td>.123</td>
<td>677</td>
<td>.255</td>
</tr>
<tr>
<td>Customer satisfaction_2</td>
<td>.077</td>
<td>.053</td>
<td>.097</td>
<td>.859</td>
</tr>
<tr>
<td>Customer satisfaction_1</td>
<td>.099</td>
<td>.271</td>
<td>.088</td>
<td>.799</td>
</tr>
<tr>
<td>Customer satisfaction_3</td>
<td>.192</td>
<td>.164</td>
<td>259</td>
<td>.660</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
<Table 8> shows the results of the likelihood ratio test of multinomial logistic regression analysis. A complete factor model was used for analysis, and satisfaction in entertainment benefits and economic benefits were significant variables at the 5% level, thereby confirming that they have influence on distinguishing the industries. The variable with considerable influence on distinguishing industries was satisfaction in economic benefits, followed by satisfaction in entertainment benefits.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Model customization standard</th>
<th>Likelihood ratio test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-2 Log likeness of scale model</td>
<td>Chi-square</td>
</tr>
<tr>
<td>Intercept</td>
<td>495.229</td>
<td>8.784</td>
</tr>
<tr>
<td>Satisfaction in social benefits</td>
<td>490.646</td>
<td>4.201</td>
</tr>
<tr>
<td>Satisfaction in entertainment benefits</td>
<td>523.108</td>
<td>36.664</td>
</tr>
<tr>
<td>Satisfaction in economic benefits</td>
<td>570.626</td>
<td>84.182</td>
</tr>
<tr>
<td>Satisfaction in customization benefits</td>
<td>486.625</td>
<td>180</td>
</tr>
</tbody>
</table>

Table 8. Likelihood Ratio Test on Users’ Satisfaction in Relational Benefits

To analyse the research question, this study selected the IT and media industry as the reference group. <Table 9> shows the results of the multinomial logistic regression analysis of satisfaction in relational benefits by industry, and the analysis model was significant at the 1% level. Pearson’s chi-square was p>0.05 at 577.867 (degree of freedom=530), thereby supporting the goodness of fit of the model. The prediction accuracy of the model was 58.9%.
### Table 9. Results of Multinomial Logistic Regression Analysis of Users’ Satisfaction in Relational Benefits

The benefits that showed significant difference in satisfaction among industries were entertainment and economic benefits. Compared with IT and media, manufacturing and FMCG showed low satisfaction in entertainment benefits and high satisfaction in economic benefits. In other words, IT and media showed higher satisfaction in entertainment benefits than the other two industries while showing lower satisfaction in economic benefits. On the contrary, FMCG showed extremely high satisfaction in economic benefits but extremely low satisfaction in entertainment benefits.

To determine the influence of benefit types on satisfaction within the same industry, this study selected four pages with high engagement and four pages with low engagement for each industry. <Table 10> shows the scope of average engagement of pages with high engagement and pages with low engagement in each industry.

### Table 10. Scope of Engagement in Selected Brand Pages
Measurement and examination of actual users’ satisfaction in each benefit type for each page revealed a difference in satisfaction in each benefit type, as shown in Figure 4, Figure 5, and Figure 6.

*Figure 4* [Manufacturing Industry] Satisfaction in Relational Benefits According to Page Engagement

*Figure 5* [FMCG Industry] Satisfaction in Relational Benefits According to Page Engagement

*Figure 6* [IT and Media Industry] Satisfaction in Relational Benefits According to Page Engagement
Binomial logistic regression analysis for each industry was conducted to test this statistically, and the results are summarized in <Table 11>, <Table 12>, and <Table 13>.

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>Significance probability</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Satisfaction in social benefits</strong></td>
<td>1.17</td>
<td>.37</td>
<td>5.67</td>
<td>.017*</td>
<td>1.931</td>
</tr>
<tr>
<td><strong>Satisfaction in entertainment benefits</strong></td>
<td>2.18</td>
<td>.33</td>
<td>5.70</td>
<td>.017*</td>
<td>2.180</td>
</tr>
<tr>
<td><strong>Satisfaction in economic benefits</strong></td>
<td>.47</td>
<td>.33</td>
<td>5.09</td>
<td>.024*</td>
<td>.478</td>
</tr>
<tr>
<td><strong>Satisfaction in customization benefits</strong></td>
<td>3.29</td>
<td>.36</td>
<td>11.25</td>
<td>.001**</td>
<td>3.291</td>
</tr>
</tbody>
</table>

---

**Step 1a**

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>Significance probability</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Satisfaction in social benefits</strong></td>
<td>1.23</td>
<td>.27</td>
<td>5.59</td>
<td>.017*</td>
<td>1.232</td>
</tr>
<tr>
<td><strong>Satisfaction in entertainment benefits</strong></td>
<td>1.1</td>
<td>.27</td>
<td>5.12</td>
<td>.017*</td>
<td>1.101</td>
</tr>
<tr>
<td><strong>Satisfaction in economic benefits</strong></td>
<td>1.79</td>
<td>.65</td>
<td>19.53</td>
<td>.000**</td>
<td>17.912</td>
</tr>
<tr>
<td><strong>Satisfaction in customization benefits</strong></td>
<td>2.75</td>
<td>.36</td>
<td>7.89</td>
<td>.005**</td>
<td>2.752</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>Significance probability</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Satisfaction in social benefits</strong></td>
<td>.56</td>
<td>.28</td>
<td>5.67</td>
<td>.017*</td>
<td>1.931</td>
</tr>
<tr>
<td><strong>Satisfaction in entertainment benefits</strong></td>
<td>2.18</td>
<td>.33</td>
<td>5.70</td>
<td>.017*</td>
<td>2.180</td>
</tr>
<tr>
<td><strong>Satisfaction in economic benefits</strong></td>
<td>.47</td>
<td>.33</td>
<td>5.09</td>
<td>.024*</td>
<td>.478</td>
</tr>
<tr>
<td><strong>Satisfaction in customization benefits</strong></td>
<td>3.29</td>
<td>.36</td>
<td>11.25</td>
<td>.001**</td>
<td>3.291</td>
</tr>
</tbody>
</table>

---

**Constant term**

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>Significance probability</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Satisfaction in social benefits</strong></td>
<td>1.17</td>
<td>.37</td>
<td>5.67</td>
<td>.017*</td>
<td>1.931</td>
</tr>
<tr>
<td><strong>Satisfaction in entertainment benefits</strong></td>
<td>2.18</td>
<td>.33</td>
<td>5.70</td>
<td>.017*</td>
<td>2.180</td>
</tr>
<tr>
<td><strong>Satisfaction in economic benefits</strong></td>
<td>.47</td>
<td>.33</td>
<td>5.09</td>
<td>.024*</td>
<td>.478</td>
</tr>
<tr>
<td><strong>Satisfaction in customization benefits</strong></td>
<td>3.29</td>
<td>.36</td>
<td>11.25</td>
<td>.001**</td>
<td>3.291</td>
</tr>
</tbody>
</table>

---

**Constant term**

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>Significance probability</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Satisfaction in social benefits</strong></td>
<td>1.17</td>
<td>.37</td>
<td>5.67</td>
<td>.017*</td>
<td>1.931</td>
</tr>
<tr>
<td><strong>Satisfaction in entertainment benefits</strong></td>
<td>2.18</td>
<td>.33</td>
<td>5.70</td>
<td>.017*</td>
<td>2.180</td>
</tr>
<tr>
<td><strong>Satisfaction in economic benefits</strong></td>
<td>.47</td>
<td>.33</td>
<td>5.09</td>
<td>.024*</td>
<td>.478</td>
</tr>
<tr>
<td><strong>Satisfaction in customization benefits</strong></td>
<td>3.29</td>
<td>.36</td>
<td>11.25</td>
<td>.001**</td>
<td>3.291</td>
</tr>
</tbody>
</table>

---

Omnibus test of model coefficient: chi-square 32.609; significance probability .000**; -2 log-likelihood 91.446

Hosmer–Lemeshow test: chi-square 9.336; degree of freedom 8; significance probability .315

Prediction accuracy: 72.2% ([manufacturing] engagement low 75.5%; [manufacturing] engagement high 68.3%)

---

Table 11. Manufacturing Industry: Results of Binomial Logistic Regression Analysis of Satisfaction and Engagement in Benefits

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>Significance probability</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Satisfaction in social benefits</strong></td>
<td>2.09</td>
<td>.27</td>
<td>5.91</td>
<td>.442</td>
<td>1.232</td>
</tr>
<tr>
<td><strong>Satisfaction in entertainment benefits</strong></td>
<td>1.09</td>
<td>.27</td>
<td>5.12</td>
<td>.729</td>
<td>1.101</td>
</tr>
<tr>
<td><strong>Satisfaction in economic benefits</strong></td>
<td>2.88</td>
<td>.65</td>
<td>19.52</td>
<td>.000**</td>
<td>17.912</td>
</tr>
<tr>
<td><strong>Satisfaction in customization benefits</strong></td>
<td>1.01</td>
<td>.36</td>
<td>7.89</td>
<td>.005**</td>
<td>2.752</td>
</tr>
</tbody>
</table>

---

**Constant term**

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>Significance probability</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Satisfaction in social benefits</strong></td>
<td>1.17</td>
<td>.37</td>
<td>5.67</td>
<td>.017*</td>
<td>1.931</td>
</tr>
<tr>
<td><strong>Satisfaction in entertainment benefits</strong></td>
<td>2.18</td>
<td>.33</td>
<td>5.70</td>
<td>.017*</td>
<td>2.180</td>
</tr>
<tr>
<td><strong>Satisfaction in economic benefits</strong></td>
<td>.47</td>
<td>.33</td>
<td>5.09</td>
<td>.024*</td>
<td>.478</td>
</tr>
<tr>
<td><strong>Satisfaction in customization benefits</strong></td>
<td>3.29</td>
<td>.36</td>
<td>11.25</td>
<td>.001**</td>
<td>3.291</td>
</tr>
</tbody>
</table>

---

**Constant term**

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>Significance probability</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Satisfaction in social benefits</strong></td>
<td>1.17</td>
<td>.37</td>
<td>5.67</td>
<td>.017*</td>
<td>1.931</td>
</tr>
<tr>
<td><strong>Satisfaction in entertainment benefits</strong></td>
<td>2.18</td>
<td>.33</td>
<td>5.70</td>
<td>.017*</td>
<td>2.180</td>
</tr>
<tr>
<td><strong>Satisfaction in economic benefits</strong></td>
<td>.47</td>
<td>.33</td>
<td>5.09</td>
<td>.024*</td>
<td>.478</td>
</tr>
<tr>
<td><strong>Satisfaction in customization benefits</strong></td>
<td>3.29</td>
<td>.36</td>
<td>11.25</td>
<td>.001**</td>
<td>3.291</td>
</tr>
</tbody>
</table>

---

Omnibus test of model coefficient: chi-square 52.389; significance probability .000**; -2 log-likelihood 91.817

Hosmer–Lemeshow test: chi-square 12.656; degree of freedom 8; significance probability .124

Prediction accuracy: 83.6% ([FMCG] engagement Hosmer–Lemeshow test 70.0%; [FMCG] engagement high 91.4%)

---

Table 12. Consumer Goods Industry: Results of Binomial Logistic Regression Analysis of Satisfaction and Engagement in Benefits
The results of binomial logistic regression analysis showed that the significance probability of -2 log-likelihood and chi-square value that verified the goodness of fit of the model for each industry was $p<0.05$, indicating that the model including each benefit type as an independent variable was significant. The Hosmer–Lemeshow test, conducted to verify the goodness of fit of the regression model, revealed the significance probability to be $p>0.05$ in all industries, indicating that the model was fit. Prediction accuracy was 72.2% in manufacturing, 83.6% in FMCG, and 92.7% in IT and media.

In manufacturing, there was a significant difference in satisfaction in all benefit types between the group with high engagement and the group with low engagement. The Exp(B) values of other factors that had a positive impact on increasing engagement showed that the influence of satisfaction in customization benefits was the strongest, followed by those of satisfaction in entertainment benefits and satisfaction in social benefits. This indicates that the difference of satisfaction in customization benefits was the biggest between pages with high engagement and pages with low engagement in manufacturing. Therefore, engagement is most likely to increase if there is higher satisfaction in customization benefits among the types of relational benefits in manufacturing. As for economic benefits, satisfaction appeared to be higher in the group with low engagement. Interpretation of this result is discussed in the conclusion.

In FMCG, there was a significant difference in satisfaction in economic benefits and customization benefits between the group with high engagement and the group with low engagement. Both factors were higher in the group with high engagement, and the Exp(B) value of satisfaction in economic benefits was 17.912, much higher than the satisfaction in customization benefits at 2.752. This indicates that satisfaction in economic benefits has a significant influence on increasing engagement. In other words, engagement is most likely to increase if there is higher satisfaction in economic benefits among the types of relational benefits in FMCG.

In IT and media, there was a significant difference in satisfaction in all benefit types between the group with high engagement and the group with low engagement. The Exp(B) value of satisfaction in entertainment benefits was extremely high at 109.280, which indicates that even slightly increasing satisfaction in entertainment benefits will lead to a much higher probability of increasing engagement. The Exp(B) values of satisfaction in customization benefits and social benefits were also high at 35.924 and 11.015, respectively. In other words, in IT and media, even slightly increasing satisfaction in the three factors other than economic benefits will increase the probability that engagement increases compared with that in other industries. Satisfaction in
entertainment benefits has the biggest influence among them. As for economic benefits, satisfaction was higher in the group with low engagement in the manufacturing sector.

6 CONCLUSION

To provide baseline data for the effective and efficient management of firm-hosted brand communities, this study suggested that the intention in using brand communities on social media is to obtain relational benefits, and examined which relational benefits are expected by potential users in each industry and how the brand community engagement of actual users varies according to their satisfaction in each relational benefit based on collected data. The results of the analysis can be summarized as follows.

First, examination of the relational benefits expected by potential users from brand communities on social media in each industry revealed that the benefits with the biggest difference in the expectation level among industries are economic and entertainment benefits. In manufacturing, users have the highest expectations for customization benefits compared with those in other industries, but the result was not statistically significant. In FMCG, users have extremely high expectations for economic benefits compared with those in other industries, while in IT and media, users have high expectations for entertainment benefits.

Second, examination of the average customer engagement of brand communities on social media in each industry revealed a significant difference in average engagement between FMCG and IT and media. The total average engagement of selected brand communities was 0.72%, and the average engagement of brands in FMCG was low at 0.42%, whereas that of brands in IT and media was high at 1.04%. This difference is due to the influence of industrial characteristics on average engagement since the IT and media industry is relatively familiar with using new media (Ben Elowitz & Charlene Li, 2012) and there is a difference in the value perceived by users in each industry (Gronholdt, 2000).

Third, an examination of the relationship between actual users’ satisfaction and engagement in relational benefit types by industry confirmed that the satisfaction level in benefits varies among industries. Satisfaction in social benefits was high in manufacturing, satisfaction in economic benefits was high in FMCG, and satisfaction in entertainment benefits was high in IT and media. In addition, this study examined the types of benefits that can increase engagement in each industry. In manufacturing, there is the highest probability that engagement will increase if there is great satisfaction in customization benefits, while in FMCG, there is the highest probability if there is great satisfaction in economic benefits. In IT and media, engagement is more likely to increase if there is great satisfaction in benefits compared with that in other industries. Satisfaction in entertainment benefits had the greatest influence. The results are summarized in Table 14.

<table>
<thead>
<tr>
<th>Relational benefits expected by potential users</th>
<th>Manufacturing</th>
<th>FMCG</th>
<th>IT &amp; Media</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average engagement</td>
<td>0.72 %</td>
<td>0.58 %</td>
<td>0.42 %</td>
</tr>
<tr>
<td>Relational benefits with highest user satisfaction</td>
<td>Social benefits</td>
<td>Economic benefits</td>
<td>Entertainment benefits</td>
</tr>
<tr>
<td>Benefits that are most effective in increasing customer engagement</td>
<td>Customization benefits</td>
<td>Economic benefits</td>
<td>Entertainment benefits</td>
</tr>
</tbody>
</table>

Table 14. Summary of Research Findings
This study confirmed that the average engagement of brand community users varies among industries and the benefits expected from brand communities also vary according to industrial characteristics. The benefit types that satisfy users in brand communities also vary among industries, revealing that benefits associated with great satisfaction are not always effective in increasing customer engagement. For example, manufacturing showed the greatest satisfaction in social benefits among the types of relational benefits, but the effective benefits for increasing engagement are customization benefits. In the end, it was confirmed that the benefits effective in increasing customer engagement in all industries are equivalent to the relational benefits most highly expected by potential users. As a result, if firms in manufacturing make efforts to increase satisfaction in customization benefits, which are most highly expected by users, then it is likely that customer engagement will increase. As for FMCG and IT and media, firms are doing a good job in satisfying users with benefits that are effective for increasing engagement. Moreover, for manufacturing and IT and media, the group with low engagement showed great satisfaction in economic benefits, confirming that satisfaction and engagement show a negative correlation. Considering the results of Research Question 1, in the brand pages of the two industries, potential users’ expectations for economic benefits were not relatively high, thereby indicating that an increase of satisfaction in economic benefits may not show a positive correlation even with an increase in engagement.

Based on these research findings, this study holds significance as it empirically confirmed the relationship between the types of relational benefits and engagement by examining brand communities that are actively operated in Korea, actual users, and potential users. In particular, this study suggested that obtaining relational benefits is the users’ intention in using brand communities, based on the uses and gratifications theory, and clarified the size of the relative influence for each benefit type by using logistic regression analysis. Thus, this study provided baseline data on elements to focus on to increase customer engagement in the brand pages of industries. Furthermore, since the benefits expected or responded to by users vary among industries in brand communities on social media, this study suggested the need for proper management. Therefore, providing the benefits desired by users will be more effective in increasing engagement than simply increasing user satisfaction.

Considering the current status in which the interest and investment of firms in social media are increasing, and social media users are still increasing despite the many users that have already been secured, this study may be used as a source of data to effectively and efficiently manage brand communities on social media. However, studies on social media have only recently begun to be relatively active, and engagement indicators on social media are not yet standardized and are defined differently depending on the view towards evaluation criteria or the perspective of those who interpret them. This study used the engagement formula calculating the engaged users provided by Facebook to collect the customer engagement data of multiple brand pages. This may have acted as an indicator reflecting customer engagement, but it has limitations as the definition of engagement may vary depending on subjective interpretation. Thus, future research may be conducted on indicators that may reflect user engagement more appropriately, and they may discuss the ways to measure user engagement based on data from social media.
References


Ben Elowitz and Charlene Li. 2009. ‘The world’s most valuable brands. Who’s most engaged?’. Wetpaint and Altimeter: 33.


DECISION SUPPORT SYSTEM FOR MULTIOBJECTIVE SUSTAINABLE MARINE SHIPPING

Saowanit Lekhavat, Brunel Business School, Brunel University London, saowanit.lekhavat@brunel.ac.uk
Nursen Aydin, Brunel Business School, Brunel University London, nursen.aydin@brunel.ac.uk
Habin Lee, Brunel Business School, Brunel University London, habin.lee@brunel.ac.uk
Ahmet Bilici, Promatech Maritime Technologies, Turkey, almet.bilici@promatech.com.tr

Abstract

Nowadays it is important to consider sustainable factors in maritime shipping. This is because they are resulting in higher greenhouse effect. One of the factors is fuel emission that may lead to more carbon dioxide emission. In order to reduce fuel emission, one possible way is to optimize vessel speed, but this method may lead to lower service level. Trading off those conflicting relationships is important for planners. Therefore, many existing literatures focus on modelling this problem to be similar to the real world situation and the algorithm that provides the optimal speed within reasonable time. However, in practice, the decision makers also need a decision support system that helps them obtain the optimal speed. Thus, this paper presents a decision support system that seeks to suggest the speed in balancing fuel emission with service level for marine shipping.

Keywords: Decision Support Systems (DSS), Multi-Objective Particle Swarm Optimization (MOPSO), Marine Shipping, Fuel Consumption.

1 INTRODUCTION

Vessel scheduling is one of the complex problems for a maritime planner to make a decision. This is because, for example, cargoes have to be assigned to available vessels in the fleet and then all the vessels in the fleet need to be scheduled to arrive in the ports within time windows. In order to assign cargoes to vessels, the capacity constraint of each vessel should also be taken into consideration. Moreover, there are also some other uncertain factors which may affect the arrival time at the ports. For example, bad weather may delay operational time at ports or sailing time in the sea. As a result, if a planner cannot plan the schedule appropriately, it may lead to an increase in operational cost due to arriving late to a port and delivering cargoes later than planned schedule. Therefore, there is a need for tools to aid ship planners in making operational decisions.

Decision support system (DSS) is a computer-based information system to help its users in decision making activities. Turban (1993) stated that the DSS supports the decision makers but it does not replace them. This is because it has semi-structured decisions where the analysis part is provided by a computer and the decision makers also can use their experience to make a decision. It should be noted that the analysis section consists of modelling techniques combining with solution approaches to guide the solution to the optimal ones. Apart from optimization part, there is a good connection between database, solution approach and output technique. Thus, this paper emphasizes the importance of the use of DSS in supporting vessel scheduling problem.

Another important issue that the ship scheduling should take into account is sustainability. Minimization of carbon dioxide (CO₂) emission is indispensable in marine shipping nowadays. This is because maritime shipping is the largest international transportation mode which is also responsible for around 3.3% of the world emission in 2007 (Buhaug et al., 2009). Consequently, it results in
climate change which is a change in temperature pattern, sea level and precipitation having an adverse influence on forestry, agriculture, ecosystems, energy production as well as consumption and human health (Buhaug et al., 2009). Therefore, consideration of the environmental factors in the vessel scheduling context is highly significant.

International Maritime Organization (IMO) focuses on reducing the emission of CO$_2$ in the shipping industry. It has set a target to reduce this emission by 2018, 2020 and 2050 up to 15%, 20% and 50%, respectively (Buhaug et al., 2009). In order to respond to this policy, many studies in the Maritime literature pay attention to reduce CO$_2$ emission in shipping operations by reducing the speed of vessels. For example, Fagerholt et al. (2010) studied the deterministic non-linear of a liner shipping to find the optimal speed for a vessel in order to minimize fuel consumption in a given time window. If a vessel cannot arrive at one port within the time window, it may propagate to the remaining ports in the same voyage. As a consequence, the vessel may not satisfy the service level of the liner shipping company.

According to the conflicting variables of fuel consumption and service level mentioned above, it is more suitable to consider both these factors as multi-objective function. As a result, the decision maker is benefited with having more than one option choice. Qi and Song (2012) studied multi-objective functions (weighting method) to find the optimal speed by considering uncertain port time in order to minimize total fuel consumption. For example, they compared the case where a 100 per cent service level was maintained with another case where they sacrificed certain amounts of service level. The result of the study highlighted that the case where service level was sacrificed showed that it can reduce a large amount of fuel consumption. Therefore, the consideration of multi-objective functions is another important element in scheduling vessels.

Although there are many studies focusing on multi-objective green vessel scheduling problem, none of them apply DSS to aid the decision maker in planning the scheduling context (Mansouri et al., 2015). Thus, the aim of this paper is to present the DSS for deciding the near optimal speed that can minimize fuel emission and maximize service level in scheduling a vessel. The rest of this paper is presented as follows. Section 2 presents the literature review about DSS in vessel scheduling problem. Section 3 demonstrates the architecture of DSS. In section 4, modelling, MOPSO solver and user interface are illustrated. Finally, conclusion and recommendations are discussed in section 5.

2 Literature Review

As highlighted by Mansouri et al. (2015), the majority of the previous studies in sustainable ship scheduling problem pay attention to only mathematical modelling and algorithms to solve the problem. Existing literature on DSS for vessel scheduling is relatively scarce and therefore this paper seeks to fill this void in the literature.

Kim and Lee (1997) were the pioneers of using optimization-based DSS for scheduling vessels. The authors created it for scheduling bulk cargoes by assigning them to a schedule in tramp shipping. LINDO optimizer was used as a tool in scheduling process in order to maximize the profit obtained from the transportation revenue of cargoes when the operating cost of vessel had been already reduced. Another similar scheduling of bulk problem in tramp shipping was proposed by Bausch et al. (1998). The authors tried to assign cargoes into the scheduling process so that all loads are transported at a minimum cost and satisfy all limitations such as time window and compatibility between port and vessel. The output of this optimisation process was presented as a schedule on a spreadsheet for the users to track it easily.

Since the study by Bausch et al. in 1998, there has been a lack of literature relating to the DSS use in vessel scheduling problem until a study by Fagerholt (2004). According to Fagerholt (2004), one of the interesting reasons why managers in marine shipping were not willing to use DSS is that because the managers thought that DSS cannot consider all of the constraints in the scheduling process. So as to address this problem, a DSS called TurboRouter was introduced.

Fagerholt and Lindstad (2007) continued developing TurboRouter to meet all the requirements for vessel scheduling problem in industrial and tramp shipping. The researchers extended the DSS
developed by Norwegian Marine Technology Research Institute. Time windows, vessel capacities, compatibility between port and vessel, bunker consumption rate, bunkering port calls are taken into account for planning the vessels to arrive at port within specific time period and with the maximum profit. As a result, the decision maker can easily see the schedule through user interface. TurboRouter also receives satellite positions from ships in real time and computes the estimated arrival times to given ports.

Apart from industrial and tramp shipping, Lam (2010) focused on designing DSS for scheduling liner shipping problem. Lam (2010) first used an integrated approach to select the location of ports and then scheduled vessels to arrive at ports within a given time window and finally financial factors were analysed. In scheduling process, a planner can add, delete, insert, edit and move the port manually and then the distance and time window of every port are updated automatically. Finally, the optimal schedule is provided for a decision maker to minimize all related costs.

Since environmental concerns plays an important role in maritime shipping, existing studies on DSS for vessel scheduling have focused on minimising CO$_2$ emissions. For example, Ballou et al. (2008) presented a DSS called Voyage and Vessel Optimization Solutions (VVOS) in order to schedule vessels to reach port of calls with minimum CO$_2$ emissions within a given time window when vessels face uncertain ocean condition such as wind, wave and currents. VVOS is considered to be user friendly as it is flexible for the user to choose whether they would like to use optimization module. Similarly, Windeck and Stadtler (2011) also focused on developing DSS for routing and scheduling problem in order to minimize cost and CO$_2$ emission by considering weather factors.

However, in the case of vessel scheduling, CO$_2$ emission can also be minimised by applying these four strategies to trade off bunker consumption and service level of port of calls in the route (Brouer et al., 2013). The first strategy which may be chosen is to omit the upcoming port of calls. Second, swapping the ports of calls and accept the delay of some ports to catch up the remaining port of calls. The disadvantage of these three strategies may be losing credibility of liner shipping from some ports, but may be safer in terms of fuel consumption. Finally, a vessel tries to speed up to reach the port within time window. The advantage of this option may be in light of increasing service level, but need to spend more fuel consumption. Therefore, there is a need to find optimal speed that can minimize fuel consumption (or total costs) and maximize service level.

Andersson et al. (2015) demonstrated the integrated fleet deployment and speed optimization in order to minimize total relevant costs such as cost at port, canal and sailing, the charter costs for spot ships, the space charter costs, and the penalty costs for delayed start of the voyages of tramp shipping. The following data are determined; the sequence of the ship, the sailing speed, the start time of each voyages, the optional voyages to sail and others so that all ports of calls are serviced within their time windows and all demands are met.

Moreover, Qi and Song (2012) proposed multi-objective functions to find optimal speed that can minimize fuel emission and penalty of vessel delays. The authors suggest that it may be better to sacrifice certain service level to gain much more reduction in CO$_2$ emissions. Nevertheless, the drawback of weighting method multi-objective approach to handle multi-objective function is that it has a narrower scope of feasible region. Therefore, in order to achieve near optimal solution, different values of weight is needed to be set. Accordingly, it may be time consuming to adjust the weighting parameters.

In order to manage this issue, Pareto concept is another existing multi-objective approach that could be applied to deal with multi-objective sustainable vessel scheduling problem. Pareto or non-dominated fronts are the sets of decision vectors (x) that are not dominated by any other decision vectors (y). It should be noted that decision vector $x$ dominates a decision vector $y$ (also written as $x \prec y$) if and only if:

\[ i \{1,2,...,M\} \land i \{1,2,...,M\} | z_i(x) < z_j(y) \] (Collette and Siarry, 2003).

The Pareto front is an example of Pareto front in case of two objective functions.
According to the advantages of DSS and non-dominated front mentioned above, it is useful to build DSS when there are conflicting objectives between trying to minimize CO$_2$ emission and trying to maximize service level. Moreover, there still is no literature on integrating DSS with sustainable factor for multi-objective functions to ship scheduling problem. Hence, this paper emphasizes that the DSS can aid captains of the ship to decide near optimal speed that can minimize fuel emission and maximize service level. The benefit of this research study does not only fill the void in the literature, but also responds to the reduction of carbon dioxide emission according to IMO policy.

3 ARCHITECTURE OF DSS

This section explains overall conceptual framework of DSS architecture. DSS architecture composes of two components; frontend and backend. Frontend section is the section that interacts with the user. It consists of four elements. First, model finding function supports users to identify the best optimisation model type for their in-hand problems. Second, model building function helps user to customise a chosen model type to be ready for a solver. This includes adding and removing decision variables and revising objective function according to the specialised needs of users. Third, model solving interacts with model solver in the backend to get optimal solution. Finally, What-If analysis tool helps user conduct sensitivity analysis of the optimal solution to test various scenarios. This frontend can interact with user via user interface based on web based and the front-end interacts with the back-end components via web service protocols. All modelling and what-if analysis conducted by a user is saved in History Base for later reuse and auditing. The architecture of DSS is depicted in figure 2.
4 MODEL SOLVER AND USER INTERFACES

Following the architecture of DSS in section 3, this section presents the implementation of DSS for multi-objective vessel scheduling problem via user interface and model solver in order to find near optimal speed to maximize service level and minimize fuel consumption. The user interface, model and solver module are presented in the subsequent sections.

4.1 Input Interface

Prior to accessing to the input interface, the planners such as operational team have to login first. The login window is shown in figure 3.

Figure 3. Login Page
Then the main page which shows the map for seeing vessel list and tracking a vessel will appear. The user can select the region on the map to monitor many vessels in all continents around the world as well. This main page can be seen in figure 4.

![Main page](image)

*Figure 4. Main page*

If the user clicks the vessel list button, it will display the page of vessel list. This page comprises of vessel name, the number of ports and port name which corresponding vessel has to go for loading or unloading purpose. For example, vessel Z has to go to 12 ports in a voyage starting from port D0 to D11. It should be noted that in the figure 5, it displays only port D1 to D11 because D0 is considered as a starting port.

![Vessel list](image)

*Figure 5. Routing of each vessel*

Then, if a user selects one vessel such as vessel Z, the data relevant to that vessel will appear, namely, a captain name of a vessel, current location which is latitude and longitude, ports that a vessel has just departed to the next coming port and the current speed as shown in figure 6.
After that, a planner has to upload all of the necessary information in order to find (near) optimal speed that can minimize fuel consumption and maximize service level. The data needed include distance between each port, a vessel’s designed speed from the minimum to the maximum, the earliest arrival time and latest arrival time which each port allows a vessel to berth, actual port time which comprises berthing, loading, unloading and ‘unberthing’ time and fuel consumption rate during staying at the port as demonstrated in figure 7. When the input data has already been updated, a planner can run the model by clicking run solver button as shown in figure 6 so as to obtain an optimal speed that minimizes fuel emission and maximizes service level.
Finally, after activating the run solver, the multi-objective sustainable vessel scheduling model and optimization engine will be implemented. Thus, the next section describes the model and the optimization tool embedded in this DSS.

### 4.2 Model and Optimization Engine

Ship speeding decision for a vessel calling at \( n \) ports along the route is conducted so as to minimize fuel emission and maximize service level. The mathematical model that is embedded in this DSS is presented as follows.

**Indexes**
- \( j \): Leg \( j \); \( j = 0, 2, \ldots, n - 1 \)
- \( n \): Number of ports-of-call

**Parameters**
- \( \xi_j \): Vessel’s port time at \( j \) ports-of-call
- \( \alpha_j \): Earliest planned arrival time at \( j \) port-of-call
- \( \beta_j \): Latest planned arrival time at \( j \) port-of-call
- \( v_{\text{min}} \): Vessel’s minimum speed (the most fuel efficient speed)
- \( v_{\text{max}} \): Vessel’s maximum speed (a technical constraint)
- \( d_j \): Distance between port \( j \) and \( j + 1 \) port-of-call
- \( t_a^j \): Arrival time at \( j \) port-of-call
- \( t_e^j \): Earliest possible arrival time at \( j \) port-of-call
- \( t_d^j \): Departure time from \( j \) port-of-call
- \( g^v_j \): Fuel consumption rate (per nautical mile) in leg \( j \) at speed \( v_j \)
- \( \phi \): Idle fuel consumption rate (per minute) during berthing, loading, unloading and “unberthing”
- \( s_j \): Service level at \( j \) port-of-call
- \( G \): Total fuel consumption along the route
- \( S \): Service level along the route

**Decision Variables**
- \( v_j \): Vessel’s average speed at leg \( j \); \( j = 0, 2, \ldots, n - 1 \)

**Objective functions**

\[
\text{Minimize } G = \sum_{j=1}^{n} \left( d_j \cdot g_{\text{\phi}}^v + (t_d^j - t_e^j) \cdot \phi \right) \tag{1}
\]

\[
\text{Maximize } S = \sum_{j=1}^{n+1} s_j \left( n + 1 \right) \tag{2}
\]

**Subject to**

\[
g_{\text{\phi}}^v = 0.0036v_j^2 - 0.1015v_j + 0.8848 \quad \text{if } \frac{t_a^j - t_e^j}{t_d^j} \leq \beta_j \tag{3}
\]
In the model, we consider two objective functions. The first one is to minimize fuel consumption during sailing in the sea, berthing, loading, unloading and “unberthing” as shown in equation (1). In the sailing part, the fuel emission depends on the speed of a vessel along the route shown in equation (3) while a vessel stays at the port, it is subject to idle fuel consumption rate. As seen in equation (2), the second objective function is to maximize service level at the ports. We can measure the service level by comparing the actual arrival time with the latest planned arrival time. If it is less than or equal to the latest planned arrival time, the service level is assigned to 1; otherwise, it is 0 as presented in equation (4). Moreover, in order to calculate the total time along the route, equation (5) assumes the arrival time at port 0 to be 0. The departure time at port 0 can be seen in equation (6). For the arrival time and departure time from port 1 to port \( n \), they can be calculated by equation (7) and (8), respectively. Finally, the optimal speed between the minimum and maximum velocity as presented in equation (9) will be obtained.

In order to solve this model, one of the efficient and effective algorithms is MOPSO (Nguyen and Kachitvichyanukul, 2010). Thus, this DSS applies MOPSO with personal knowledge of the swarm to solve the sustainable vessel scheduling problem.

![MOPSO optimization engine](image)

**Figure 8. MOPSO optimization engine**

The procedures of this method begin with reading input data related to the vessel scheduling problem and MOPSO parameters from the user. The input data relevant to vessel scheduling has already been obtained via the user interface as stated in section 4.1. Moreover, the input parameters related to MOPSO are also set as follows.

- Number of iterations = 150 iterations
Number of particles in the swarm = 12 particles
Minimum value of inertia weight = 0.4
maximum value of inertia weight = 0.9
Number of neighbour in the swarm = 0.5
Personal best position acceleration constant = 1
Global best position acceleration constant = 1
Local best position acceleration constant = 1
Neighbour best position acceleration constant = 1
Maximum number of particles in elite group = 10
Number of replication = 10

After inputting the parameters, the process tries to map such a problem to MOPSO by a direct solution representation way. This means that each position of particle can be represented by the speed of a vessel. Following this, the process moves to evaluate the fitness values of the mathematical model. The fitness values which are fuel consumption and service level are then achieved. The next step is to determine global non-dominated front and copy some particles in non-dominated front into elite group so that the particles in this group help guide the next position of the particle. After determining the Pareto front, the stopping criteria of the algorithm is checked. If it meets the stopping criteria, the optimization process is ended. On the other hand, if the stopping criteria are not yet met, the next phase of the procedure is to select a particle from the elite group to be a guidance of the movement. Finally, the particles move to a new position with the velocity that contains the term of guidance in elite group. The output of this stage is the new input value for sustainable vessel scheduling problem and to evaluate again. The process repeats until the stopping criteria are met. All of these stages are shown as figure 8.

In order to link vessel scheduling problem with MOPSO, the solution representation is needed to be decided. This DSS applies the speed of the ship in each leg directly to each dimension of each position, position best, local best and neighbour best of particle as shown in figure 9. Then the personal knowledge from movement strategy 0 presented by (Nguyen and Kachitvichyanukul, 2010) is used as a guidance to the optimal solution.

![Solution representation of marine shipping problem in MOPSO](image)

### 4.3 Output Interface

Once the program finishes running, the output page which recommends the fuel consumption and service level will appear. All of the non-dominated points are shown in figure 10. For example, in this case, there are four near optimal solutions for a planner to choose whether he/she pays more attention to maintain high service level or fuel consumption. If the planner gives priority to service level, he/she may select the speed for each leg that provides service level equal to 75 percent although this plan will
consume more fuel consumption which is equal to 640.36 tonnes. However, if a planner would like to save more fuel consumption, he/she may choose the speed that gives 516.32 tonnes but the service level is only 50 percent and other choices are also provided. This can be done by clicking show button in speed and scheduling column. Then the speed and schedule which tells about the actual arrival and actual departure time as shown in figure 11 will appear.

![Nondominated solution](image10)

![Speed of each leg and the Expected Arrival Time](image11)

5 Discussion and Conclusion

Since speed decision in ship liner scheduling problem is a complex problem which cannot be determined by manual computation, DSS equipped with effective and efficient algorithm as well as user friendly interface is needed. In this paper, the DSS embedded with MOPSO algorithm can provide many possible speeds that minimize fuel consumption and service level for the decision maker. Moreover, it is also convenient for the users to add the speed by using their own experience to check whether the better results will be obtained.

However, the result obtained from simulation optimization algorithm embedded in this DSS tends to provide better schedules than those decided by experience of a captain. Another advantage is that it also provides many choices of the schedule for a captain to choose the appropriate speed whether that
port should pay more attention to service level or to fuel consumption due to Pareto approach in handling multi-objective function. According to this point, Mansouri et al. (2015) highlighted that there is a lack of studies surrounding multi-objective sustainable vessel scheduling problem; therefore, we consider it as one of the key contributions to the literature.

Finally, the DSS presented in this paper not only benefits maritime shipping practitioners in terms of their economics and credibility but also enhance environmental sustainability to the world. In conclusion, this paper focuses on the description of DSS for deciding the optimal speed of a vessel in marine shipping that can minimize fuel consumption and maximize service level.

The recommendations for further research are to consider other relevant factors that have an effect on the speed of a vessel. Factors such as weather condition, shortest distance, vessel’s weight due to container weight, the ship weight and uncertain port time need to be considered so that the model will be more similar to the real world problem and be able to provide more accurate speed and schedule to the planner.

References


FACTORS INFLUENCING ADOPTION OF GREEN SHIPPING PRACTICES AND INFLUENCE ON ORGANISATIONAL PERFORMANCE

Oluwakayode Aluko, email: cbpgoaa@brunel.ac.uk, Brunel Univeristy, UK
Habin Lee, email: Habin.Lee@brunel.ac.uk, Brunel Univeristy, UK

Abstract

Green shipping Practices (GSPs) are a new and increasing trend in the shipping industry. This appears to be a response of the shipping industry to the increasing demand for sustainability in this industry sector. While these practices are gradually becoming acceptable, it is unclear what motivates shipping firms to engage in these practices. This research proposes Institutional Theory as an organizational tool to help provide a generic understanding of the subject area. The research ultimately seeks to understand what drives the adoption of GSPs. Through robust literature review the conceptualization of GSPs as presented by (Lai et al, 2011) is further developed into a proposed conceptualization model. This research proposes a case study approach to exploring the subject area. This is because the research area being explored is still very new and there is little data and literature in this area. It is expected that the findings from this research will be helpful for developing a systematic theoretical framework to help the successful implementation of GSPs.

Keywords: GSPs, adoption, Institutional Theory
1. INTRODUCTION

The clamor for sustainable shipping has been on for over a decade now. Very recently, actions are now being taken to mitigate maritime emissions. These actions are in the form of policies and strategies that strive to reduce maritime emission through the cooperative interaction of technical, operational and market forces. Foremost in the campaign for maritime sustainability is the International Maritime Organization (IMO), an arm of the United Nations (UN) that is responsible for ensuring sustainability in maritime shipping. A recent IMO study carried out in 2009 showed that if actions are not taken maritime emission is likely to increase by up to 250% in the year 2050. On the contrary however, the successes of proposed actions are expected to reduce maritime emission by as much as 75% of its present value which will be a significant achievement in the strides towards maritime emission reduction. Several academics are also in agreement with the prospect of sustainable shipping as will be seen in subsequent sections of this research.

On the other hand, the shipping industry is beginning to respond to environmental pressures for sustainability through the adoption of Green Shipping Practices (GSPs) but this is still in the early stages and the motivation behind this new trend is still unclear (Lai et al, 2011). The dearth of literature in this regard as identified by Lun et al, (2011) further stresses this point. This research through industry case study seeks to fill the existing void in literature by exploring what factors influence the adoption of GSPs and how this affects organizational performance (service and financial). This knowledge will be valuable for the development environmental policies targeted towards the shipping industry (Lai et al, 2011), as well as development of a systematic framework for the adoption of GSP, which will also contribute to the limited volume of literature in this regard.

It has been recently observed that organizations are increasing engaging in voluntary environmental measures on the presumption that such actions will yield both environmental and economic returns (Lai et al., 2010a). Studies on green practices in different industrial sectors however show that many firms are still lukewarm in this regard (Zhu et al., 2008) owing to many reasons the most prominent of which is the lack of strong incentives for adopting green practices to justify the investment of the required time and resources. In the shipping sector, there are evidences of environmentally friendly actions of shipping firms; however, there is a lack of systematic framework for the adoption of green practices in the shipping industry. This research is further motivated by the presumption that contrary to the general belief that firms mainly act to maximize financial gains (Smith and Grimm, 1987) there may be other contributory reasons for firms adopting Green practices. This research seeks to investigate what other factors (asides economic incentives) influence the adoption of GSPs. The main research question to be answered is; why are shipping firms beginning to adopt GSPs? In order to answer this question, the following questions will have to be answered: What is responsible for the adoption of GSPs (e.g. institutional variables i.e. external/internal pressures such as regulations, social norms and customer pressures)? If yes, what are those “institutional variables” and how are they contributing to the diffusion of GSPs? And if not, what theory can adequately explain this trend? What is the effect of the adoption of GSPs on firm performance (service & financial performance)?

The peculiarity and complexity of the shipping industry makes the problem of adoption of environmental practices even more demanding. These complexities are largely economic,
regulatory and global and can possibly have butterfly effect across global supply chain. This largely constrains the possibility of implementation of environmentally friendly practices to organizational culture basis. With the increasing demand for environmental accountability it has become imperative for shipping industry to comply through the implementation of some form of environmental friendly practices, which is still in the early stages. The motivation behind this is however elusive. While it could be assumed that it is due to regulatory imposition, it is unclear how, if and what other factors contribute to this new and growing trend. This research looks into investigating the motivation behind GSPs adoption in the shipping industry and the possible impact of this on firm performance.

Having already stated that the complexity of shipping industry largely constrains the implementation of environmental policies to organizational level we presume that organizational theory will be helpful for understanding this trend. This opinion is also supported by Lai et al, (2011) who proposed institutional forces, which are elements of institutional theory as a possible explanation for the adoption of GSPs. This research takes a clue from their proposition, which is however not yet validated as a starting point. The pending question here is whether or not institutional theory will be able to fully explain what factors are responsible for the adoption of GSPs and to what extent. Further more, if institutional theory is unable to explain the factors responsible for the adoption of GSPs, what other theory can and finally if no theory can explain this trend, this research will attempt to propose a viable theoretical framework.

The rest of the document is organized as follows; section two covers literature review as well as theoretical background for the research. This leads on to section three, which is the methodology adopted for carrying out the research. Section four is the proposed conceptualization of the adoption of GSPs while section five presents a conclusion of the paper with expected results.

2. LITERATURE REVIEW

2.1 Green Shipping Practices

The definition of GSPs is still maturing. GSPs has been considered as technological efforts made by shipping firms that focus on efficiency in energy consumption hence minimizing waste and negative impact on environment (Krozer et al., 2003). It has also been considered as business practices that improve environmental performance of shipping firms which are sometimes demanded by stakeholders e.g. obtaining ISO 14000 certification (in compliance with the ISM Code) (Celik, 2009). (Lai et al, 2011) simply defined GSPs as the "sustainable handling and distribution" of cargoes and although there are seemingly diverging perspectives to GSPs there is convergence in the fact that they are helpful for minimizing negative environmental impacts that result from shipping activities.

(Lai et al, 2011) pioneered studies on GSPs, building on previous studies in environmental management (Lai et al., 2010a; Zhu et al., 2008), they conceptualized GSPs identifying six dimensions in which GSPs can be effectively illustrated. These six dimensions are as follows: Company policy and procedure (CPP), Shipping documentation (SD), Shipping equipment (SE), Shipper cooperation (SC), Shipping materials (SM) and Shipping design and compliance (SDC). They presume that the effective adoption of GSPs would have to encompass these dimensions. This conceptualization however requires empirical validation, which is a part this research’s contribution. This research aims to provide empirical validation
to Lai et al's conceptualization as well as study/explore case studies to understand what factors motivate the adoption of green shipping practice.

As already stated in the introduction, this research begins its exploration from an organizational perspective proposing organizational theory as a possible explanation for the adoption of GSPs. We therefore provide a background on Organizational theory. Sarkis et al, (2010) attempting to define Organizational theory stated that it is not easily definable. This is because it has been influenced by several other disciplines (sociology, political science, engineering etc.) (Hatch, 2006;Pfeffer, 1997). They however still managed to put forward the following definition: “management insight that can help explain or describe organizational behaviors, designs, or structures”. Management studies as a whole has greatly benefitted from application of Organizational theory (Etzion, 2007;Hoffman and Ventresca, 2002) however application in environmental management is still in early stages (Sarkis et al, 2010).

Studies show that the application of organizational theory to operations management is only beginning to be rampant (Ketchen and Hult, 2007). This research taking a cue from Lai et al (2011) begins its exploration of the adoption of GSPs from an Organizational theory (Institutional theory) perspective. This is because GSPs is just emanating hence a holistic view will be helpful to develop a generic undersatnding of the subject area. This opinion is also upheld by (Scott 2003) who stated, “Institutional perspective provides a useful theoretical lens to study the organizational response to environmental issues” because it sheds light on forces beyond organization that compel organizations to respond to its interest. There is evidence of the application of Institutional Theory to environmental management practices (SCM, IT etc.) and it has proven to be helpful.

Institutional theory is one of the profound organizational theories, which studies the influence of internal/external pressures on an organization. It is valuable for understanding how organizations attend to green issues (Jennings and Zandbergen, 1995, Campbell 2007; Chen et al. 2008;) and it is becoming increasingly relevant in explaining organizational perspectives to environmental management practices (Lounsbury, 1997). Institution theory assumes that the existence and functionality of organizations is embedded in social networks (Lin and Sheu, 2012) hence firm behavior is largely influenced by its social interaction, which is also a function of interrelated organizational networks (Anderson et al. 1994; Iacobucci and Hopkins 1992; Meyer and Rowan, 1977; Scott, 1987). Institutional theory tries to explain the reason for observed similarities amongst firms in a specific field (DiMaggio & Powell, 1983). According to institutional theory, institutions are “regulative, normative, and cognitive structures and activities that provide stability and meaning for social behavior” and they are a vital part of an environment (Scott, 1995, p. 33). It is observed that institutions exert a constraining force (isomorphism) on organizations, which cause organizations in a specific class to act in a similar pattern as other under the same type of influence (Hawley, 1968). According to (DiMaggio & Powell, 1983), three types of Isomorphic influences are identified. These are coercive, normative, and mimetic. Coercive isomorphism translates to pressures from entities upon which firms/organizations depend for resources. Mimetic isomorphism arises as a result of firms/organizations imitating or copying other successful ones. This happens when firms are in uncertain situations. Normative isomorphism is as a result of firms having to follow standards and/or practices as established by certified organizations (certification and training methods, professional networks etc.).

In Supply Chain Management, Institutional theory has been used to study the adoption of Green Practices and impressive findings have been made showing the effect is coercive pressures ((Rivera, 2004), (Clemens and Douglas, 2006), (Kilbourne et al., 2002)), normative
pressures from customers and the market in general (Carter et al., 2000), Ball and Craig (2010), Harris, (2006), Christmann and Taylor, 2001) and mimetic pressures (Aerts et al., (2006), Zhu and Liu, (2010), Christmann and Taylor, (2001)). There are similar studies in Information Technology & Information System (Orlikowski et al. 2001, Tingling et al. 2002, Teo et al. 2003, Liang et al. 2007). (Murugesan 2008) emphasized the legal, ethical and social obligation we have to green our IT processes. (Adela et al. 2009) found that coercive and mimetic forces played important roles in IT and IS adoption. The influence of normative pressure was omitted in their model because they could not differentiate between its effects and mimetic effects. (Chen et al. 2010) cited (Lampe et al. 1991) to have suggested that a combination of pragmatic (e.g. financial and legal) and idealistic (e.g. moral and ethical) are mostly responsible for firms adoption of Green IT.

Institutional theory is taken as a guide into exploring factors influencing adoption of green practices in shipping and based on this a proposition of adoption is presented in subsequent sections. This research does not emphatically posit Institutional theory as responsible for the adoption of GSPs; it however takes it as a starting point in exploring the adoption of GSP. This research adopts an exploratory/explanatory case study (qualitative) approach, which allows for the introduction of a theory even though it is interpretivism perspective. More on this is discussed in the methodology section.

2.2 Relationship between environmental performance and firm performance

There is some evidence of relationship between environmental and firm performance in literature. It started as a conceptual idea however it has gradually progressed into empirically proven facts over time mostly in the manufacturing sector. (Porter, 1991) was one of the first to conceptualize this idea using the terminology “win–win”, which at the time challenged the traditional mindset that environmental regulations are a burden to the health of organizations; Porter’s arguments were based on observatory evidence, which suggested that there exists benefits for environmental practices implemented by firms and these benefits outweigh the cost of implementing these practices hence stricter regulations are likely to inspire innovation which will eventually translate into improved firm performance. This notion was carried forward by (Porter and van der Linde 1995a) through the idea of “innovation offsets”. The limitation to their works was that their studies only took into account environmental regulatory influence, which is only one dimension of influences/pressures that organizations are subjected to.

There has been further works in this regard with the contribution of other authors who have considered the influence of broader factors. Berry and Rondinelli (1998) strongly emphasized the immense value of firms’ internal environmental strategies and its effects on firm performance stating lowered regulatory cost and new business opportunities as benefits of environmental proactivity. Hanna and Newman (1995) also corroborated the “win-win” and similar arguments were also put forward by ((Royston, 1980), (Bonifant, 1994), (Bonifant and Ratcliffe, 1994)) albeit without empirical proof. Florida (1996) provided empirical validations that supported Porter’s “win-win” arguments using surveys in the manufacturing sector. Similar finding were made by (Klassen and McLaughlin (1996), Russo and Fouts, (1997), Klassen and Whybark, (1999) and Orlitzky et al. (2003)

All the studies presented above all show that relationship exist between environmental and firm performance. Empirical evidence exists for this relationship in the manufacturing sector. The influence of external (e.g. regulatory, customer etc.) and internal (managerial) pressures has also been considered. This research will look to find if similar relationship exist in the shipping sector.
3.0 METHODOLOGY

This section discusses the methodical approach of this research. It briefly describes the existing research paradigms with emphasis on the paradigm considered for his research. This research takes a case study approach due to the current early stages of the subject area. The use of this approach is further justified because case studies are used to explore single phenomenon, which in this case is the adoption of GSPs. Other requirements of the use of case study are also satisfied by the research idea as will be seen in subsequent sections making the approach legitimate.

Typically, there are two extreme research paradigms; positivism and interpretivism. These two are based on two contrasting opinions on worldview and hence the approach to undertaking research. Positivism emanates from natural sciences and hence upholds a deductive approach to research emphasizing the use of theories to explain/understand social phenomena (Collis and Hussey, 2009, p 56). On the other hand, interpretivism developed as a result of inadequacy of natural science perspective explanation of social phenomena. It focusses on subjectively exploring social phenomena with a view to understand its complexity without the necessary constraints of beginning with a theory as positivism recommends (Collis and Hussey, 2009, p 57). Due to the varying perspective of these paradigms, they require different analytical methods. Positivism, which is mostly associated with large data, requires surveys, cross sectional studies, experimental studies etc. Interpretivism goes along with methodologies such as case studies, ethnography, action research etc.

This research takes an interpretivism position and specifically adopts a case study approach (exploratory/explanatory case study). Case studies are used to explore phenomena in their natural setting and there are different types. Exploratory case studies are particularly used when there are few theories applicable or where a deficiency in body of knowledge exists, which characterizes this research. In addition, an element of Explanatory case study is included in which it is permitted to have an underlying theory for the development of foundational understanding of the research area (Collis and Hussey, 2009, p 82). The subject area being explored is very new and lacks a systematic theoretical framework and literature. Hussey (2009) further suggests that the different types of case studies are much differentiated hence one type can be combined with another.

At the time of this writing, contact is being made three of the foremost shipping firms who have begun to adopt GSPs and they are willing to make themselves available as case studies for the research. The data collection will be embarked upon in three simultaneous run phases, semi structured interviews will be conducted with management level staff of these shipping firms. This will be combined with an analysis of the business process to ascertain if GSPs has been duly implemented and to what extent and finally these will compared in parallel with the companies self-reported performance metrics.
4.0 CONCEPTUAL PROPOSITION FOR THE ADOPTION OF GREEN SHIPPING PRACTICES

From previous sections we already understand coercive pressures to emanate from institutions upon which firms depend for resources or institutions that exercise regulatory influence on organizations. In the shipping industry, foremost amongst this type of institutions is the IMO, which has recently imposed an emission reduction target for the maritime sector as a whole. The International Maritime Organization (IMO), an arm of the United Nations (UN) that is responsible for ensuring sustainability in maritime shipping. This institution is considered as the plausible foremost coercive influence on the adoption of GSPs. In addition to regulatory bodies, Toe et al, (2003) also pointed out that customers could exert coercive forces on organisations hence this influence is also considered in the proposed model.

Normative pressures as has been already discussed arise from formulated norms that somehow compel organisations to act in a certain way. Lee et al, (2013) termed it as “a collective struggle of a people in a certain profession to define the conditions of their work”. DiMaggio and Powell (1983) identifies two sources of normative pressure namely; formal education and growth of professional network of personnel within an organisation. This research will explore the possible influence professional networks and level education of shipping firms as driver of adoption of GSPs.

The influence of mimetic forces has been described to emanate from organisations looking to replicate the success of others particularly when they are in uncertain situations. It is desirable to know whether the adoption of GSPs is driven simply by the shipping industry’s desire to replicate the success of the implementation of Green Practices seen in other business sectors like GSCM, Green IT & IS etc. Having discussed the practical representation of these institutional elements in shipping, the proposed conceptualization of GSPs is presented below followed by the description of the propositions.

4.1 Conceptual Proposition

![Figure 4.0: Conceptual Proposition for the adoption of GSPs](image)

**Proposition 1:** Coercive Isomorphism (Regulatory forces) strongly influences the adoption of GSPs.

Environmental regulations are systematic guidelines that help firms implement environmentally responsible practices. This has been observed to be effective from previous environmental studies that have highlighted the importance of regulations in environmental management (Lai et al, 2011). The enforcement of these regulations imposes “coercive
isomorphism” on organizations. A profound regulatory body in shipping is the IMO, which oversees all maritime regulatory activities all over the world. Recently, IMO modified (MARPOL 73/78) protocol of 1978 to include all forms of maritime environmental pollutions (accidental, chemical, packaging, sewage, air etc.). It has been previously argued that sloppy regulatory enforcements negatively affect firm’s implementations (Economy and Lieberthal, 2007). Hence to reap the benefit of regulations, it must be firmly enforced. It is presumed that coercive isomorphism will influence the adoption of GSPs amongst shipping firms. This is however only an assumption that requires empirical validation as noted by (Lai et al, 2011). In addition to this, it is necessary to know the extent of its influence (absolute or partial).

**Proposition 2: Mimetic Isomorphism (industry institutionalized practices) strongly influences the adoption of GSPs.**

Industry instituted norms/practices often play a profound role in ensuring its own sustainability. This is obvious through efforts of industrial associations. A good example is the Marine Environmental Protection Committee (MEPC), which has often encouraged the implementation of environmentally sustainable practices in shipping. Recently, MEPC proposed for the reduction of carbon footprint of the shipping industry through a program that ensured environmental friendly recycling of vessels at the end of its service life. Maersk Line has already adopted this proposition and being a leader in the shipping business, this can influence other shipping firms to follow suit in a bid to achieve some form of legitimacy as well as measure up to stand a good competing chance. Similarly the shipping and environmental code of practice as proposed by International Chamber of Shipping (ICS) (http://www.marisec.org/ics/index.htm, 8 January 2010) suggests ten environmentally friendly practices to encourage sustainability in shipping. This form of influence is presumed to contribute to the adoption of GSPs as shipping firms can be motivated to adopt GSP due to benefits that can be gained from assistance offered by these organizations. An empirical verification of this type influence is required.

**Proposition 3: Normative Isomorphism (stakeholder influences) strongly influences the adoption of GSPs**

Pressure from stakeholders upon which shipping firms largely depend for the business needs is another factor likely to influence the adoption of GSPs by shipping firms. This pressure is interpreted as normative Isomorphism as shown on the conceptual model above. This pressure is due to increasing awareness of about environmental sustainability. Staffs of shipping firms in a bid to measure up to their professional colleagues and hence attain legitimacy are likely to trigger their organisations to adopt environmental friendly practices in these GSPs. This research will examine if this plays a role in the adoption of GSPs and to what extent.

**Proposition 4: Adoption of GSPs leads to increased firm (financial and Service) performance**

(Lin and Sheu, 2012) clearly pointed out that there is a profound difference between firms submitting to pressures in the adoption of green practices and the exploitation of these practices for the improvement of firm performance. Some research (e.g., Anderson et al., 1999; Ketokivi and Schroeder, 2004; Prajogo, 2011) also suggests that intent to implement certain management practices (e.g. TQM and ISO9000) could actually influence the effectiveness of those practices and the overall performance of the firm. Similarly, (Choi and Eboch, 1998; Prajogo, 2011; Zhu and Sarkis, 2007) examining the influence of internal and external pressures in Green Supply Chain Management found that that “existence and response to internal motives and external pressures are likely to cause the relationships between green practices and performance to vary”. Could this finding be true for the shipping industry? Striking a good balance between these two functions is usually a challenge for organizations (Lai et al., 2008, 2006, 2010b). The influence of the institutional forces
mentioned above will need to be appropriately balanced in order for shipping firm to reap economic gains. This is the striking advantage of GSPs over other environmental management practices in that it strives to create a balance between the reduction of adverse environmental activities and performance gains for shipping firms. This not only increases profitability but also enhances compliance with international regulatory trade requirements hence increasing possibility of business expansion (Lai et al., 2011). GSPs consciously incorporate environmentally friendly practices into the end-to-end process of cargo transportation. This is reflected in decision making at every stage of process. There is some evidence that adoption of environmentally friendly practices can enhance organizational/firm performance (Lai et al., 2010a). (Vachon and Klassen, 2008; Yang et al., 2009b; Zhu et al., 2010) all suggest that some organizations reaped positive performance benefits through the cooperation of suppliers and customers on environmentally related issues. All these assumptions however lack empirical validation particularly in the shipping industry, which is only beginning to adopt green practices hence this research is interested finding out what influence/contribution does the adoption of GSPs have on firm performance.
5.0 CONCLUSION

This section presents an overview of the whole paper. Section one of the paper introduces the research aim and objective with a description of the research context. Sections two presents a literature review and theoretical framework of the research. The proposed methodology for the research is briefly presented in section three.

GSPs are considered a new an increasing trend in the shipping industry and very little is known about what drives this trend. More so, only literature on the subject is available. Through this research it is anticipated that mystery behind the motivation for the adoption of GSP amongst shipping firms would be unraveled. It is also anticipated that the impact of the adoption of GSPs on firm performance would also be determined. The accumulated knowledge from this research would be contributing to literature; furthermore, it would be useful for the development of systematic framework for the adoption of GSPs. At the moment, preliminary finding show that very few large shipping firms are implementing GSPs hence it is still largely an unfamiliar terrain. Developing a systematic frame would help with the diffusion of the concept in the industry. Lastly, understanding the driving force of GSPs will be beneficial for the development of business friendly environmental regulations which will also encourage firm to partake in such practices.

This research proposes adopts an exploratory/explanatory case study approach to research what factors drive the adoption of GSPs in the shipping industry. Understanding these factors is vital to the success of implementing GSPs. The complexity of the shipping is envisaged to alter the replication of any existing adoption systems in other sectors. It is also desirable to know if the finding from thus research will be similar or different from those from other environment management practices. Justification for the adopted methodology has been provided in the methodology section, which largely includes the fact that the area being research is relatively new and there are not many literature and data amongst other reasons.

Institutional theory having been very useful in helping to understanding adoption of environmental practices in other sectors is considered to help with providing a generic understanding of the subject area. This is not to say however that this theory is fully adopted for this research however it is used to first to explore the subject area. While Institutional theory provides a holistic view of the pressures that can influence organizational practices/behaviors, It is the interest of this research to find out if these pressures are responsible for the adoption of GSPs and if not what exactly drives the adoption of GSPs and how adoption of GSPs affects firm performance. Further more, it will be interesting to see if findings from this research will be similar to those from other environmental related research and/or if the complexity of the shipping industry will play any role in altering the findings obtained.

At the moment there are no results presented in this paper, this is because the research is only just approaching the data collection stage. There has previously been what can be considered as a preliminary data collection, which gives an incline that adoption of GSPs, may largely motivated by financial reasons. How and why this is so is still unknown and results cannot be generalized hence the expected results from this research cannot be emphatically stated at the moment.
References
Celik M. A hybrid design methodology for structuring an integrated environmental management system (IEMS) for shipping business. Journal of Environmental Management 2009; 90(3): 1469–75


Wong CWY, Lai KH, Teo TSH. Institutional pressures and mindful information technology management: the case of a container terminal in China. Information & Management 2009c; 46(8): 434–41


Healthcare Information Systems

Chairs: Associate Prof. Andriana Prentza and Associate Prof. Flora Malamateniou
Table of Contents

**Track – Healthcare Information Systems**

4. [PID-115] Leandros Stefanopoulos, “A SMARTPHONE APP FOR EFFICIENT NOTE KEEPING IN CONCURRENT MULTI-PARTICIPANT BEHAVIOURAL STUDIES”
OPERATIONAL AND INNOVATION COLLABORATION
AND CLOUD COMPUTING

Euripidis Loukis, Department of Information and Communication Systems Engineering, University of the Aegean, Greece, eloukis@aegean.gr

Niki Kyriakou, Department of Information and Communication Systems Engineering, University of the Aegean, Greece, nkyr@aegean.gr

Konstantinos Pazalos, Department of Information and Communication Systems Engineering, University of the Aegean, Greece, kpaz@aegean.gr

Abstract

A major trend in the modern economy is the increasing collaboration among firms having complementary resources, both at the operational and at the product/service and process innovation level. At the same time another major trend in the area of information and communication technologies (ICT) is the emergence of cloud computing (CC), which changes radically the way firms access and use ICT for supporting their activities. It has been argued that there is an association between these two trends: that CC can significantly support and facilitate business collaboration at a low cost. However, there has been no empirical investigation of this association based on large datasets. This paper contributes to filling this research gap, by presenting an empirical investigation of the effects of firm’s operational and innovation collaboration with other firms, and also the use of ICT for supporting it, on firm’s propensity to adopt CC. It is based on a dataset collected in the e-Business Survey of the European Commission from 676 European firms from the glass, ceramics and cement industries. It has been concluded that innovation collaboration with other firms, and also the use of ICT for supporting it, have positive effects on the propensity to adopt CC; positive effects has also, mainly in the smaller firms, the use of ICT for supporting some forms of operational collaboration with other firms.

Keywords: Cloud Computing, Collaboration, Network, Innovation, Operations.

1 INTRODUCTION

A major trend in the modern economy is the increasing collaboration among firms having complementary resources, both at the operational and the product/service and process innovation level (Rycroft, 2007; Zeng et al., 2010; Xie et al., 2013; Majava et al., 2013). The globalization, the strong competition, the continuous emergence of new technologies, the fast changes that characterise the modern business environment, as well as the high expectations and demands of consumers for high value-added products and services, and also for continuous renewal and improvement of them, make it difficult for individual firms to survive on their own, relying only on their internal resources. This drives firms to enter in various types of collaboration with other firms, in order to gain access to additional resources (e.g. equipment and production facilities, human skills, knowledge). The simplest form of this collaboration is buying products and services from other firms, however there have been developed more complex forms of collaboration as well, such as business networks, clusters, ecosystems, innovation hubs, keiretsu, and triplehelix (Majava et al., 2013), in which participate suppliers, customers, partners, and even universities and government agencies, and collaborate in order to design, produce and promote innovative products and services. These multiple forms of collaboration among firms necessitate extensive exchange of information, both ‘structured’ and ‘unstructured’, among collaborating firms, which can be greatly supported through the use of information and communication technologies (ICT) (Baraldi and Nadin, 2006; Loukis and Charalabidis, 2013).
At the same time a major trend in the area of information and communication technologies (ICT) is the emergence of cloud computing (CC), which changes radically the way firms access and use ICT for supporting their activities (Armbrust et al., 2010; Zhang et al., 2010; Marston et al., 2011; Venders and Whitley, 2012). Marston et al (2011) define CC as ‘an information technology service model where computing services (both hardware and software) are delivered on-demand to customers over a network in a self-service fashion, independent of device and location’. There are three main categories of CC services currently offered: infrastructure as a service (IaaS) (= remote use of provider’s storage and computing facilities), platform as a service (PaaS) (= remote use of provider’s platform, including also operating system support and software development environment, for the development and deployment of applications) and software as a service (SaaS) (= remote use of software applications running on provider’s systems and supported/maintained by them). CC can provide significant benefits to firms, associated mainly with the reduction of their ICT ownership and operation costs, conversion of related capital investments to operating costs, and also quick and low cost development of the ICT support required for product/service and process innovations. However, CC can pose some risks as well, associated mainly with data security (concerning lack of control of, unauthorized access to or modification of firm’s data resources) and performance (e.g. availability). For these reasons the adoption of CC by firms has been lower than the initial expectations (Saya et al., 2010; Benlian and Hess, 2011; Hsu et al., 2014; Oliveira et al., 2014); this has motivated considerable research on the factors affecting positively or negatively CC adoption, which is briefly reviewed in the next section.

It has been argued that there is an association between these two major trends: a recent study conducted by the London School of Economics (LSE) based on interviews with practitioners revealed that CC has a strong potential to support and facilitate business collaboration at a low cost (Willcocks et al., 2014). In particular, the existing business computing paradigm is based on firms’ internal information systems (IS), which are protected from the external world through highly secure firewalls, acting as clear and strict borders and ‘gates’ of the firm, inhibiting electronic collaboration with the external world (requiring complex, costly and inflexible security changes in firewall settings, security upgrades with external communication channels, etc.). On the contrary the new business computing paradigm emerging through CC moves firm’s ICT support infrastructure outside its firewall, and enables much easier authorized access to appropriate parts of it (e.g. to some of the data, or some of the functionality) by external entities (e.g. customers, suppliers, business partners, etc.), anytime and from anywhere, as it happens with its employees, without having to make complex security changes in the settings of firm’s firewall; this enables and facilitates a much richer collaboration with the external world, easily and at a low cost. The above study argues that this will gradually lead blur the boundaries of firms, and in general lead to structural changes of them, giving rise to the ‘cloud corporation’, which has much more ‘amorphous’ and less stict boundaries with the external world, is much more collaborative with external stakeholders, flexible and ‘fluid’. Furthermore, in recent years have been developed a variety of cloud-based collaboration tools (Forbes, 2013; Tan and Kim, 2015; Ross and Blumenstein, 2015), offered through a SaaS model, which enable firms to access and use a rich collaboration support functionalities (e.g. centralized content storage, forums, instant messaging and other interaction and productivity applications, social media type applications, project management, etc.), available to both firm’s employees and also external entities anytime and from anywhere, rapidly (requiring only minimal initial settings and customizations) and at a low cost, with high levels of availability and flexibility.

However, there has been no empirical investigation of the association between business collaboration and CC based on large datasets. This paper contributes to filling this research gap, by presenting an empirical investigation of the effects of firm’s operational and innovation collaboration with other firms, and also the use of ICT for supporting them, on firm’s propensity to adopt CC. Our research objective is to study CC adoption from a business collaboration perspective, and address the following research question: ‘Is operational and innovation collaboration with other firms, and also the use of ICT for this purpose, a driver of CC adoption ?’ or equivalently ‘Do the firms view the CC as a cost-effective means of supporting operational and innovation collaboration with other firms, and as a means of reducing the cost and increasing the capabilities and flexibility of existing operational and innovation oriented electronic collaboration they have with other firms?’. Our study is based on a
dataset collected in the e-Business Survey of the European Commission from 676 European firms from the glass, ceramics and cement industries.

This paper consists of six sections. The introduction is followed by section 2 outlining the background of this study. Then the research hypotheses are formulated in section 3, and the data and method are described in section 4. The results are presented and discussed in section 5, and finally section 6 summarizes the conclusions.

2 BACKGROUND

2.1. Business collaboration – networking

As mentioned in the introduction, the globalization and the strong competition that characterise the modern business environment, as well as the continuous emergence of new technologies and the fast changes, and also the high expectations and demands of consumers, lead firms to increasingly collaborate with other firms, in order to gain access to wider range of resources (both hard ones, such as equipment and production facilities, and soft ones, such as human skills and knowledge) beyond their own (Rycroft, 2007; Zeng et al., 2010; Xie et al., 2013; Majava et al., 2013). This can initially take simpler forms, such as buying products and services from other firms. Today firms tend to expand their activities to wider geographical areas, buying, and also selling as well, products and services beyond their regional and country markets, in wider international markets. Furthermore, there is an increase of firms’ outsourcing of some parts’ production or services provision from other specialised firms all over the world, in order to take advantage of their resources and economies of scale, and focus on their core activities (Gusmano et al., 2009; Navghavi and Ottaviano, 2010; Arvanitis and Loukis, 2013). Then they proceed to more complex forms of collaboration, such as development of business networks, clusters, ecosystems, innovation hubs, keiretsu, and triplehelix structures (Majava et al., 2013), in which participate a variety of stakeholders, such as suppliers, customers, partners, and even universities and government agencies, and collaborate in order to design, produce and promote innovative products and services. Firms are increasingly creating various types of business collaboration structures, which comprise different and heterogeneous organizations, having various types of relationships among them and also economic and social exchanges, which aim at the design, production, marketing and distribution of complex products and services. These collaboration structures have become of critical importance in the modern economy (Rycroft, 2007; Busquets, 2010; Zeng et al., 2010), so competition in many industries tends to be more among such collaboration structures than among individual firms.

The participation of a firm in such collaboration structures offers significant business benefits (Baraldi & Nadin, 2006; Kajikawa et al., 2010; Zeng et al., 2010): access to complementary resources and capabilities, new markets and technologies, diverse knowledge, and also opportunities to achieve economies of scale, to share the costs and risks of firm’s activities, and to cope with market and technological complexities. Also, they facilitate learning through transfer of knowledge among participating firms, acting as ‘conduits’ for moving and processing knowledge, so they increasingly become the ‘locus’ of combination of diverse knowledge and complementary resources, and collaborative creation of novel knowledge and innovation. Previous research in the innovation domain has shed light on the increasing importance of these business collaboration structures for innovation activity in the last decade (Mancinelli and Mazzanti, 2009; Zeng et al., 2010; Huizingh, 2011; Salavisa et al., 2012). It has revealed that there has been a fundamental change in the way firms design and implement innovation; while previously this has been viewed as a predominantly internal task, in the last decade it increasingly becomes a more ‘open’ and collaborative process based on interactions among different firms.

The relationships among firms as part of such structures necessitate specific coordination actions at three layers (Hakansson & Snehota, 1995; Baraldi & Nadin, 2006): ‘activity links’ (i.e. mutual adaptations in their activities), ‘resource ties’ (i.e., technical connections and mutual orientations of their physical and organisational resources) and ‘actor bonds’ (i.e. social interactions between individuals and organisational units of collaborating firms). These coordination actions require
extensive exchanges of information, both ‘structured’ and ‘unstructured’, which can be greatly supported through the use of appropriate ICT, and especially through the use of appropriate CC services, taking into account the strong potential of the latter to support and facilitate business collaboration at a low cost, as mentioned previously in the Introduction. This has lead to a big growth of the business collaboration software market, which is expected to boom from $0.85 billion in 2011 to $ 4.5 billion in 2016 (Redwood Capital, 2013). Numerous platforms have been recently developed in order to support such collaboration, by providing centralized content storage, forums, instant messaging and other interaction and productivity capabilities, and also project management and social media type capabilities; most of them are offered through a SaaS model (Forbes, 2013; Tan and Kim, 2015; Ross and Blumenstein, 2015). An examination of the recent Forrester report in business collaboration confirms these trends (Koplowitz, 2014). However, the association between business collaboration and CC has not been empirically investigated using large datasets. Our study makes some first contribution towards filling this research gap.

2.2. Cloud Computing Adoption

As mentioned in the Introduction, the adoption of CC by firms has been lower than the initial expectations, and this has motivated considerable research on the factors affecting positively or negatively CC adoption. This research can be divided into three main streams based on the underlying theoretical foundation.

The first stream of empirical CC adoption studies were based on various adaptations of the Technology Acceptance Model (TAM). Wu (2011a) developed an explorative model of SaaS adoption factors, which includes classical factors from TAM (perceived usefulness, perceived benefits, perceived ease of use, attitude, behavioral intention of future use) and its extensions (social influence, marketing efforts), and also CC specific factors (security and trust). Using data collected from 42 Taiwanese managers a structural equation model was estimated connecting the above factors. It lead to the conclusion that the main factors affecting intention to use CC in the future are perceived ease of use, followed by perceived usefulness, which are both affected by social influences (such as mass media, expert opinions and word-of-mouth) and marketing. Wu (2011b) applied data mining techniques (rough set theory) in order to extract relations among the above TAM-based factors of the previous study, based on data collected from 246 Taiwanese managers. It has been concluded that expert opinions are very influential for CC adoption, which is also affected significantly by the perceptions concerning CC effectiveness. Gupta (2013) extend the TAM and develop a five factors model of the inclination of small and medium firms to use CC, which includes perceived ease of use, cost savings, support of collaboration and data sharing, security and privacy reliability, and reliability as independent variables. Using data from 211 small and medium firms a structural equation model has been estimated, which lead to the conclusion that the perceived ease of use has the strongest effect, followed by security and privacy, and cost reduction; on the contrary the perceived reliability and support of collaboration and data sharing did not have statistically significant effects.

A second stream of empirical CC adoption research has been based on a wider theoretical foundation: the Technology, Organization and Environment (TOE) theory (Baker, 2011). This theory identifies three groups of factors that affect the adoption of technological innovations by firms: technological (= perceived characteristics of the technological innovation), organizational (= firm’s characteristics) and environmental (= characteristics of firm’s external environment) ones. Using TOE theory as their theoretical foundation Low et al. (2011), and based on data from a sample of 111 Taiwanese high-tech industry firms, examine the effect of a set of technological factors (relative advantage, complexity and compatibility), organizational factors (top management support, firm size and technology readiness) and environmental factors (competitive pressure and trading partner pressure) on CC adoption. They found that perceived relative advantage, top management support, firm size, competitive pressure and trading partner pressure have statistically significant effects on CC adoption. Another TOE-based study has been conducted by Hsu et al. (2014), which examines the effect of perceived benefits and business concerns (technological factors), IT capability (IT personnel and budget - organizational factor) and external pressure (environmental factor) on CC adoption intention, using data from 200
Taiwanese firms. It concluded that the first three of these factors are significant determinants of CC adoption while the fourth is not. Mangula et al. (2014), using data from 147 Indonesian firms, examine the effect of a set of technological factors (relative advantage, compatibility, complexity, trialability, observability), organizational factors (organizational readiness, top management support) and environmental context (market pressure, market competition vendor marketing, trust in vendor, government support) on the adoption of Software as a Service (SaaS) services. They found that compatibility, observability, market competition and government support have a positive correlation with SaaS adoption, while complexity has a negative correlation with it. Oliveira et al. (2014) using data from 369 Portuguese firms examine the effects of three CC innovation characteristics (relative advantage, complexity and compatibility), two organizational context characteristics (top management support, firm size), one technological context characteristic (technological readiness) and two environmental context characteristics (competitive pressure, regulatory support). They conclude that relative advantage, technological readiness, top management support and firm size have positive effects on CC adoption, while complexity has a negative effect.

A third stream of CC adoption empirical studies is based on the synthesis of other theoretical frameworks. Benlian et al. (2009), developed a SaaS adoption model by combining and adapting for SaaS three theoretical perspectives: transaction cost theory (including in their model the application specificity and perceived uncertainty), resource-based view of the firm (including application strategic value and inimitability) and theory of planned behaviour (including the attitude towards SaaS and also social influence). Using data collected from 297 German firms they estimated structural equation SaaS adoption models, from which it has been concluded that social influence, adoption uncertainty and application strategic value are the most consistent SaaS adoption drivers across all application types. Saya et al. (2010), based on the institutional theory and the real options theory, and using data collected from 101 ICT professionals from Singapore and Japan, formulated and estimated a four layers structural equation CC adoption model. They conclude that institutional influences (e.g. from government, customers, suppliers, competitors, strategic partners, industry and trade organizations, professional bodies) affect organizations perceptions about the technological characteristics of CC (perceived accessibility, scalability, cost effectiveness and lack of security), and through them affect the perceptions on the provided real options by CC adoption (concerning ICT applications growth, abandonment and deferral) and finally the intention to adopt CC. Benlian and Hess (2009), having as theoretical foundation the theory of reasoned action in combination with previous research on ICT outsourcing and application service provision (ASP), using data collected through a survey of 349 IT executives at German companies, examine the effects of perceived SaaS opportunities (cost advantages, strategic flexibility, focus on core competencies, access to specialised resources and quality improvements) and SaaS risks (performance, economic, strategic, security and managerial ones) on the intention to increase the level of its adoption. They conclude that the perceived cost advantages have the strongest positive effect, followed by strategic flexibility and the quality improvement; the focus on core competencies and the access to specialized resources do not have statistically significant effects. Furthermore, the security risks have the strongest negative effect, followed by the performance, economic and strategic risks; the managerial risks do not have statistically significant effects. Wu et al. (2013) study empirically the effects of two information processing requirements related factors (business process complexity and entrepreneurial culture) and two information processing capacity related factors (applications, functionality and compatibility) on the intention to adopt CC, using data from 289 USA manufacturing and retail firms. The theoretical foundations of this study are the innovation diffusion theory (DOI) (focusing mainly on its relative advantage and compatibility dimensions) and the ‘information processing view’ (IPV) of the firm. They conclude that business process complexity and also applications compatibility have negative effects on CC adoption intention while, on the contrary, entrepreneurial culture and applications functionality have positive effects.

From the above review of previous empirical literature on CC adoption it is concluded that most of the examined factors as to their impact on CC adoption are firms’ perceptions concerning the benefits and capabilities of CC, as well as its risks. However, limited empirical research has been conducted on the effect of firm’s characteristics on CC adoption, despite their influence of the former on the benefits and risks of CC perceived by the firm. The effects of only a small number of firm characteristics
(mainly firm size, top management support, technological readiness and organizational readiness) on CC adoption has been empirically investigated. This study contributes to filling this research gap, by empirically investigating the effects of firm’s operational and innovation collaboration with other firms, and also the use of ICT for supporting various forms of them, on firm’s propensity to adopt CC.

3 Research Hypotheses

Our first research hypothesis concerns the effect of operational collaboration with other firms on the propensity to adopt CC. As mentioned in the introduction and in 2.1, in the modern economy firms tend to increase their operational collaboration with other firms. On one hand they outsource part of their previously internal activities, in order to focus on their core activities, and take advantage of resources and economies of scale of other firms (Gusmano et al., 2009; Navghavi and Ottaviano, 2010; Arvanitis and Loukis, 2013); this leads to an increase of their external procurement. On the other hand, they are expanding their selling activities in wider geographic areas, in order to increase their sales revenues, economies of scale, and finally profitability. Firms tend to increase the geographing scope of their sales and procurement, moving from regional markets to country, and then to international markets. These increase significantly firms’ operational complexity and workload, especially in the cases of international expansion of sales (since selling to customers beyond firm’s country necessitates the management of many different legislations, regulations, taxations, specific needs, etc., and also the organization of complex shipments to many geographic locations), and international procurement (since having suppliers beyond firm’s country poses similar challenges). The above increase the requirements for storage, processing and exchange of relevant information. This leads to high costs for the development, maintenance and operation of supporting IS, and also for interconnecting them with the ones of main customers and suppliers, which can increase firm’s operating costs and reduce its competitiveness. CC can be quite valuable in such cases, as it can drastically reduce the above costs, and also transform them to ‘operating expenses’. Also, as mentioned in the introduction, CC moves firm’s ICT support infrastructure outside its firewall, and enables much easier authorized access to appropriate parts of it (e.g. to some of the data, or some of the functionality) by external entities (e.g. customers, suppliers, business partners, etc.), anytime and from anywhere; this can provide an effective and at the same time easy, rapid to implement and low cost support of operational collaboration with these external entities. For the above reasons we expect that firms having extensive operational collaboration with other firms will have a high motivation and propensity to adopt CC. So our first research hypothesis is:

H1: The extent of firm’s operational collaboration with other firms is positively associated with its propensity for cloud computing adoption.

As mentioned above operational collaboration with other firms necessitates extensive storage, processing and exchange of relevant information, which needs various forms of ICT support (e.g. electronic exchange of orders, invoices, inventory levels and other data) in order to be efficient. This ICT support has high operation, support, maintenance and upgrade costs, so it can be highly beneficial to use CC services in order to reduce these costs (e.g. by using IaaS and PaaS services for hosting some of these applications, or even by using SaaS for replacing some older and/or bespoke applications with more modern standard software packages). Also, the electronic exchange of orders, invoices, inventory levels and other data can be conducted much easier and at a lower cost if the firms we are collaborating with are given access to appropriate parts of our cloud based/hosted ICT infrastructure (data and functionality). This can provide an efficient support of operational collaboration with other firms, which has also high flexibility for addressing changes in our business collaboration networks (new firms can be easily given such access if required, and this will activate immediately electronic collaboration with them). For the above reasons we expect that firms using ICT for supporting various forms of operational collaboration with other firms will have a high motivation and propensity to adopt CC. So our second research hypothesis is:

H2: The use of ICT for supporting firm’s operational collaboration with other firms is positively associated with its propensity for cloud computing adoption.
Our third research hypothesis concerns the effect of innovation collaboration with other firms on the propensity to adopt CC. As mentioned in the introduction and in 2.1, in the modern economy innovation becomes increasingly collaborative: firms are increasingly collaborating with other firms, which possess complementary resources (e.g. equipment and production facilities, human skills, knowledge), in order to design, produce and promote innovative products, services, or even production and business processes (Rycroft, 2007; Salavisa et al., 2012; Zeng et al., 2010). This necessitates extensive exchange of information (both structured and unstructured) in all the phases of innovation life-cycle: generation of innovation ideas, selection of the most promising ones, innovation implementation, evaluation, and marketing. This can be significantly supported and facilitated through the use of appropriate ICT. The use of CC services enables the development, operation and maintenance of this ICT support of innovation rapidly, at a low cost, without having to make additional investments. As mentioned in the introduction, in the recent years have been developed a variety of cloud-based collaboration tools (Forbes, 2013; Tan and Kim, 2015; Ross and Blumenstein, 2015), offered through a SaaS model, which provide a wide range of collaboration support functionalities. For the above reasons we expect that firms having innovation collaboration with other firms will have a high motivation and propensity to adopt CC. So our third research hypothesis is:

**H3:** Innovation collaboration with other firms is positively associated with propensity for cloud computing adoption.

Firms already using ICT for the electronic support of their innovation collaboration with other firms can substantially reduce the operation, support, maintenance and upgrade cost of it, and also gain access to better and more extensive collaboration support functionality, by using appropriate CC services (e.g. by replacing existing on-premises collaboration support IS with the use of cloud-based collaboration tools offered through a SaaS model). For the above reasons we expect that firms using ICT for supporting innovation collaboration with other firms will have a high motivation and propensity to adopt CC. So our fourth research hypothesis is:

**H4:** The use of ICT for supporting firm’s innovation collaboration with other firms is positively associated with its propensity for cloud computing adoption.

### 4 Data And Method

The data used in this study has been collected through the “e-Business Survey 2009”, which has been conducted by the e-Business Market W@tch (www.ebusiness-watch.org) initiative of the European Commission, from a sample of 676 firms, from the glass, ceramic and cement sectors of six European countries (Germany, France, Italy, Poland, Spain, UK); 53.8% of the sample firms were small (with 1-49 employees), 33.6% were medium (with 50-249 employees) and the remaining 12.6% were large firms (with more than 250 employees).

The questions of the aforementioned survey used in this study are shown in the Appendix. As dependent variable has been used the propensity for CC adoption, which is measured in a three levels scale: very relevant, partly relevant or not relevant. We have used four groups of independent variables, which concern operational collaboration, electronic operational collaboration, innovation collaboration and electronic innovation collaboration respectively. The first group includes two independent variables concerning the extent of firm’s operational collaboration with other firms: the geographic scope of firm’s sales and procurement; they are both measured in a three levels scale: regional, country or international. The second group includes binary (yes/no) variables assessing whether or not the firm is using four types of IS for supporting four particular forms of operational collaboration with other firms: supply chain management (SCM) systems, electronic sharing of information on inventory levels with suppliers, electronic orders from customers, and electronic invoicing. The third group includes two binary (yes/no) independent variables assessing whether or not in the development of firm’s product/service innovations and process innovations respectively are involved other firms or external experts. The fourth group includes one binary (yes/no) independent variable assessing whether or not the firm is using software applications to collaborate with other firms in the development of product/service innovations or process innovations.
In order to test the research hypotheses H1 – H4 we calculated the association between the dependent variable and each of the abovementioned independent variables, using two measures of ordinal variables’ association: the Somers’ d and the Kendall’s tau-b (both ranges from -1 to 1, with their absolute values indicating the strength of association (larger absolute values means stronger association), and their sign indicating the direction of the association. It should be noted that we did not estimate a regression model because there were high correlations between our independent variables (according to the econometric literature (e.g. Greene, 2011; Gujarati, 2008) if we have high levels of correlation between the independent variables of a regression (multi-collinearity problem), then the regression coefficients are not reliable measures of the impact of the independent variables on the dependent variable).

The above association measures were calculated initially for the entire sample, and then for two sub-samples of it: the first one included the small firms (with 1-49 employees – 53.8% of the sample), and the second one the medium and large firms (with 50 or more employees – 46.2% of the sample), in order to examine to whether the effects of the abovementioned collaboration related variables on the propensity to adopt CC depend on firm size. It should be noted that we did not create separate sub-samples for the medium and for the large firms, because they would be much smaller (especially the large firms’ one) than the small firms’ sub-sample, making comparison difficult, as smaller sample size increases the confidence intervals of the estimated association measures, and therefore affects their significances).

5 Results

We can see the calculated Sommer’s D and Kendall tau-b values between the propensity to adopt CC and each of the abovementioned independent variables, initially for the whole dataset in Table 1, and then for the small firms’ sub-sample in Table 2 and for the medium and large sub-sample in Table 3. The statistically significant values (having significance lower than 10%) are shown in bold.

With respect to the operational collaboration we remark that for both variables there is not statistically significant association with the propensity to adopt CC (neither in the whole sample, nor in the two sub-samples). Therefore hypothesis 1 is not supported. This indicates that having a wide geographic scope of sales and procurement, which necessitates extensive operational collaboration with a big number and variety of firms, and therefore extensive storage, processing and exchange of relevant information, is not a driver of CC adoption; firms of the three examined sectors do not view CC as a cost effective means of providing or increasing ICT support of their operational collaboration with customers and suppliers. A possible reason might be in the three examined sectors these operational collaboration processes exhibit significant specificities and complexities, leading to high levels of ‘asset specificity’ (e.g. need of highly specialised and customised software applications in the CC services provider side, and also extensive communication and cooperation between highly knowledgeable personnel of the CC services provider and the CC services user) and ‘uncertainty’ (as to whether the CC services provider can meet all the special needs with satisfactory service levels and price). This higher asset specificity and uncertainty, according to the transaction cost theory (Williamson, 1985; Benlian and Hess, 2009; Arvanitis and Loukis, 2013), make CC outsourcing of electronic collaboration with customers and suppliers more difficult and costly to manage, and less attractive and beneficial, in comparison with the on-premises alternative. Another possible reason might also be that the adoption of CC for supporting critical everyday activities (such as the operational collaboration with ) is risky, and requires a certain level of ‘cloud computing maturity’ along several technological and organizational dimensions (Oracle, 2011); there is a chance that the three examined sectors do not possess sufficient maturity for this.

On the contrary, there is statistically significant positive association of the use of SCM systems with propensity adoption in the whole sample, and in the two size sub-samples; the same holds for the e-invoicing, but only in the whole sample and in the small firms’ sub-sample, and also for the electronic orders from customers, but only in the small firms’ sub-sample. However, there is not statistically significant association with the propensity to adopt CC of the electronic sharing of information on inventory levels with suppliers (neither in the whole sample, nor in the two sub-samples). Therefore
hypothesis 2 is only partially supported (only for some types of ICT use for supporting particular forms of electronic collaboration, and mainly for the small firms).

These results indicate that though the firms of the three examined sectors do not view CC as a cost effective means of providing ICT support of their operational collaboration with customers and suppliers, they view it however as a means of reducing the cost and increasing the flexibility of existing ICT support of some particular forms of operational collaboration with other firms, especially highly sophisticated ones, such as the SCM. A possible explanation for this might be that already using ICT for supporting particular forms of electronic collaboration leads to the generation and accumulation of relevant experience, knowledge and maturity within the firms, which makes it easier to define their specialised needs, and communicate them to CC services providers, to evaluate their offerings, and makes it easier to manage their relationships with CC service providers, while it also reduces uncertainty; these reduce the ‘transaction costs’ with the CC services provider, making this CC outsourcing more attractive and beneficial.

Also, the above results indicate that the small firms view CC as a means of reducing the cost and increasing the flexibility of existing electronic commerce related capabilities, concerning electronic customer ordering and electronic invoicing. A possible reason for this might be that the costs of operating and maintaining on-premises such ICT infrastructures are high for the small firms, as they cannot have the economies of scale that a CC services provider can have; also the small firms do not have sufficient specialised ICT personnel for the required continuous improvement and evolution of these applications, in order to exploit new technologies and satisfy new needs. However, these do not hold for the larger firms (as they can have high levels of economies of scale, and usually have sufficient specialised ICT personnel).

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Sommer’s D</th>
<th>Kendall tau-b</th>
<th>Research Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Operational Collaboration</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geographic scope of sales</td>
<td>0.011</td>
<td>0.013</td>
<td>H1</td>
</tr>
<tr>
<td>Geographic scope of procurement</td>
<td>0.016</td>
<td>0.018</td>
<td>H1</td>
</tr>
<tr>
<td><strong>Electronic Operational Collaboration</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of SCM systems</td>
<td>0.190</td>
<td>0.190</td>
<td>H2</td>
</tr>
<tr>
<td>Electronic sharing of information on inventory levels with suppliers</td>
<td>0.052</td>
<td>0.052</td>
<td>H2</td>
</tr>
<tr>
<td>Electronic orders from customers</td>
<td>0.053</td>
<td>0.054</td>
<td>H2</td>
</tr>
<tr>
<td>Electronic invoicing</td>
<td>0.124</td>
<td>0.131</td>
<td>H2</td>
</tr>
<tr>
<td><strong>Innovation Collaboration</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Involvement of other firms in product/service innovations</td>
<td>0.130</td>
<td>0.131</td>
<td>H3</td>
</tr>
<tr>
<td>Involvement of other firms in process innovations</td>
<td>0.162</td>
<td>0.166</td>
<td>H3</td>
</tr>
<tr>
<td><strong>Electronic Innovation Collaboration</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of software applications to collaborate with other firms for product/service or process innovations</td>
<td>0.153</td>
<td>0.153</td>
<td>H4</td>
</tr>
</tbody>
</table>

Table 1. Sommer’s D and Kendall tau-b values between propensity to adopt CC and independent variables (for the whole sample)
<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Sommer’s D</th>
<th>Kendall tau-b</th>
<th>Research Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Operational Collaboration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geographic Scope of Sales</td>
<td>-0.069</td>
<td>-0.084</td>
<td>H1</td>
</tr>
<tr>
<td>Geographic Scope of Procurement</td>
<td>-0.003</td>
<td>-0.004</td>
<td>H1</td>
</tr>
<tr>
<td></td>
<td>Electronic Operational Collaboration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of SCM systems</td>
<td><strong>0.110</strong></td>
<td><strong>0.110</strong></td>
<td>H2</td>
</tr>
<tr>
<td>Electronic sharing of information on inventory levels with suppliers</td>
<td>0.090</td>
<td>0.091</td>
<td>H2</td>
</tr>
<tr>
<td>Electronic orders from customers</td>
<td><strong>0.113</strong></td>
<td><strong>0.117</strong></td>
<td>H2</td>
</tr>
<tr>
<td>Electronic invoicing</td>
<td><strong>0.190</strong></td>
<td><strong>0.208</strong></td>
<td>H2</td>
</tr>
<tr>
<td></td>
<td>Innovation Collaboration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Involvement of other firms in product/service innovations</td>
<td><strong>0.117</strong></td>
<td><strong>0.118</strong></td>
<td>H3</td>
</tr>
<tr>
<td>Involvement of other firms in process innovations</td>
<td><strong>0.216</strong></td>
<td><strong>0.220</strong></td>
<td>H3</td>
</tr>
<tr>
<td></td>
<td>Electronic Innovation Collaboration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of software applications to collaborate with other firms for product/service or process innovations</td>
<td><strong>0.119</strong></td>
<td><strong>0.120</strong></td>
<td>H4</td>
</tr>
</tbody>
</table>

Table 2. Sommer’s D and Kendall tau-b values between propensity to adopt CC and independent variables (for the small firms’ sub-sample)

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Sommer’s D</th>
<th>Kendall tau-b</th>
<th>Research Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Operational Collaboration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geographic Scope of Sales</td>
<td>0.041</td>
<td>0.044</td>
<td>H1</td>
</tr>
<tr>
<td>Geographic Scope of Procurement</td>
<td>0.014</td>
<td>0.015</td>
<td>H1</td>
</tr>
<tr>
<td></td>
<td>Electronic Operational Collaboration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of SCM systems</td>
<td><strong>0.216</strong></td>
<td><strong>0.217</strong></td>
<td>H2</td>
</tr>
<tr>
<td>Electronic sharing of information on inventory levels with suppliers</td>
<td>0.009</td>
<td>0.009</td>
<td>H2</td>
</tr>
<tr>
<td>Electronic orders from customers</td>
<td>-0.008</td>
<td>-0.007</td>
<td>H2</td>
</tr>
<tr>
<td>Electronic invoicing</td>
<td>0.043</td>
<td>0.045</td>
<td>H2</td>
</tr>
<tr>
<td></td>
<td>Innovation Collaboration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Involvement of other firms in product/service innovations</td>
<td><strong>0.112</strong></td>
<td><strong>0.113</strong></td>
<td>H3</td>
</tr>
<tr>
<td>Involvement of other firms in process innovations</td>
<td><strong>0.093</strong></td>
<td><strong>0.095</strong></td>
<td>H3</td>
</tr>
<tr>
<td></td>
<td>Electronic Innovation Collaboration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of software applications to collaborate with other firms for product/service or process innovations</td>
<td><strong>0.159</strong></td>
<td><strong>0.160</strong></td>
<td>H4</td>
</tr>
</tbody>
</table>

Table 3. Sommer’s D and Kendall tau-b values between propensity to adopt CC and independent variables (for the medium and large firms’ sub-sample)

With respect to the innovation collaboration, from the above Tables 1, 2 and 3 we can see that it is much stronger associated with propensity to adopt CC than the operational collaboration. In particular, there is statistically significant positive association of innovation collaboration, both for
product/service innovation and for process innovation, and also of the use of ICT for supporting it, with propensity for CC adoption, in the whole sample, and in the two size sub-samples. Therefore hypotheses 3 and 4 are both supported. These results indicate that firms of the three examined sectors view CC as a cost-effective means of providing electronic support of innovation oriented collaboration with other firms and external experts, and also as a means of reducing the cost and increasing the capabilities and flexibility of existing ICT support of innovation collaboration. A possible explanation for these is that innovation collaboration has a much smaller scale and is less critical for the everyday operation of the firms (though quite important for their future performance, or even survival) in comparison with the operational collaboration with other firms; therefore the business uncertainty generated from the use of CC services is lower for the innovation collaboration support than for the operational collaboration support. This, according to the transaction cost theory (Williamson, 1985; Benlian and Hess, 2009; Arvanitis and Loukis, 2013), leads to higher propensity for CC adoption for the former than for the latter.

6 CONCLUSIONS

Two important and widely debated trends in the modern economy is the increasing collaboration among firms, leading to the development of extensive business networks and the increase of their importance, and also the emergence of cloud computing (CC), as a new more efficient paradigm/model of business computing (i.e. of using ICT for supporting firms’ activities). In the previous sections of this paper has been presented an empirical investigation of the association between these two trends. Four research hypotheses have been formulated concerning the effects of firm’s operational and innovation collaboration with other firms, and also the use of ICT for supporting it, on firm’s propensity to adopt CC. They have been tested using a dataset collected from 676 European firms from the glass, ceramics and cement industries through the e-Business Survey of the European Commission.

Our results provide some first evidence concerning the existence of association between the above important trends: we have found that collaboration, mainly innovation oriented, has positive impact on propensity for CC adoption. In particular, it has been concluded that innovation oriented collaboration with other firms (for the development of product, service and process innovations), and also the use of ICT for supporting it, are drivers of CC adoption; this holds for both smaller and larger firms. On the contrary, the operational collaboration with other firms is not a driver of CC adoption in the three examined sectors; however, the use of ICT for supporting some forms of operational collaboration with other firms is a driver of CC adoption, mainly in the small firms.

Our results indicate that firms of these sectors view CC as a cost-effective means of supporting collaboration with other firms in their innovation development activities, but not in their critical every day operations. This might be due to specificities and complexities of the processes and collaboration practices of the three examined manufacturing sectors, which result in limited supply of corresponding specialised SaaS applications by CC providers. Also, the importance of this operational collaboration with suppliers and customers for the everyday activities of these firms makes them hesitant to use external providers of ICT support of them. However, the firms of these sectors view CC as a means of reducing the cost and increasing the capabilities and flexibility of existing ICT support of innovation oriented collaboration, and also of highly sophisticated forms of operational collaboration, such as SCM; furthermore, the small firms view CC as a means of reducing the cost and increasing the flexibility of their electronic commerce capabilities (e.g. concerning electronic customer ordering and electronic invoicing).

Further empirical and theoretical research is required on the association between business collaboration and CC. It should investigate on one hand different forms of business collaboration (both operational and innovation oriented ones), and on the other hands adoption of different categories of CC services (e.g. IaaS, PaaS, SaaS), in various sectoral and national contexts. Also, future relevant research should use a more detailed measurement of the extent of CC services adoption than the binary one used in this study. Furthermore, it would be useful to conduct research for identifying moderators of the relationship between business collaboration and CC adoption (e.g. various characteristics of the
firm, including firm’s ‘cloud computing maturity’ along various technological and organizational dimensions (Oracle, 2011), and also its external environment, that increase or decrease the strength of this relationship).

References


**Appendix: Variables Definitions – Questions**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent variable</strong></td>
<td></td>
</tr>
<tr>
<td>Propensity for cloud computing adoption</td>
<td>How relevant is cloud computing for your company?</td>
</tr>
<tr>
<td><strong>Independent variables: operational collaboration</strong></td>
<td></td>
</tr>
<tr>
<td>Geographic scope of sales</td>
<td>What is your company’s most significant sales market?</td>
</tr>
<tr>
<td>Geographic scope of procurement</td>
<td>Do you procure primarily from suppliers in your region, in your country of from an international supplier base?</td>
</tr>
<tr>
<td><strong>Independent variables: electronic operational collaboration</strong></td>
<td></td>
</tr>
<tr>
<td>Use of SCM systems</td>
<td>Do you use an SCM system (Supply Chain Management)?</td>
</tr>
<tr>
<td>Electronic sharing of information on inventory levels with suppliers</td>
<td>Does your company share information on inventory levels electronically with suppliers?</td>
</tr>
<tr>
<td>Electronic orders from customers</td>
<td>Can customers order goods or services from your company online on the internet or through other computer networks, not counting manually typed e-mails?</td>
</tr>
<tr>
<td>Electronic invoicing</td>
<td>Does your company use e-invoicing, that is sending or receiving invoices electronically?</td>
</tr>
<tr>
<td><strong>Independent variables: innovation collaboration</strong></td>
<td></td>
</tr>
<tr>
<td>Involvement of other firms in product/ service innovations</td>
<td>Were external experts or business partners involved in developing new products or services?</td>
</tr>
<tr>
<td>Involvement of other firms in process innovations</td>
<td>Were external experts or business partners involved in developing new processes?</td>
</tr>
<tr>
<td><strong>Independent variables: electronic innovation collaboration</strong></td>
<td></td>
</tr>
<tr>
<td>Use of software applications to collaborate with other firms for product/ service or process innovations</td>
<td>Does your company use online software applications other than e-mail to collaborate with business partners in the development of new products, services or processes?</td>
</tr>
</tbody>
</table>
CHALLENGES OF EVOLVING PINCLOUD PHR INTO A PHR-BASED HEALTH ANALYTICS SYSTEM

Michaela Poulmyomenopoulou, Department of Digital Systems, University of Piraeus, Piraeus, Greece, email: mpoul@unipi.gr

Flora Malamateniou, Department of Digital Systems, University of Piraeus, Piraeus, Greece, email: flora@unipi.gr

Andriana Prentza Department of Digital Systems, University of Piraeus, Piraeus, Greece, email: aprentza@unipi.gr

George Vassilacopoulos Department of Digital Systems, University of Piraeus, Piraeus, Greece, email: gvass@unipi.gr

Abstract

State of the art technologies like cloud computing, service-oriented architectures and NoSQL databases enable the creation of new generation Personal Health Record (PHR) systems which are provided as cloud services and enable collecting life-long cross-institutional information from various sources. Based on this concept, the PINCLOUD PHR is a cloud-based service that seeks to integrate patient health and social care information from various sources such as the patient, non-healthcare providers, home care systems (that store medical information transmitted from Internet connected medical devices to the patient) and other healthcare information systems (e.g. primary care systems, electronic medical record - EMR systems, e-prescription and e-referral systems). Such PHR services contain integrated big patient data from many sources that can be used in multiple health analytic scenarios in order to provide insights for improving diagnosis and treatment accuracy, cutting down healthcare costs and improving healthcare delivery. Hence, such PHR services, if enhanced with a health analytic engine, can evolve into a PHR analytics system. However, there are many challenges in the big data integration that should be performed before the data analysis. Those involve data trustworthiness, interoperability and security. To this end, in this paper, an overview of the PINCLOUD PHR is presented and the challenges related to the development of the PINCLOUD PHR with the insight to evolve into a health analytics system are also discussed.

Keywords: PHR, health analytics, systems evolution, cloud computing, NoSQL.

1 INTRODUCTION

In healthcare, the continued pressure to comply with regulatory demands, healthcare-specific compliance requirements, and guidelines like the U.S. Health Insurance Portability and Accountability Act (HIPAA) and Health Information Technology for Economic and Clinical Health (HITECH) legislation and the European Union Directives, requires transformation in the healthcare information technology (HIT) to effectively handle exponential data growth, improve operational efficiencies, and manage limited budgets (Behara, 2014). Moreover, big data analytics in healthcare is evolving into a promising field for providing insight from very large data sets and improving outcomes while reducing costs. Hence, health analytics evolution has resulted in a swift in HIT from departmental solutions to encompass larger solutions at the enterprise level, and from standalone systems that provide limited and localized solutions to more interconnected ones that provide comprehensive and integrated solutions (Zarrouk, 2014).

Personal Health Record (PHR) services that integrate patient health and social information from various, heterogeneous data sources is expected to play a key role in the health analytics evolution
since those can provide a data source with integrated patient data to be used by analytic engines that use a set of techniques such as machine learning and text mining in order to analyze patient data. For example, advanced analytics to patient profiles (e.g., segmentation and predictive modeling) can proactively identify individuals who would benefit from preventative care or lifestyle changes or fraud can be predicted and minimized by implementing advanced analytic systems for fraud detection (Raghupathi, 2014, Gearon, 2007).

Based on these concepts, this paper presents a cloud-based PHR service that seeks to integrate different information systems and patient data into a central repository (e.g. data warehouse), leading to the provision to authorized actors of integrated patient data anytime and anywhere. In particular, the PINCLOUD PHR service, stems from our involvement in the PINCLOUD project, and consists of the following data components: a) a non-healthcare component containing health and social information collected by either the patient or non-healthcare providers (e.g. family members friends, social care providers); b) a home care component containing health information existing in home care systems (e.g. medical information transmitted from Internet connected medical devices to the patient); and c) a healthcare professional component containing health information stored in various healthcare information systems (e.g. primary care, electronic medical records – EMRs, e-referral and e-prescribe systems). This PHR service follows the patient-oriented paradigm which means that patients are the owners of their information and are empowered to authorize other subjects to access it.

The PINCLOUD PHR service is built in a reliable and rugged platform warranting stakeholder collaboration and enjoying public trust, by utilizing cloud infrastructure. The PINCLOUD PHR service inevitably contains big data in the sense that data volume can be too large, data variety can comprise of unstructured (e.g. audio, video), semi-structured (e.g. patient generated data or sensor generated data) and structured (e.g. healthcare provider data) forms and data velocity can be too fast while also including incomplete or noisy data due to external factors (Bohlouli, 2013, Kononengo, 2001). The analysis of PINCLOUD PHR big data with the use of appropriate analytic techniques is expected to provide ample opportunities for the improvement of healthcare delivery. Hence, PINCLOUD PHR can be enhanced with a health analytics engine that uses PHR integrated data and a set of analytic algorithms in order to transform patient data into actionable knowledge.

Despite the great potential of health analytics, healthcare analysis and interpretation of data is a highly difficult task. In this paper, the challenges related to the big data integration, like security and interoperability, are discussed. For instance, a factor that hinders big data interoperability is that most healthcare information systems are designed to meet local needs (Raghupathi, 2014, Kumar, 2013). To face these challenges there is a need to provide appropriate solutions at technological, organizational and environmental level. To this end, in this paper, an overview of the PINCLOUD PHR is presented and the challenges related to the PINCLOUD PHR evolution into a PHR-based health analytics system are also discussed. Generally, cloud-based PHRs like the PINCLOUD proposed here provides a solution for achieving big data integration that is a key factor for the realization of health analytics systems in the future.

2 BACKGROUND

Big data analysis typically refers to the analysis of large complex data set that yields substantially more information when analyzed as a fully integrated data set as compared to the outputs achieved with smaller sets of the same data that are not integrated. Healthcare providers have started to realize the value of health analysis and therefore have started to provide their patient data into large data repositories to be used for gaining insights for making better-informed health-related decisions. Currently, there is a series of initiatives that make efforts to build a cloud-based platform to permit secure data sharing among the institutions so as to improve the quality of consumer care. Some examples for electronic, personally-controlled health records delivered as Software as a Service (SaaS) are the Microsoft Corporation Health Vault Program and the Intel Corporation Dossia. Cloud-based PHR services increase the amount of patient data produced and processed resulting in the big data phenomenon (Bohlouli, 2013). The amount of data in cloud-based PHRs like the one presented in this paper is expanding at an extreme pace thus those can help to the health analytics evolution by realizing the value of healthcare analysis of larger integrated datasets (Poulymenopoulou, 2014).

Health analytics systems usually consist of a health analytics engine that analyze a set of patient data by using analytic techniques like machine learning and data mining algorithms. However, since data is by definition large, those systems are mostly implemented on cloud and processing is broken down
and executed across multiple nodes. Cloud computing saves on the costs for storing big datasets while also enables the delivery of increased IT efficiency and service levels. Furthermore, open source platforms such as Hadoop/MapReduce, available on the cloud, have encouraged the application of big data analytics in healthcare (Zarrouk, 2014, Bohlouli, 2013).

In big data analysis of multiple data sets from dispersed data sources there is also a variety of security issues that need to be considered. These include the regulations that protect patient data and prevent patient re-identification by any means, the healthcare providers security policies, the agreements based on consent forms and patient data sharing preferences as imposed by the patient-oriented paradigm. In big data analysis it is widely suggested to aggregate patient data in order to ensure non-identification and anonymization while the original data remain safe from any modifications (Kamateri, 2014). However, even with the use of those methods, the healthcare provider security policies and patient sharing preferences should be considered.

Big data interoperability is also a key problem that hinders big data integration in cloud-based PHRs. Big data might come from many and diverse sources like electronic health records, clinical decision support systems, government sources, laboratories, pharmacies and insurance companies residing at multiple locations (geographic as well as in different healthcare providers’ sites) in numerous legacy and other applications (transaction processing applications, databases, etc.) (Crisholm, 2015). Therefore patient data exist in multiple formats (flat files, .csv, relational tables, ASCII/text, etc.) and with diverse semantics that should be semantically integrated before its use by analytic techniques in order to provide valuable results. Hence, several decisions should be made with regard to the data collection approach, the distributed design, the data integration scheme and the data semantics.

Although there are definite benefits to using cloud computing as a means for storing integrated big patient data, there are also many potential risks, because cloud computing combines new technology with many unproven vendors and service providers. Security, reliability and manageability need to be key elements in the planning and selection processes for the use of private and public cloud services. When sensitive information like patient medical information is to be transferred on a public cloud, the security issues are sometimes a barrier for cloud computing adoption (Poulymenopoulou, 2014, Crisholm, 2015). However, it is possible with the use of appropriate security techniques like encryption, access control policies, data backups, audit logs etc. to safeguard data in the cloud and maximize the healthcare benefits of cloud offerings. Although regulatory and security concerns have held back the healthcare industry from widespread adoption of public clouds, the overall cloud computing market in healthcare is expected to grow to $5.4 billion by 2017 (Zarrouk, 2014).

3 THE PINCLOUD PERSONAL HEALTHCARE RECORD

3.1 Overall architecture

The PINCLOUD PHR service offers to authorized users access to integrated patient data and enables healthcare professionals to perform e-prescriptions and e-referrals as well as to inform and advice patients and the patients to electronically arrange medical appointments, to communicate with healthcare professionals and manage authorizations to patient data.

As show in Figure 1, the PINCLOUD PHR exists at the cloud and communicates through the Internet with other systems like hospitals, medical offices, health insurance companies, diagnostic centers, pharmacies, social care providers, home care systems, e-referral and e-prescribe systems. Moreover, the PINCLOUD PHR integrates patient data from different organizations and systems and provides integrated services to healthcare professionals, thus resulting to improved patient care. Other systems connected to PINCLPUD might also exist at the same or another cloud provider or implemented at local infrastructures and incorporated into the PINCLOUD PHR through the use of web services.

At each source (e.g. social and health care provider, home care systems) connected to the PINCLOUD PHR exist web and/or Representational State Transfer (REST) services that extract a pre-specified subset of patient social and health information (e.g. an extended discharge summary including citizen critical factors like allergies extracted from EMRs) from social care providers and various healthcare systems, accordingly, as JSON documents that are stored to the PINCLOUD PHR data repository. In particular, web/REST services are used in this project as a means of exchanging patient data among the connected systems. This results in lower cost for the connected healthcare organizations since they retain their existing systems and infrastructures without the need for investing on new technologies.
Moreover, the PINCLOUD uses a NoSQL data repository for storing patient data from multiple sources in the form of JSON documents based to a JSON schema defined according to the continuity-of-care (CCD) document schema, extensively used in PHR systems (Bonnet, 2011).

More specifically, as shown in Figure 2, the PINCLOUD PHR service comprises the following:

---

**Figure 1.** The PINCLOUD PHR service interaction with multiple organizations.

**Figure 2.** The PINCLOUD PHR service interaction with multiple organizations.
A NoSQL data repository that stores patient information in the form of JSON patient documents,

A web portal, through which, patients can access and manage their lifelong health and social information and set their sharing preferences,

A set of web services that implement the PINCLOUD PHR functionality and are called by authorized users and

An authorization system that enables modelling and enforcing patient-centered authorization policies

As regards to security, at each connected to the PINCLOUD organization, are defined access control rules of the form which patient data (objects) is allowed to be exported to the PINCLOUD service (subject) and under what circumstances. For example, a healthcare organization access control rule may imply that patient identification data, critical medical data, health problems, hospital encounters and emergency incidents information can be shared with the PINCLOUD PHR. Moreover, access requests to patient data existing in PINCLOUD PHR are subject to patient sharing preferences. Hence, patients should specify its sharing preferences for use by health analytics systems.

### 3.2 PINCLOUD implementation

The PINCLOUD PHR service has been implemented on a private cloud based on Microsoft Windows Azure and is provided as a SaaS. Moreover, the MongoDB has been used as a NoSQL database for storing patient information and the Liferay as the web portal of the PINCLOUD. In addition, a set of web and REST services has been implemented for the realization of PINCLOUD PHR functionality.

At this stage the PINCLOUD PHR service has been connected through a Service Oriented Architecture (SOA) with the following healthcare information systems:

- The healthcare information system of the private hospital Ygeia
- The home care system of Vidavo
- A healthcare system for physicians called 4doctors of SingularLogic
- An e-prescribing service developed
- An e-referral service developed

As regards patient data coding, in this system implementation, standard terminology systems have been used like the International Classification of Diseases (ICD-10) as well as national standards for coding medications. In addition, the epSOS patient summary has been used as a basis for the creation of a patient summary to be exchanged among the connected systems.

### 4 PHR-BASED HEALTH ANALYTICS EVOLUTION CHALLENGES

The PINCLOUD PHR main objective is to promote knowledge and excellence through the development and investigation of innovative integrated ehealth services using cloud computing, SOA and advanced patient monitoring technologies and set the basis for the realization of a health analytics system. From the technological point of view, PINCLOUD is expected to contribute to the realization of health analytics systems, as an innovative area, deserving to be extensively analyzed and studied, to be applied in the future as real and commercial services. In terms of the economy, PINCLOUD is expected to reduce cost in health care and free up resources (e.g. hospital beds). Last but not least the project is of great importance for the society as patients health data will be documented by the system that can be used by health analytics scenarios in order to provide medical insights. As a result, medical errors associated with the non-integrated nature of health services will be eliminated and, as a result, less people will be affected by these errors.
For the realization of large scale health analytics systems, the lag between data collection and data analysis with the use of analytic algorithms has to be addressed. Hence, important issues and challenges for the realization of integrated big patient data should be identified and appropriate solutions should be found. In particular, the main challenges faced during the PINCLOUD design and implementation related to organizational motivation, technology, interoperability and concerns over data privacy and security are summarized as follows:

- Participant motivation to join the health analytics effort and give permissions for extracting patient summaries from locally stored data to be transferred to the cloud. Critical to the successful implementation of health analytics systems is the engagement and support of the wide range of healthcare stakeholders like healthcare providers, EMR vendors, governments, pharmacies and insurance companies.

- Lack of a practical mechanism to uniquely identify participants. The integrity and value of patient data depends fundamentally on being able to unambiguously link it to one and only one individual. This is sometimes achieved by matching records based on several characteristics (probabilistic matching) instead of using a unique identifier.

- Missing well established laws or regulations mandating the electronic capture of patient data in addition to law covering issues of protection and security of this data. Hence, the EU directives that establish rules on the protection of individuals personal data as well as other international regulations like HIPAA privacy rule should be taken into account and appropriate security measures should be taken in order to safeguard the confidentiality of patient data (Crisholm, 2015).

- Health care data is rarely standardized, often fragmented, or generated in legacy IT systems with incompatible formats like csv, text, XML. Appropriate mechanisms are required in order to transform patient data from one format to another.

- The existence of multiple e-health standards (e.g., DICOM, ISO/TC 215, HL7/CDA) developed by numerous standardization bodies (e.g. ISO, HL7). Many of these are not interoperable or not directly coordinated with each other at an organizational level. Hence, an e-health standard should be selected and then the retrieved patient data mapped into this standard.

- Semi-structured or unstructured data usually demonstrate significant noise and should be properly prepared (e.g. using cleaning and stemming algorithms) before integrated or used for analysis.

Currently, the main benefits gained by the use of the existing PINCLOUD service are summarized below:

- Healthcare providers retained and connected internal data sources and systems with external virtual private cloud computing resources. This is extremely useful for small and medium sized healthcare providers where they can utilize advanced IT infrastructures and services to support their healthcare operations without facing high initial and operational costs.

- Better control of HIT costs and the right sizing of IT investments based on the nature of the workload involved

- Creation of an agile IT environment to support the PINCLOUD PHR, capable of dynamically scaling to meet healthcare needs

- An integrated medical data repository with data existing in a standard format stored as JSON files (according to the JSON schema defined) that can be used by analytics tools in order to transform big data into actionable knowledge. For example, the integrated data can be easily utilized to develop data mining models to discover new medical facts and to conduct medical research to enhance medications, treatments and healthcare services.

The technological solution selected in the PINCLOUD PHR development enable the realization of an integrated big data repository that can then be used by health analytics services. However, regardless the availability of big data and advanced analytics solutions,
there is a need to build analytics competencies to harness ‘semantically integrated data’ beyond ‘big data’ to obtain business insights and improve outcomes (Ferguson, 2012). Hence, our future work in this area is intended to focus more on the data interoperability challenge for patient data from dispersed and heterogeneous resources (e.g. physical resources, streams from cloud, medical sensors and devices). Semantic web technologies like ontologies have been used for exploiting data heterogeneity and for transforming source patient data into the format required to the data warehouse where analysis is performed.
5 CONCLUSIONS

Personal healthcare records are expected to result in a dramatic increase of healthcare data availability. If analyzed properly, this data has the potential to transform healthcare, to contain cost while increasing quality. For example, regulators and policy makers can better define policies that increase healthcare value and safety, pharmaceutical companies can deepen their understanding of diseases and treatments and better direct the design of products; improve their mechanism for recruiting patients for clinical trials and lead innovative smart solutions, such as clinical decision support for personalized treatments, personalized dosage optimization, and adherence analysis (Ferguson, 2012, Raghupathi, 2014).

Quality decisions come from quality data, hence data pre-processing is critical and considerable work is needed to ensure data consistency and validity across sources, platforms and systems (Figo, 2010). Several challenged should be faced before the realization of health analytics systems that include organizational, policy, technological and environmental issues. In this paper, the PINCLOUD PHR service is presented which is a cloud-based PHR that integrates patient data from multiple connected sources. Moreover, the challenges realted to evolving PINCLOUD PHR into a health analytics system are presented.

From a technological perspective, cloud-based ehealth solutions in conjunction with SOA and NoSQL databases have the potential to enable the realization of integrated big data repositories. However, it is up to future research in this field to create more advanced methods and tools, as well as the readiness of the healthcare field to accept and apply these findings and techniques to improve healthcare quality.

References


Zarrouk M. 2014. ‘Cloud computing: Lowering cost complexity barriers in the healthcare industry’.


Raghupathi W. and Raghupathi V. 2014. ‘Big data analytics in healthcare: promise and potential’.

Health Information Science and Systems, 2:3.


MHEALTH: COMMON USABILITY AND USER EXPERIENCE PRACTICES AND FLAWS

Alexandros Mourouzis, Ioanna Chouvarda and Nicos Maglaveras

Lab of Medical Informatics/School of Health Sciences, Aristotle University, Thessaloniki, Greece
{mourouzi; iochou; nicmag}@med.auth.gr

Abstract

Mobile devices show promise as tools for delivering health information and helping users manage their health. This paper aims at identifying common practices among modern mobile health (mHealth) applications available and bring forward typical design errors that spoil the user experience. mHealth, as the practice of medicine and public health supported by mobile devices, is still a relatively new and rapidly developing medium, but that cannot excuse some of the awful user experience issues that users are exposed to on a daily basis. The goal of this effort is to identify any differences between such applications and general purposes mobile applications. A number of mobile applications of various types were inspected by means of heuristic evaluation and cognitive walkthrough, offering useful insights regarding common usability flaws. The analysis offered here, shall help researchers and developers to create more usable applications as well as providing beneficial information for anyone interested in the implementation and assessment of mHealth applications.

Keywords: mHealth, usability, UX, heuristics, evaluation methods.

1 INTRODUCTION

Mobile technologies, particularly high performance smartphones and ubiquitous access to the Internet, are demonstrating these days an unprecedented penetration in all aspects of life. One of the areas expecting a tremendous growth in terms of mobile application development and usage is healthcare industry. Today, through wireless and mobile technologies we have the opportunity to connect information in the real-world via numerous sensors, wearable or fixed sensors embedded in the environment, and produce continuous streams of data on changes of an individual’s health condition and surrounding environment. Such data have the potential to improve our capacity to optimise our health services, in terms of diagnosis and treatment, but also improve our understanding and knowledge of the diseases from a medical perspective.

A number of emerging trends, already happening in healthcare, are nowadays pushing towards more systematic and universal adoption of mobile technologies in healthcare, opening a whole new era of opportunities and challenges for electronic Health (eHealth). For instance, ageing populations and chronic illness are driving a regulatory reform and the need for more accessible, faster, better and cheaper healthcare. At the same time, the foundations for a shift towards “ubiquitous care” care are already in place - industrialisation of healthcare, electronic health records, remote monitoring and communications. And most importantly, healthcare, like other industries, is getting personal for predictive, participatory and preventative care. The need and growth for healthcare mobile applications is related to the willingness of both hospitals and physicians to integrate electronic health records (EHR) in the very near future, the need to integrate the medical records to the pharmacy systems, and the large number of healthcare mobility service providers who are eager to integrate their products into Electronic Healthcare Systems.
This new area of mobile and anywhere healthcare, or simply mobile health (mHealth), has the 
potential to be a transformative force and change when, where, and how healthcare is provided. In 
contrast to the Internet digital divide that limited for years, the development of computerized health 
interventions for lower socioeconomic groups, mobile phone use has been rapidly and widely adopted 
among virtually all demographic groups. In fact, mobile phone usage appears greater among those 
populations most in need of such interventions (Pew Internet, 2010). Mobile penetration in developing 
countries, where wireless technologies have leapfrogged the wired computer infrastructure, have 
produced considerable excitement in the global health community with the prospects of reaching and 
following individuals who were previously unreachable (Kossaraju, et al. 2010).

In figures, the growth mHealth application market has had a continuous upward arc. It is estimated to 
be worth $6.4 billion in 2015, and more than double that in 2016 ($13.5 billion), up from $4 billion in 
2014 and $2.4 billion in 2013. The launch of Apple HealthKit and Google Fit in the latter half of 2014, 
along with the launch of the Apple Watch in the first quarter of 2015, have the world of mobile health 
applications poised to truly go mainstream in 2015. The number of mHealth applications published on 
iOS and Android has more than doubled in two and-a-half years, reaching more than 100,000 apps 
available for iOS and Android in the health and fitness categories in 2014, according to 
research2guidance's fourth annual study on mHealth app publishing. That doesn't mean all 
applications are successful. The competition is fierce. According to research2guidance, 68% of 
mHealth application publishers make less than $10,000 in revenue, and only the top 5 percent make 
more than $1 million. Eighty-two percent of application publishers generated less than 50,000 
downloads with their mHealth portfolios in 2013 while the top 5 percent reached more than 500,000.

So it is not that difficult to create an application on a handheld device that could be used for mHealth. 
The question is which are actually going to be the ones to emerge that are going to be highly useful 
and highly different. It is clear that consumer demand for mHealth applications and sensors has far 
outpaced the science needed to understand their benefits, risks and impact on health outcomes. In the 
small amount of mHealth research conducted to date, issues of safety, privacy, confidentiality, 
regulatory control, and interoperability have been known to hamper researchers’ and developers’ 
efforts.

On the other hand, usability becomes a key factor in the adoption of these applications, which are 
often used by people who have problems when using mobile devices and who have a limited 
experience of technology. Furthermore, in many cases in the field of mHealth, the degree to which 
applications are usable may impact their effectiveness.

This paper aims at identifying common practices among modern mobile health (mHealth) applications 
available and bring forward typical design errors that spoil the user experience. The goal of this effort 
is to identify their differences in relation to practices with general purposes mobile applications.

To this end, a number of mobile applications of various types were inspected by means of heuristic 
evaluation and cognitive walkthrough, offering useful insights regarding common usability flaws and 
resulting into practical guidance for the design and development of more appealing and competitive 
mHealth applications.

2 BACKGROUND AND RELATED WORK

2.1 mHealth

The mHealth field has emerged as a sub-segment of eHealth. Mobile applications and services can 
include, among other things, remote patient monitors, video conferencing, online consultations, 
personal healthcare devices, wireless access to patient records and prescriptions. To date, no 
standardized definition of mHealth has been established. For the purposes of a large scale survey, the 
Global Observatory for eHealth of the World Health Organization (WHO) defined mHealth as
“medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices”. mHealth involves the use and capitalization on a mobile phone’s core utility of voice and SMS, as well as more complex functionalities and applications including general GPRS, 3G and 4G systems, GPS, and Bluetooth technology.

A growing number of countries are using mobile technology to address health needs. The mHealth field is remarkably dynamic, and the range of applications being designed is constantly expanding. Healthcare providers are well aware of the benefits of the mobile apps. Providers are looking for ways on how to improve the care to their patients and make such services more cost efficient. The key applications for mHealth are:

- Communication between individuals and health services
- Communication between health services and individuals
- Consultation between health care professionals
- Intersectoral communication in emergencies
- Health monitoring and surveillance
- Access to information for health care professionals at point of care

### 2.2 mHealth applications

Health applications serve many different purposes, such as providing medical information through a mobile device, mobile wellness applications, and the applications designed to access electronic health records (EHR) and personal health records (PHR). Mosa et al. (2013) reviewed several healthcare applications for smartphones which are documented on MedLine. They categorized the users and purposes of the health apps into three major types:

1. applications for healthcare professionals on disease diagnosis, drug reference, medical calculators, literature search, clinical communication, hospital information system client applications, medical training, and general healthcare;
2. applications for medical or nursing students on medical education; and
3. applications for patients focusing on disease management with chronic illness and other conditions.

<table>
<thead>
<tr>
<th>General category</th>
<th>Sub-category</th>
<th>Share (%)</th>
<th>Share (position)</th>
</tr>
</thead>
<tbody>
<tr>
<td>General healthcare and fitness</td>
<td>Fitness &amp; nutrition</td>
<td>30.9</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Nutrition</td>
<td>7.4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Health tracking tools</td>
<td>*</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Managing medical conditions</td>
<td>6.6</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Medical Compliance</td>
<td>1.6</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Wellness (traditional and corporate)</td>
<td>15.5</td>
<td>3</td>
</tr>
<tr>
<td>Medical information</td>
<td>Reference</td>
<td>16.6</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Diagnostic tools</td>
<td>1.4</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Continuing Medical Education (CME)</td>
<td>2.1</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Alerts and awareness</td>
<td>1.1</td>
<td>10</td>
</tr>
</tbody>
</table>
Remote monitoring, collaboration and consultation

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Remote monitoring (safety)</td>
<td>0</td>
</tr>
<tr>
<td>Remote consultation</td>
<td>0.6</td>
</tr>
<tr>
<td>Remote collaboration</td>
<td>0</td>
</tr>
</tbody>
</table>

Healthcare management

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistical &amp; payment support</td>
<td>0</td>
</tr>
<tr>
<td>Patient health records</td>
<td>2.6</td>
</tr>
</tbody>
</table>

Table 1. Taxonomy of mHealth according to the Research2guidance (2014)\(^1\).

A recent study by research2guidance (2014) has revealed that there are now 100,000 apps which are listed in the mHealth sections of major platform operators: Apple and Android. Both platforms are by far the leading mobile operating systems for mHealth apps today. Over the last two years Android, similarly to the total app count, has seen a tremendous growth of the number of apps which are listed in the Health & Fitness and Medical sections in Google Play. To develop their apps, mHealth app publishers also make use of other mobile SDKs like WindowsPhone and BlackBerry.

The biggest group of mHealth apps could be categorized as fitness apps. More than 30% of all apps that are listed in the Health & fitness and Medical app sections of the various OS stores are fitness trackers or exercise guides. The second and third largest groups are Medical reference (16.6%) and Wellness apps (15.5%). Medical reference apps provide information about drugs, diseases, symptoms and give advice on how to take drugs or what to do in case of experiencing pain. They also show locations of pharmacies and medical centers/doctors. Wellness apps summarize all kinds of relaxation solutions, yoga instructions and beauty tips. Nutrition apps help their users keep track of their diet, inform them about e.g. vitamins, calories and fat content as well as socio-economic aspects of food products (e.g. fair trade). Medical condition management apps represent the 5th largest group of mHealth apps (6.6%). This group consists of all apps which track, display and share user’s health parameters, medicament intake, feelings, behaviour or provide information on a specific health condition e.g. diabetes, obesity, heart failure. Even though they capture notable event and press coverage, all other mHealth app categories (PHR, CME, Diagnostics, Compliance, Reminders and Remote monitoring apps) are significantly smaller in size than the ones mentioned above.

2.3 Usability of mHealth applications

Usability is the effectiveness, efficiency and satisfaction with which specific users can achieve a specific set of tasks in a particular environment.\(^3\)\(^9\) In essence, a system with good usability is easy to use and effective. It is intuitive, forgiving of mistakes and allows one to perform necessary tasks quickly, efficiently and with a minimum of mental effort. Tasks which can be performed by the software (such as data retrieval, organization, summary, cross-checking, calculating, etc.) are done in the background, improving accuracy and freeing up the user’s cognitive resources for other tasks.

Usability evaluation is far broader than the simple process of measuring user satisfaction. Just as importantly, usability metrics include measures of efficiency, effectiveness, cognitive load and ease of learning. Usability emerges from understanding the needs of the users, using established methods of iterative design, and performing appropriate user testing when needed. There are a wide range of design and evaluation methodologies, both subjective and objective, which are continually growing in sophistication. Built-in webcams on modern laptop PCs, robust wireless networking, remote testing software, and compact, inexpensive video recorders make it increasingly easier to “test” in live clinical settings.

\(^1\) The study represents around 10% of mHealth apps.
2.3.1 Usability of mobile applications at large

The advent of mobile devices has presented new usability challenges that are difficult to model using traditional models of usability. Zhang and Adipat (2005) highlighted a number of issues that have been introduced by the advent of mobile devices:

- Mobile Context: When using mobile applications the user is not tied to a single location. They may also be interacting with nearby people, objects and environmental elements which may distract their attention.
- Connectivity: Connectivity is often slow and unreliable on mobile devices. This will impact the performance of mobile applications that utilize these features.
- Small Screen Size: In order to provide portability mobile devices contain very limited screen size and so the amount of information that can be displayed is limited.
- Different Display Resolution: The resolution of mobile devices is reduced from that of desktop computers resulting in lower quality images.
- Limited Processing Capability and Power: In order to provide portability, mobile devices often contain less processing capability and power. This will limit the type of applications that are suitable for mobile devices.
- Data Entry Methods: The input methods available for mobile devices are different from those for desktop computers and require a certain level of proficiency. This problem increases the likelihood of erroneous input and decreases the rate of data entry.

The usability features are determined by four contextual factors:

- Users: This factor is related to the user demographics and profiles of mobile application users. The user’s culture, age, experience with technology and mobile devices, perceptions, etc. will influence the way they use mobile devices and technology.
- Environment: Physical location and environmental types and conditions will affect how users use their mobile devices and access the mobile apps in them.
- Technology: The devices types, interfaces, the access to networks and other technology related factors will affect the way users access their mobile devices and the apps downloaded in them.
- Task/Activity: Based on the task users want to perform in their mobile devices, the mobile apps may be usable or not. Some tasks may be predefined, simple, and closed ones, which increases the likelihood of using mobile apps. Other tasks that the user wants to accomplish may be difficult, complex, open, interactive tasks, and therefore the users may not be inclined to use his mobile device.

3 Method (Procedure)

This work, having in mind to goal of coming up with a list of common usability pitfalls in modern mHealth applications and, ultimately, the development of series of recommendations for designers and developers has designed a review of a number of representative mHealth applications available today on the various OS application stores. Part of this activity was planned as part of a student assessment for an MSc course in medicine at the Aristotle University of Thessaloniki\(^2\). As a starting point, the

\(^2\) under the supervision of the Professors N. Maglaveras and I. Chouvarda at the department of Medical Informatics.
students of the course were asked to formulate teams of two members, and select an application that they found interesting after searching among the various available applications at the major mobile OS stores. They were asked to justify their selection, and keep not of the perceived usefulness and ease of use, based on the information made available by the providers at the corresponding store. Then, the students were asked to provide a profile of their selected application, based on a number of fields and taxonomies predefined by the primary research team. As a next step, each member in team was asked to work individual and inspect independently their application in terms of its usability.

All students have been taught in advance to use common usability inspection technics, such as heuristic evaluation and cognitive walkthrough (see next section). Each team member, was asked to review their application using both these technics, and by making for each inspection at least three turns: one to receive a general understanding of the application navigation, content and functions, one to review in detail specific parts, and one to search for issues missed during the previous two turns. All inspectors (students), apart from identifying and documenting usability issues, they also rated the issues and proposed potential design solutions to address each issue in question. Then, a usability expert (the lead author), reviewed each inspection in order to remove any issues that were not relevant to usability and UI design, merged and harmonised the two inspection reports, and produced an executive summary for each study.

3.1 Heuristic evaluation

Heuristic evaluation is one of the most commonly used discount evaluation methods due to its low cost (Nielsen, 1994). In this method, usability experts evaluate a system using a set of design principles and guidelines, called heuristics. It is a low cost that tends to addresses both local (interfaces) and global (system) usability problems. It is an inspection method that can prospectively uncover problems with a user interface, indicate the severity of the problems, and make suggestions for fixing the problems. Heuristic evaluation can uncover both major and minor problems not necessarily found with user testing. Although major problems are generally easier to discover than minor problems and are the most important to fix, minor problems can just as easily contribute to data entry errors and are easier to find via heuristic evaluation than by other evaluation methods. Heuristic evaluations that are performed by two or more usability experts can identify more than 50 percent of the usability problems with an interface (Nielsen, 1993). The method requires that a small set of 3 to 5 experts evaluate a user interface based on their knowledge of human cognition and interface design rules of thumb or heuristics (ibid.). Once the experts identify the heuristics violated within an interface, experts rate the problems in terms of severity on a scale from 1, indicating a cosmetic problem (fix can wait), to 4, indicating a catastrophic problem (immediate fix).

Some examples of heuristics include visibility (users should always be informed of the system state), consistency (interface design standards and conventions should be employed), match (user model matches system model), minimalist (limited use of extraneous information), memory (minimize memory load by using recognition versus recall), flexibility (shortcuts to accelerate performance), message (good error messages), error avoidance (prevent errors), closure (clear closure on all tasks), reversible actions (undo functions), control (avoid surprising actions), feedback (provide informative feedback about actions), language (utilize the users’ language), and documentation (help options) (Nielsen, 1994). These heuristics were used in this study (see next section).

However, there are some minor drawbacks with this method. Heuristic evaluations are good at exposing the majority of usability problems within an interface. However, heuristic evaluations cannot reveal all problems within a system. Using this technique along with other methods may reveal both local (particular interface screens) and global problems (system issues) (Dumas & Redish, 1999).
3.2 Cognitive walkthrough

Cognitive walkthrough is a usability inspection method that compares the users’ and designers’ conceptual model and can identify numerous problems within an interface (Wharton, et al. 1994). It can be used to evaluate an interface for ease of learning (Polson et al., 1992) and to disclose many problems that a first-time user would encounter with system functionality and ease of system use. It defines how well the interface supports “exploratory learning,” or how well the first time user can perform a task without formal training. It is a technique that focuses on errors in design that would interfere with the users performing a task. It also explains mismatches between the users’ and the designers’ conception of a task. A cognitive walkthrough is conducted by an analyst using predefined scenarios addressing four steps that identify the users’ goals and how easy it is for users to meet these goals. Before beginning this type of analysis, the designer must know the users, the respective tasks they will be performing, and the accurate order of actions for each task. As the evaluator steps through the actions to accomplish a task, he or she tries to answer four questions: (1) will the user try to achieve the correct effect, (2) will the user notice that the correct action is available, (3) will the user associate the correct action with the desired effect, and (4) will the user notice that progress is being made toward accomplishment of his or her goal. If the evaluator answers “no” to any of the questions regarding the action leading to the goal, then that action is considered to have usability problems.

Cognitive walkthroughs tend to find more severe problems, but find fewer problems than a heuristic evaluation, are labor intensive, and require a usability expert. Cognitive walkthroughs have been successfully used to identify problems with health IT applications.

3.3 Overview of the reviewed applications

In total twenty two mHealth applications were studied and analysed, covering a range of application categories.

Ten of them were targeted to patients and citizens (see Table 2):

- **Blood Pressure (My Heart):** For storing blood pressure measurements. Helps collect and analyse blood pressure data.
- **Medicine Alarm Reminder:** Reminds the user to take his / her pills / medicines on time. Multiple notifications and multi-user support.
- **Θερμιδομετρητής:** A calories meter. The user can create favorites lists add items to a basket in order to calculate daily calories intake.
- **Lab Values Pro.** Contains three components for quick medical reference and lab values: Lab Values, Medical Abbreviations, and Medical Prefix/Suffix. On pay (2,99$ ). Number 1 rated medical reference application.
- **Weight Challenge:** A tool for a successful diet project. The user can set challenges to reach specific goals. Track daily calorie intake, exercise and the waist circumference and see how they correlate with weight.
- **Pillboxie:** Aims to make it easy to remember one medicines. Available on pay (0,99$)
- **Diabetes - Glucose Diary:** Tracks and analyses key diabetes data like glucose level, description, tags, which one can edit, view on interactive graphs or send them directly to the doctor.
- **Color Blindness Test:** Online test for colour blindness (short test or long test available).
- **Blood Pressure (BP) Watch:** Collect, track, analyze and share blood pressure records. Get reminded at the right time. Share and backup using Google Drive or Dropbox.
• **Instant Heart Rate**: A heart rate monitoring application, for optimising exercise and track progress. The best Health & Fitness app on Mobile Premier Awards 2011 according to jury of industry experts.

<table>
<thead>
<tr>
<th>Title / Developer</th>
<th>Score / Ratings</th>
<th>Operating system / Business model</th>
<th>Downloads</th>
<th>Category</th>
<th>Main category / Sub-category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood Pressure (My Heart) / Klimaszewski Szymon</td>
<td>4 / 18.748</td>
<td>Multiple / Free</td>
<td>1.000.000</td>
<td>Medical</td>
<td>General healthcare and fitness / Health tracking tools</td>
</tr>
<tr>
<td>Medicine Alarm Reminder / Abhisoft Tech</td>
<td>3,5 / 103</td>
<td>Android / Free</td>
<td>10.000</td>
<td>Medical</td>
<td>Medical information / Alerts and awareness</td>
</tr>
<tr>
<td>Θερµιδοµετρητής / Pinapps.com</td>
<td>3,5 / 644</td>
<td>Android / Free</td>
<td>100.000</td>
<td>Health &amp; Fitness</td>
<td>General healthcare and fitness / Fitness &amp; nutrition</td>
</tr>
<tr>
<td>Lab Values Pro / Hipposoft, LLC</td>
<td>4,5 / 746</td>
<td>iPhone / iPad / On pay</td>
<td>0</td>
<td>Medical</td>
<td>Medical information / Reference</td>
</tr>
<tr>
<td>Weight Challenge / motiMind</td>
<td>4 / 994</td>
<td>Windows mobile / Free</td>
<td>0</td>
<td>Health &amp; fitness / diet &amp; nutrition</td>
<td>General healthcare and fitness / Fitness &amp; nutrition</td>
</tr>
<tr>
<td>Pillboxie / Jared Sinclair</td>
<td>4 / 1.042</td>
<td>iPhone / iPad / On pay</td>
<td>0</td>
<td>Medical</td>
<td>General healthcare and fitness / Medical Compliance</td>
</tr>
<tr>
<td>Diabetes - Glucose Diary / Klimaszewski Szymon</td>
<td>4,2 / 2.399</td>
<td>Android / Free</td>
<td>100.000</td>
<td>Medical</td>
<td>General healthcare and fitness / Managing medical conditions</td>
</tr>
<tr>
<td>Color Blindness Test / CherryCode</td>
<td>3,6 / 2.659</td>
<td>Android / Free plus donate (no ads) version</td>
<td>1.000.000</td>
<td>Medical</td>
<td>Medical information / Diagnostic tools</td>
</tr>
<tr>
<td>Blood Pressure (BP) Watch / NumbersMatter2Me</td>
<td>4,3 / 12.773</td>
<td>Android / Free</td>
<td>1.000.000</td>
<td>Health &amp; Fitness</td>
<td>General healthcare and fitness / Health tracking tools</td>
</tr>
<tr>
<td>Instant Heart Rate / Azumio Inc.</td>
<td>4.3/ 232.921</td>
<td>Multiple / Free</td>
<td>10.000.000</td>
<td>Health &amp; Fitness</td>
<td>General healthcare and fitness / Fitness &amp; nutrition</td>
</tr>
</tbody>
</table>

*Table 2. List of review mHealth applications that were targeted to patients and citizens at large (ordered by number of ratings).*

The rest twelve were mainly addressed to clinicians and health professionals (see Table 3):

• **ACLF Calculator**: This application defines the prognosis of a cirrhotic patient who is admitted to the hospital with acute deterioration with a complication of cirrhosis. The ACLF calculator helps in identifying whether a user has Acute on Chronic Liver failure.

• **Stroke Track**: Stroke track is for tracking and logging stroke cases. It enables the user to capture details of individual stroke patients included detailed NIH stroke scales, potential complications and tPA dose calculation. Individual cases can be saved to a log and reviewed at a later date.

• **Pediatric Growth Charts by Boston Children's Hospital**: A growth charts application for parents caring for their own child or a paediatricians. It allows to track children's growth over time and display the data points on any of the included WHO and CDC growth charts. Can track height, weight, head circumference and BMI and determine the corresponding percentiles.

---

3 According to the OS store taxonomy
4 As categorised further to the research2 guidance taxonomy
• **Medical Calculator**: A medical calculator for helping health professionals with the most common calculations (LDL / VLDL Cholesterol, body Mass Index – BMI, Mean Arterial Pressure, Creatinine clearance, Glomerular Filtration Rate, Glasgow Coma Scale, Pregnancy due date, Cardiovascular Disease Risk Score, etc.).

• **A & P Visible Body**: Presents visually and interactively the core concepts of an undergraduate anatomy and physiology course. The complete app includes cells, tissues and all 11 body systems presented with hundreds of 3D, dissectible model sets, 75+ animations, illustrations, and hundreds of quiz questions. Available on pay (1,11 € per downloaded course unit in-app purchase).

• **eGFR Calculators**: Allows medical professionals to estimate kidney function using five separate eGFR calculators.

• **Prognosis-Rheumatology**: Allows to explore 16 varied clinical cases based on actual patients and update your knowledge on the latest therapeutic guidelines. The clinical cases and discussions have been reviewed by our editorial board comprised of 120 physicians across 27 specialties. Designed to update busy physicians while being an educational tool for residents, medical students and other healthcare professionals studying for academic and licensure exams.

• **Clinical Skills**: An evolving, evidence-based guide to history-taking and the physical exam and simple ECG.

• **Calculate**: Medical calculator and decision support tool, freely available to the medical community. Essential tools in General Practice, Internal Medicine, Cardiology, Surgery, Obstetrics, Nephrology, Hematology, Orthopedics, Pediatrics, Gastroenterology, Neurology, Neurosurgery, Respiriology, and more.

• **Prognosis - Your Diagnosis**: Test out decision making skills in a risk-free environment. Self-assess clinical knowledge on the go. Learn about diseases within minutes. This is an award winning app for practicing physicians to develop and maintain their clinical acumen in a risk-free setting.

• **Medscape**: Is the leading medical resource most used by physicians, medical students, nurses and other healthcare professionals for clinical information. The Medscape app is the highest rated, fastest growing free mobile app for healthcare professionals with over 4 million registered users.

• **Epocrates**: More than 1 million active members, including 50% of U.S. physicians, rely on Epocrates to enable better patient care by delivering the right information, right when it's needed. Disease information, alternative medications, lab guides and more clinical tools and content are available by upgrading to an Epocrates Essentials subscription.

<table>
<thead>
<tr>
<th>Title / Developer</th>
<th>Score / Ratings</th>
<th>Operating system / Business model</th>
<th>Downloads</th>
<th>Category⁵</th>
<th>Main category / Sub-category⁶</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACLF Calculator / CyberLiver</td>
<td>4 / 4</td>
<td>Multiple / Free</td>
<td>100</td>
<td>Medical</td>
<td>Medical information / Diagnostic tools</td>
</tr>
<tr>
<td>Stroke Track / Jim Burke</td>
<td>4,5 / 18</td>
<td>iPhone / iPad / Free</td>
<td>168</td>
<td>Medical</td>
<td>General healthcare and fitness / Managing medical conditions</td>
</tr>
<tr>
<td>Pediatric Growth Charts by Boston Children's Hospital / Pascal</td>
<td>4,5 / 19</td>
<td>iPhone / iPad / Free</td>
<td>0</td>
<td>Medical</td>
<td>General healthcare and fitness / Health tracking tools</td>
</tr>
</tbody>
</table>

⁵ According to the OS store taxonomy
⁶ As categorised further to the research2 guidance taxonomy
Table 3. List of review mHealth applications that were targeted to clinicians and health professionals (ordered by number of ratings).

<table>
<thead>
<tr>
<th>Application</th>
<th>Platform</th>
<th>Rating</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pfiffner Medical Calculator / jhinccap</td>
<td>Android / Free (limited) plus on pay version (full)</td>
<td>3.9 / 110</td>
<td>Medical Information / Diagnostic tools</td>
</tr>
<tr>
<td>eGFR Calculators / National Kidney Foundation</td>
<td>Android / Free</td>
<td>4 / 447</td>
<td>Medical Information / Diagnostic tools</td>
</tr>
<tr>
<td>Prognosis : Rheumatology / Medical Joyworks LLC</td>
<td>Android / Free</td>
<td>4,5 / 798</td>
<td>Medical Information / Continuing Medical Education (CME)</td>
</tr>
<tr>
<td>Clinical Skills / Khalil Rumman</td>
<td>Android / Free</td>
<td>4,3 / 851</td>
<td>Medical Information / Continuing Medical Education (CME)</td>
</tr>
<tr>
<td>Calculate / QxMD Medical Software Inc.</td>
<td>Android / Free</td>
<td>4,5 / 4.886</td>
<td>Medical Information / Diagnostic tools</td>
</tr>
<tr>
<td>Prognosis – Your Diagnosis / Medical Joyworks LLC</td>
<td>Android / Free</td>
<td>4,5 / 8.615</td>
<td>Medical Information / Continuing Medical Education (CME)</td>
</tr>
<tr>
<td>Medscape / WebMD, LLC</td>
<td>Android / Free</td>
<td>4,4 / 34.282</td>
<td>Medical Information / Reference</td>
</tr>
<tr>
<td>Epocrates / Epocrates, Inc.</td>
<td>Android / Free (limited) plus on pay version (full)</td>
<td>3,4/ 68.836</td>
<td>Medical Information / Reference</td>
</tr>
</tbody>
</table>

4 TOWARDS MHEALTH OF BETTER QUALITY AND HIGHER ADOPTION

4.1 Design tips

The last years have seen dramatic changes in the mobile platform landscape. New user interface (UI) paradigms have emerged, screens and processors are becoming as advanced as desktop computers, and input mechanisms have been revolutionized. Within these shifting constraints, designers should always try to create a look and feel that is cutting-edge, memorable and high-performing. This is not just a matter of reducing image sizes. Decisions made at various levels of the design and the design process will have a significant impact. Building the most appealing design is like navigating a terrain with many hurdles. It is a continual balancing act between functionality, aesthetics, usability and performance. Whilst the usability guidelines for designing for mobile devices are still evolving (based on the audience’s developing familiarity with – and our usability knowledge of – the medium), there are some basic guidelines which appear reasonably robust. These guidelines have been derived from a combination of our literature review and our experience of conducting user research, usability testing and user-centred design projects for mobile devices - both for mobile sites and also mobile applications. These guidelines may help usability professionals to design usable and useful mobile applications.
4.1.1 Create and stick to a UI brand identity

Each user interaction with an application should reflect the story of the brand and should increase recognition, loyalty and satisfaction. Identifying which elements contribute most to the brand’s identity is essential. Examples are features, visuals, wording, fonts and animations.

4.1.2 Display content clearly

Displaying content clearly includes grouping information in a clear, consistent way and making sure that information is easily viewable and accessible. Information grouping, or “chunking,” means organizing information in short sections or groups of related items. Chunking reduces the amount of information in one place and helps users better retain the information. Another common pitfall under this guideline is the misuse or abuse of colours, often leading to colour polluted and hard for eye-scanning UIs.

4.1.3 Less is more

Overstuffing an interface can result in an application that is cluttered and difficult to navigate. In the worst cases, critical parts of the interface may actually be downright impossible to see. Apparently, mobile applications work best when they focus on a single main piece of content or functionality at any one time. This usability guideline can be understood to directly relate to two of the main defining characteristics of mobile devices: the screen sizes on the one hand and the user attention are relatively limited. Most mobile devices are currently used while the user is ‘on the go’ or multi-tasking in some form. Thus, design needs to be clearly focused on the presentation and communication of the primary goal at any given point.

4.1.4 Select the right structure and navigation

Mobile devices have their own set of information architecture patterns, too. While the structure of a responsive site may follow more “standard” patterns, native applications, for example, often employ navigational structures that are tab-based. Depending on the nature of the content, there is often a “right” way to architect the mobile site or application, by selecting among the most popular patterns: Hierarchy, Hub & spoke, Nested doll, Tabbed view, Bento box and Filtered view. Use consistent navigation to reduce the burden on the user and make it easier to move to and from information sources. Low-literacy users can be unfamiliar with the function of common navigation features like dropdown menus and often ignore them, preferring “back” and “forward” buttons (ODPHP, 2010).

4.1.5 Ensure user control

Mobile applications represent self-contained environments that allow designers a great deal of freedom with – and control over – the user experience. Designers are often overwhelmed and forget to provide some key options that allow the user to keep control of the interaction dialogue and flow, including the necessary ‘Back’ buttons, and undo and exit options.

---

7 See: http://www.uxbooth.com/articles/designing-for-mobile-part-1-information-architecture/
4.1.6 Minimise user input

Mobile applications designs should generally require the minimum possible data input from users. The main usability reasons for this are that data entry - selection and text entry - on most mobile devices not very easy and requires significant time and attention.

4.1.7 Write actionable content, in simple user language

Action-oriented language is characterized by short statements with positive messages that quickly explain the benefit of a behavior, followed by information on how to take action. Because mHealth applications are often intended to help users adopt healthy behaviors and/or manage their health, action-oriented language is important because it helps users act on the information they find in the application. Action-oriented content should also be written in plain language, which is recognized as one of the keys to health literacy by government, academia, nonprofits, and hospitals (Stableford, 2007). Plain language is communication that allows users to find what they need, understand what they find, and act on what they find (PLAIN, 2013). It isn’t “dumbed-down” language; it is grammatically correct and utilizes full sentences and proper sentence structure, while also making the information easy to understand.

4.1.8 Optimise UI dialogues for performance

Do not leave users hanging. Leaving the user out of the loop when the application is loading or processing could cause users to think the app is malfunctioning. It is also just a poor experience. Google puts “Every millisecond counts” as the second principle of its user experience. Optimizing individual screens, flows and UI elements will reduce waiting times and keep users from thinking that they are wasting their time. Every UI element affects performance. And because every optimization contributes to overall performance, all UI elements should be considered. The number and type of UI elements on the screen will affect the performance of that screen. Furthermore, the characteristics of an element, such as its resolution or image depth, affects drawing time. Then, the way a UI element is drawn by the application affects screen-loading time. Finally, intro animations must not be gratuitously used. Those fun little animations when an application first opens can be really nice, but it is important not to go overboard with them.

4.1.9 Optimise UI dialogues for touch across all screen sizes

Given that the past two months smartphones by major manufacturers ranged from 3.5” to 5.5” two things need to be carefully considered: touch target sizes and placement of controls and information. Any navigation system that needs to work with touch needs to have menu options that can be comfortably used with imprecise fingers. Remember that most users’ index fingers are 1.6 to 2 cm wide. Take into account the width of a finger, plus the fact that users are moving quickly and aren’t able to reliably tap a tiny area of the screen. It also needs to be positioned in a way that aligns with how people hold and use touch-enabled devices.

4.1.10 Optimise UI dialogues for orientation changes

Most users expect a mobile application to respond to device orientation changes - for example, from portrait to landscape - with an appropriate change in layout. Ensuring that mobile designs behave in

---

8 See: http://www.lukew.com/ff/entry.asp?1649
this way has the benefit of not only meeting users’ expectations, which is an important usability principle in itself, but also optimising for the available screen size as mentioned above.

4.1.11 Optimise UI dialogues for the operating system
Bad conversions from one mobile OS to another can confuse and annoy users. Every mobile OS has its own style and the OS’s creator has probably published detailed UI guidelines that have codified their unique aesthetic. It is not necessary to make every application look like it was built by the operating system’s creator, but be careful not to blindly copy style from other OS and turn up with an application that looks like it does not belong on the platform.

4.1.12 Optimise UI dialogues for gesture, according to user expectations
A common pitfall is using a gesture-accessed menu or action as the only point of access, especially in situations that this is not expected and understood by the user. Yet, not every single element of the interface has to be fully visible or easy to get to immediately. In other words, keep gestures in mind, but don’t become overly reliant on them.

4.1.13 Optimise UI dialogues for accessibility and usability for all
One need to consider all the above mentioned guidelines for diverse and the widest possible types users and contexts of use, including for people with disabilities and chronic conditions, the elderly, etc. Adaptivity and adaptability are two powerful tools at the hands of designers to achieve this.

4.2 Development strategy
With such a huge opportunity and fierce competition, the question raised is how are mHelath application developers supposed to stand out and join the top tier of publishers with more than 500,000 downloads? Our study brought forward approach mistakes as well as trends driving mobile health application success in the years to come.

4.2.1 Study the user and seek their feedback
The motivation for the application development is often misguided. Regardless of the elegance, ease of use, enjoyable experience, or other appeal of a health application, if it does not address a specific problem, it will not be considered useful and subsequently not adhered to. People searching for health applications and health information in general are likely doing it because of a health problem. Data must be collected and filtered in a way that it translates a message to the end-user, whether that be a patient or clinician. Co-design, which involves end users in testing an application and applies their feedback, is one of the primary principles in designing user-friendly tools. Co-design is one way developers can learn about users and conceptualize, evaluate, and revise the application. It is unique from usability testing in that it encourages developers to engage end users from the onset—building the application from the start with the end user in mind. Developers often assume that everyone will use an application in mind the same way they do themselves. Usability testing is a must, no matter how good an application looks. Consider organizing a closed beta to small group of trusted people (including a few experienced designers) and update the interface before releasing the app to the public.
4.2.2 Involve clinicians and health professionals

The lack of clinician involvement in development of mobile health technologies continues to dominate the landscape today notwithstanding vendor promises of achieving better patient outcomes at a lower cost and better patient experience. Expert clinician input is necessary on a number of levels. It assures accurate and reliable content, and it leads to a better UX for the clinician with regard to how data is obtained, presented and incorporated into clinical workflow. There are processes that the technology fits into which might very well need to be totally redesigned around the technology (this is a good thing, for many processes need changed). These processes may range from someone’s personal schedule to instituting hospital case managers who advise patients on mobile apps. Knowing the healthcare landscape is critical to determining a strategy of adoption. It is imperative, therefore, to have clinician input into the design of the technology.

4.2.3 Involve representative patients

Achieving the final construction of an application must include an in-depth consideration of the experience a user with the need for the application has. Just as clinician involvement is important in the development phase, so is that of the patient and/or caregiver, who are the data sources. If they are not engaged by good UX design, the technology never takes off and no one knows why.

4.2.4 Study the healthcare landscape and the regulatory specifications

It doesn’t matter how much wow factor the app has, if it doesn’t meet regulatory requirements - security, HIPAA, FDA (if necessary), etc., it will need to be reworked as a significant cost. New proposed regulations regarding handling of data from apps might affect development as well and these should be followed in the news closely. Of course the FDA final guidance document is anxiously being awaited.

4.2.5 Identify and focus on core use cases

Teams often face several project kick-offs in which the initial list of requested features is lengthy, unfocused and impossible to build within the requested timeline. When brainstorming on what a product should do, designers often lose sight of the fact that customers look for solutions that help them with very particular needs. The users need to be able to find and purchase quickly, regardless of whatever other functions that enrich the overall experience.

4.2.6 Prioritise and focus among different versions of the same application

Whether a company wants to launch a product quickly, or develop a product portfolio (i.e., multiple products on one platform, the same product on multiple platforms, or both), or if facing limited time and resources, hard choices have to be made. Design and optimization efforts should be targeted at those products in the portfolio that matter most. A design priority matrix helps us understand where design efforts will pay off the most.

4.2.7 Offer connectivity to other applications

Most mobile health apps today are islands. They don not interact with other applications beyond simple sharing to social networks. This lack of integration hurts mobile health applications in more ways than one. First, it limits their value. Data from other applications cannot flow into or out of the application, limiting utility for the user and creating redundant tasks that require input into multiple application databases.
4.2.8 Offer access to medical databases and cloud services

The demand for apps that allow for monitoring, compliance, and consultation will by default require connectivity to medical databases and patient records at the physician, hospital and service provider offices where care is administered, even if done remotely. This connectivity will be driven in part by hospitals and healthcare organizations and their electronic medical records management services, and by a host of new providers building connectivity tools that act as a secure bridge between application and patient databases. This is the frontier of mobile health applications. Apps that can hook into patient records and patient management systems will enable the valuable interactions and functionality that will drive long-term value and growth. Developers who take advantage of this growing new layer of interoperability can bring new solutions to the market and create a whole new category of powerful mobile health solutions.

5 DISCUSSION AND CONCLUSIONS

There is good reason to be excited over mHealth. Mobile technology can enable much-needed, thoroughgoing change in healthcare systems worldwide and in turn bring significant social and economic benefits. Yet, mHealth is still a work in progress and is growing and changing along with healthcare needs. The mHealth marketplace has lived up to its hype, but where it goes next depends on our changing healthcare needs. While fitness applications are hot now, the future for mHealth applications lies in remote monitoring and consultation.

Access and use of mHealth applications is still on the rise and the opportunity for application developers is clear. Yet, our studies demonstrate the importance of adapting health applications to users’ need. One way to maximize the potential of mHealth applications to improve health is to ensure that these are designed to deliver health information that is simple, engaging, and easy to use for patients and health professional of all literacy levels. There must be age, healthcare and educational literacy and ethnic language appropriateness. The value of an application is simple, intuitive and pleasurable interaction. Efficiency of the presentation, interaction and feedback are important to a good user experience.

Today, all users are reported to experience usability issues, including in finding the right applications for them, and pinpointing the added value in these applications (Franko & Tirrell, 2012). Application consumers are overwhelmed by the great number of tools available for download. Lots of them choose not to download any of them because they cannot decide which ones will meet their needs and they do not want to have too many (health) applications on their mobile devices (van Velsen, et al. 2013). So the initial and potentially one of the greatest challenge for health applications is the quality of the perceived usefulness and ease of use in the process of the selection from the existing large inventory and then of integrating those selected into the user’s mobile device and its associated use routines.

New applications developers need to comprehend that from now on their applications success will not only rely on what their applications do but to a great degree also on how they actually do their thing. Developers and research need to work more towards understanding common usability pitfalls and what may affect the user experience of an mHealth application user. They need to consider appropriate design guidelines and development strategies that fit the specific characteristics of this emerging medium for health care delivery and management.
Acknowledgments

Part of the work reported here has been carried out in the framework of national project Providing Integrated eHealth Services for Personalized Medicine utilizing Cloud Infrastructure (PinCloud) [4], led by the University of Piraeus, conducted in the context of the National Strategic Reference Framework NSRF 2007-2013, and co-funded by the EC.

References


A SMARTPHONE APP FOR EFFICIENT NOTE KEEPING IN CONCURRENT MULTI-PARTICIPANT BEHAVIOUR STUDIES

Leandros Stefanopoulos, Aristotle University of Thessaloniki - Lab of Medical Informatics, lstefano@auth.gr
Christos Maramis, Aristotle University of Thessaloniki - Lab of Medical Informatics, chmaramis@med.auth.gr
Ioannis Ioakeimidis, Karolinska Institutet – Division for Applied Neuroendocrinology, Ioannis.Ioakimidis@ki.se
Ioannis Moulos, Aristotle University of Thessaloniki - Lab of Medical Informatics, joemoul@med.auth.gr
Nicos Maglaveras, Aristotle University of Thessaloniki - Lab of Medical Informatics, nicmag@med.auth.gr

Abstract

When an experiment with multiple participants is undergoing in the context of a behavioural study, there is a need for taking notes in an easy yet organized manner. Moreover, the scientific staff needs to easily review and evaluate each note for the benefit of the experiment. Modern technology can help in this direction by providing computer applications (e.g., Smartphone or Tablet apps) that replace manual note-keeping. In the current paper we present the design and implementation of Memorandum, a novel Android application for Tablets that facilitates real-time note-keeping in concurrent multi-participant behavioural experiments. Memorandum has been developed to cover the computer-assisted note-keeping needs of the behavioural studies of SPLENDID, an EU research programme aiming to monitor and modify the eating and physical activity behaviour of young people at risk of obesity and eating disorders. Memorandum emphasizes note keeping speed and usability by providing a visual participant identification mechanism and selectable lists of the most common observations, accompanied with a free text note option.

Keywords: smartphone application, note-keeping, behavioural studies, multi-participant experiments

1 INTRODUCTION

Making observations during the execution of a scientific experiment (also known as note keeping) is an indispensable step in all good-practice scientific processes. Traditionally, this has been executed manually. However, when speed and accuracy are the primary objectives, manual note keeping might prove to be inadequate. This is the case when we have to deal with concurrent, multi-participant experiments, i.e., experiments where multiple participants execute tasks concurrently. In such experiments, note keeping speed and accuracy become imperative, especially if the assistants that are present in the experiment are limited in number.

Computer-assisted note keeping, i.e., ICT applications that allow their users to make electronic observations is the perfect solution for the aforementioned problem. The computer applications that have been designed for facilitation of note keeping context in an experimental environment are called electronic lab notebooks (ELN). Recently, the advances in mobile technologies have provided a new, more convenient family of devices for electronic note-keeping, with Smartphones and Tablets being the more prominent examples. However, although there are plenty of general-purpose note keeping apps for Smartphones and Tablets, none has been reported for scientific/experimental application.
This shortage in apps has been recognized by SPLENDID, an EU-funded programme that aims to develop a novel preventive intervention for young people at risk for obesity and eating disorders (ED) (Maramis et al., 2014): In order to evaluate the developed intervention, SPLENDID has to conduct a series of multi-participant pilot studies where high-school students concurrently consume – and record with a help of sensor – a joint meal in the cafeteria of their school. To annotate the meal recordings and ensure the quality of the collected recordings an efficient note keeping tool has to be employed during all the studies/experiments. For this purpose, we have designed and developed Memorandum, an Android app best suited for Tablets to allow the assistants of the experiments to quickly and accurately take notes about the behaviour of the participants during the experiment. Memorandum has been designed from scratch and implemented with the aforementioned requirements in mind, in addition to the objective of offering a high usable and friendly User Interface (UI).

The paper is structured as follows: Section 2 presents the previous work in computer-assisted note keeping. Section 3 describes the methodology that has been followed for designing and implemented Memorandum, while Section 4 provides the end results, i.e., the Memorandum App for Android Tablets. Section 5 concludes this work.

2 RELATED WORK

In (Walsh and Cho, 2012), the authors provide an extensive report on employing Evernote as an ELN in their laboratory. The results from their evaluation indicate that users of paper lab notebooks were primarily discouraged by the inability to freehand into a notebook. On the other hand, the advantages of using Evernote were cardinal, including ease of use, flexibility, ease of searching, ease of information sharing, accessibility, online storage, drawing and audio recording, limited cost (free basic edition, and low-cost pro edition), security and legality, and simplicity and unstructured note keeping.

BURRITO is another example of ELN. It is a Linux-based provenance collection and note-taking system. The platform consists of several plugins with functions like audio recordings (microphone), command invocations (Bash, Python, MATLAB), text editor interactions (Vim), web browsing history (Firefox, Chrome), sticky notes and to-dos (Xpad), digital sketches (Wacom pen tablet), etc. A sidebar residing on the left portion of the user’s desktop background represents the activity feed, i.e., a near real-time stream of the user’s actions as a list of feed events in reverse chronological order, with the user interface template inspired by Facebook and Twitter. The feed displays the following events: Bash commands, website visits, file modifications, digital sketches CRUD events, status updates (created by the user), as well as Checkpoint CRUD events (created by the user).

The BURRITO system is most useful when scientists try to work through months-old data. In such situations, the annotated feed events can help them rearrange their thoughts in a faster and more detailed manner. Moreover, BURRITO provides interaction capabilities with file modification events, allowing the users to retrieve older versions of the files they created. Also, it includes two tools that prove to be of much help: The first is the Computational Context Viewer, which connects the source code file with the output code file disregarding any modifications that happened afterwards. The second tool is the Activity Context Viewer, which captures and logs the changes made from one modification to another. Finally, the Lab Notebook Generator functionality exports everything the users select in HTML files for sharing or printing (Guo and Seltzer, 2012).

InkSeine is a Tablet PC application that supports active note taking by coupling a pen-and-ink interface with an in-situ search facility that flows directly from a user’s ink notes (Hinckley, Ken, et al., 2007). InkSeine integrates four key concepts: (1) it leverages preexisting ink to initiate a search; (2) it provides tight coupling of search queries with application content; (3) it persists search queries as first class objects that can be commingled with ink notes; (4) it enables a quick and flexible workflow where the users may freely interleave inking, searching, and gathering content. InkSeine offers these capabilities in an interface that is tailored to the unique demands of pen input, and that maintains the primacy of inking above all other tasks. A short walkthrough of an InkSeine usage scenario is as follows:
The user inks “BumpTop” in his electronic notebook, and then decides he needs a PDF document.

The user lassos the ink to create a breadcrumb.

The user then opens the breadcrumb. Since the breadcrumb is persistent, this could have been opened either immediately or sometime after it was created.

Opening it brings up the results and the user taps the PDF filter to narrow down the results set and locate the desired document.

At the next step, the user drags the PDF icon into his notebook for future reference. The user then decides to grab a piece of content from this PDF document.

He employs the document’s icon to open the PDF and a circular tracking menu appears with some pen-operated controls for manipulating the underlying document. He touches the pen down in the outer ring of the tracking menu and circles with the pen to scroll through the document.

When he sees the desired content, he taps the Capture icon in the tracking menu. This places a transparent gray overlay on the screen and he sweeps out a rectangular selection region.

The user hits the Close icon in the tracking menu to close the PDF and return to InkSeine.

The clipping appears in the user’s notes, and

The user draws several ink annotations that refer to it.

3 METHODOLOGY

3.1 Designing Memorandum

The application has been developed to assist the note-keeping procedure during the pilot studies of the SPLENDID programme. The user of the app is one of the experiment assistants. The target device is an Android Tablet with a screen of approximately 10 inches, assuming a landscape holding position.

In order to ensure the speed and accuracy of the note-keeping process, two primary challenges that had to be tackled during the design of the app were: (1) the identification of the student/participant for which we want to keep a note, and (2) to allow the quickest possible entry of observations into the system.

To tackle the first challenge, we employed a sophisticated identification procedure based on QR codes and a visual representation of the room where the experiment took place. More specifically, each participant of the experiment was supplied with a QR code containing an alphanumeric string that served as the unique identifier of the participant throughout the experiment. Moreover, we incorporated in the app a site layout plan with clickable tablets and/or seats. Before the start of the actual experiment (i.e., meal recording) the user of the app has to assign a QR code to each taken position. This assignment is stored in the database of the app. When the actual experiment start, the user can click on the same site layout plan to select the participant for which a note need to be taken. Since the assignment of the selected seat to a QR code is known, each note that is recorded is automatically associated with the appropriate participant via his/her QR code.

With respect to the second challenge, several alternatives (freehanding, typing, audio recording, etc.) were considered. Our conclusion was that a predefined list of carefully selected common observations would have the maximum impact of the note-keeping speed. For this purpose, we asked a few collaborating researchers and assistants with great experience in this kind of experiments to provide the most common observations for such experiments. Based on their input, we compiled the list of available observations, taking into account to keep it short by including the five most common observations. However, to avoid disregarding least possible observation, we decided to include a sixth option for free-text insertion of the observation.
Memorandum includes 3 main functionalities. All of them can be initiated from the Home screen of the app. The first functionality is the registration of the individuals into the database by means of the QR codes and also their assignment to specific seats (Register Participant). When this functionality is initiated, the user of the app selects from the table layout plan the table where the participant seats and then – in the next screen – the exact seat of the participant on the selected table; then, the user can scan the QR code of the participant to register him/her to the selected position. To facilitate the registration process of many participants, the app minds to indicate a seat that has already been registered by displaying it enclosed in a red box.

After registration is completed, the actual note-keeping functionality can start (Keep Note). This functionality is well synchronized with the actual experiment and, as a consequence notes have to be taken in “real-time”, especially if many events that need to be noted happen during the experiment. Each time the assistant thinks that a participant action need to be noted, he/she selects – same as in the first functionality – the table and then the exact seat where the participant seats. This opens the common observations as a multiple-choice list for the assistant to select the appropriate entry or, alternatively, input the observation in free text.

Lastly, there is the Export Notes functionality, which is meant to be used after the end of the experiment. Here, the assistant has the option to export the notes he/she made in a CSV file and afterwards to view the exported report with any CSV Viewer application currently available on the tablet. The flows of screens that have been designed to carry out the three functionality of Memorandum are shown in Figure 1.

![Flowchart](image)

**Figure 1.** The flows of screens that carry out the three functionalities of Memorandum. From left to right: Register Participant, Keep Notes, Export Notes.

In the back-end, Memorandum uses an SQLite database to store the registration and note-keeping information. This is a very simple relational database consisting of two tables. The first table (Registration) stores information about the registration process (i.e., the assignment of QR codes to seats); it includes the following fields: table number, seat number and QR code. The second table (Note) holds information about the notes that are made for a particular position/seat; this table includes the fields: note and timestamp of the note/entry as well as a registration identifier acting as a foreign key in the Registration table. The schema of the employed database is presented in Figure 2.
Figure 2. The schema of the SQLite database that is employed by Memorandum.

When the Export Notes functionality is executed, three CSV files are created. One file is created for each table of the database, while a third file is created from the result-set an SQL query for those QR codes that have at least one note associated with them, along with the notes themselves and their timestamps. This file is one that can be viewed from within the app in the third screen of the Export Notes functionality (Show Report screen).

From each screen of the app, the user can choose the option to return to the Home screen. Register Participant and Keep Note are considered repeated functionalities; so when one “iteration” of these functionalities is completed, the app returns once more to the beginning of the same functionality. The exact appearance of the discussed screens is given in Section 4.

3.2 Implementing Memorandum

Memorandum has been developed in Android Studio for devices with minimum SDK: API 16, i.e., for devices that run Android 4.1 (JellyBean). Nowadays, approximately 88.7% of the android devices support applications compiled for JellyBean (Dashboards| Android Developers, May 2015). Moreover, all the auxiliary functionalities of the application are carried out with the help of JAVA libraries embedded in the code of the application; specifically ZXing for QR code reading and OpenCSV 2.4.0 for CSV files manipulation. Concerning the minimum hardware requirements of Memorandum, a low-resolution rear camera is required for QR tagging and less than 10MB of available storage space in the tablet’s internal (or external) SD card for storing the database and the CSV files.

Several popular Android classes have been employed for the development of the UI of the app: Buttons, ImageButtons, RadioButtons (Radio Group), ImageView and EditText, etc. The scaling and the position of these visual elements are defined statically by the corresponding XML layout files, while the runtime associated runtime actions are defined by JAVA methods. All the graphic icon that are used in the Memorandum were designed from scratch for the app using GIMP (GNU Image Manipulation Program) v2.8.14.

In total seven Android Activities were developed, one for each screen that is listed in Figure 1. The creation/update of the Info (Notes) table is undertaken by the last Activity of the Register Participant (Keep Note) functionality.

4 Results

During the development, debugging was performed on a Samsung Galaxy Tab 2 (GT-P5100). Subsequently, the app was tested on an Asus Transformer Pad TF103C and a Powertech TAB-04 and no problem and no bugs/issues were detected. On March and April 2015, Memorandum was successfully used for keeping notes in 2 pilot studies performed at the Internationella Engelska Gymnasiet Södermalm in Stockholm (Sweden) in the context of the SPLENDID programme. One instance of the app was used to write down notes about forty participants in the two experiments.
The screens of Memorandum are presented in the figures below, following the naming and order of Figure 1. As shown in Figure 4, this screen displays the exact layout plan of the room where the experiment takes place. The numbered boxes are clickable and represent the tables. When a table is clicked, a toast informs the user of the made selection and then the screen depicted in Figure 5 appears: The seats of the selected table – its number is shown in the centre of the table image – are displayed with their corresponding numbers and, if they have already been registered, with a red-squared outline. By selecting a seat, the camera is activated waiting for a QR code. When the QR code is detected, it gets decrypted and the resulting alphanumeric string is stored in the database along with the corresponding seat and table numbers. Automatically, the user is redirected to the activity with the table layout plan, in order to continue the registration process with minimum delay. When the registration is complete the user selects the *Done Registering* button and gets redirected to the Home Screen (Figure 3).

![Figure 3. Home Screen - The three buttons correspond to the three main functionalities of the app.](image)

Once the registration process is over, the assistant can proceed to the actual note keeping by selecting the *Monitor Session* button of the *Home Screen*. On click the user is redirected to a screen identical to the one shown in Figure 4 (table layout plan) with the exception that the *Done Registering* button has been replaced by the *Done Monitoring* button. The procedure for selecting table and seat number is identical to the one described above and the same holds for the red-squared outline of the already registered seats. Clicking on a seat, the assistant is redirected to screen of Figure 6, which displays the selected table and seat number and presents a list of fixed observations. The assistant selects the one that best suits the current event; he/she can also add custom text to a fixed sentence or select *other* from the list and type a custom sentence of his/her own. When the assistant presses the *Submit* button, the app gets redirected to the table layout plan of *Keep Note* functionality, so that he/she can repeat the procedure quickly to submit as many notes as needed while the experiment takes place. When the last observation is submitted, the assistant presses the *Done Monitoring* button in order to return to the *Home Screen*.

Once the *Export Notes* button is selected on the *Home Screen*, Memorandum presents to the assistant the screen depicted in Figure 7. The *Export to CSV* button exports the three CSV files that were mentioned in Section 3 and stores them in the filesystem of the Tablet’s OS. If the export is successful, the *Show Report* button becomes clickable. By selecting, the CSV file containing all the notes submitted along with their corresponding QR codes, the table numbers, and seat numbers is displayed in a new screen within the app.
Figure 4. Select Table Screen – The table layout plan with clickable tables.

Figure 5. Select Seat Screen – Table seat layout with clickable seats. Red boxes are used to indicate already registered seats.

5 CONCLUSION AND FUTURE WORK

In this paper, we have presented the design and implementation of a novel Android app for Tablet, aiming to facilitate note-keeping in the context of concurrent, multi-participant experiments. Memorandum covered a real need in scientific/research practice by providing functionalities to fast
and accurate registration of observation, contributing towards the efficient annotation of the datasets resulting from the experiments.

Figure 6. Choose/Insert Note Screen – Multiple-choice list including the most common observations. For custom observations the user first selected the “Other” option and then inserts the observation either by typing or – if supported by the hardware – by freehanding. A red circle indicates the seat for which the note is being taken.

Figure 7. Export Notes Screen – By pressing the “Export to CSV” button the three CSV files mentioned in Section 3.1 are exported; after that, the “Show Report” button becomes available.

The next step will be to evaluate the usability of the app in real experiments (e.g., by means of SUS questionnaires). The analysis of the evaluation results will provide valuable feedback for the improvement of the app. Moreover, we consider the addition of new note-keeping features in the app.
such as audio/video recording and image capturing. Finally, we intend to enhance Memorandum with customization capabilities: this refers to the table layout plan, the number of seats per table, the list of common observations etc. All these variables can be viewed as configuration parameters that, in the future, should be fed to the app to customize it at execution time.

REFERENCES


Hinckley, Ken, et al. 2007. 'InkSeine: In Situ search for active note taking.' Proceedings of the SIGCHI conference on Human factors in computing systems. ACM.


DashBoard Android Developers :
Innovative Research Projects

Chairs: Dr. Demosthenis Kyriazis
# Table of Contents

**Track – Innovative Research Projects**

4. [PID-110] Vasileios Anagnostopoulos, Nikos Vlastaras and Dimosthenis Kyriazis, “IMPLEMENTING FAILURE DETECTION AND REPORTING IN A LIBVIRT CLOUD THROUGH AD-HOC MECHANISMS”
6. [PID-3] Leonidas Katelaris “E-Referral - PINCLOUD”
A NOSQL DATABASES COMPARISON FOR E-HEALTH PROJECTS

Nikolaos Chelmis: Department of Digital Systems University of Piraeus, Greece
helnik@unipi.gr
Dimosthenis Kyriazis: Department of Digital Systems University of Piraeus, Greece
dimos@unipi.gr
Marinos Themistocleous: Department of Digital Systems University of Piraeus, Greece
dimos@unipi.gr

Abstract

The continuous development of Cloud Computing and the emerging interest on Internet of Things has led to new applications with greater demands from databases. New needs, for storing and processing big data effectively, high performance and availability when reading and writing, led traditional databases to face many new challenges. To this end NoSQL databases have been created. In this paper NoSQL databases are categorised based on their data model and a comparison between Document Oriented NoSQL databases is presented.

Keywords: NoSQL, Performance, Cloud Computing, Big Data.

1 INTRODUCTION

The concept of relational model was introduced with E.F.Codd’s paper (Codd, 1970) which made data modelling and application programming much easier. The proposed model was well suited to client-server programing and have proved to be predominant technology for storing structured data in web and business applications.

The term NoSQL was first introduces by Carlo Strozzi for his RDBMS, Strozzi NoSQL (Strozzi, 2013) in 1998. The term was used only to distinguish Strozzi’s solution from other RDBMS that utilize SQL but his solutions was still addressing to the relational model. Recently the term NoSQL with his current meaning (not only SQL) has come to describe a large class of databases which do not have properties of traditional relational databases and which are generally not queried with SQL (structured query language). The term revived in the recent times with big companies like Google / Amazon using their own data stores to store and process huge amounts of data as they appear in their applications and inspiring other vendors as well on these terms.

Storing healthcare information electronically was introduced several decades ago since paper-based records could not meet the requirements of advanced health care systems. Electronically stored healthcare information has been identified by a number of different names such as Electronic Patient Records (EPR), Computerised Patient Records, Electronic Medical Records and Electronic Health Records (EHR) (ISO2004, Narayan et al., 2010). Establishing a nation-wide EHR system requires a significant investment as well as extensive system design and project management (Hoerbst et al., 2010, Vest, 2012). Poorly designed architecture not only poses a substantial failure risk for the implementation of EHR systems, but also can cause significant losses of financial and human resources.

There are numerous obstacles and challenges in relation to EHR systems mentioned in the literature, such as standardisation of vocabulary, security, privacy and data quality which are out of scope of this paper. In addition to these matters (Orfanidis et al., 2004) claims that the expanding size of healthcare data also creates an obstacle for EHR systems. The increasing diffusion of information systems in healthcare delivery and increasing size and heterogeneity of the healthcare data has resulted in a bottleneck for storage, retrieval, high availability and analysis aspects of traditional relational
databases. NoSQL database systems might be the solution to this bottleneck in addition to providing many other advantages (Jin et al., 2011, Schmitt and Majchrzak, 2012).

In this paper both a qualitative and quantitative comparison is conducted in the existing noSQL solutions in order to decide which one better suits the needs of Pincloud project (Koumaditis K. et al., 2014). Pincloud aims at providing integrated e-Health services for personalized medicine utilizing cloud infrastructure. Technologies like cloud computing and Service Oriented Architectures are used to provide efficient, scalable, portable, interoperable and integrated IT infrastructures that are cost effective and maintainable.

The rest of the paper is structured as follows: Section 2 presents the background theory needed by shortly presenting the main theorems which relational and noSQL databases are based on, while Section 3 discuses why noSQL databases are needed for e-health projects. Section 4 presents the classification of NoSQL databases and the comparison between the selected ones, while Section 5 concludes the paper.

2 BACKGROUND THEORY AND RELATED WORK

2.1 CAP Theorem

NoSQL database systems have received much attention from research community (Leavitt, 2010, Cattell, 2011, Stonebraker, 2010, Fan, 2010, Han et al., 2011a). As suggested in literature current research mainly focuses on scalability, fault tolerance and performance advantages fault-tolerance and performance advantages of the NoSQL / distributed database systems, while criticising the weak consistency approach. The issue of consistency is explained in the context of CAP (consistency, availability, and partition-tolerance) theorem which was introduced by Eric Brewer in 2000. It suggests that there is always a trade-off between consistency, availability and partition-tolerance. Consistency means that each server returns the right response to each request while availability means that each request will eventually receive a response and finally partition-tolerance means that the service can continue operating normally even when communication between some nodes are lost. The underlying idea of CAP theorem is that the communication between servers is prone to network errors and failures, thus it is not possible to have all three features working together.

In figure, Icomparative strengths of NoSQL databases and relational databases in terms of CAP theorem are summarized.
Figure 1. Comparison of three main data model types used in NoSQL databases with relational databases in terms of CAP Theorem.

2.2 ACID
In 1981 Gray (Gray, 1981) suggested four properties for database systems to achieve reliable transaction processing which are known as Atomicity, Consistency, Isolation and Durability (ACID). Atomicity stands for the status of a transaction which can be either completed entirely or failed (i.e. there is no partial completion). Consistency is a property which makes sure that every transaction changes a database into a valid new state, incorporating all rules, constraints and triggers. Isolation means that each transaction exists totally independent and do not affect each other while being executed. Durability is the property that means if a transaction has been completed, the new state of database is guaranteed to be durable regardless of any potential failures such as power loss, network errors, etc. afterwards.

2.3 BASE
Due to their distributed nature noSQL databases cannot offer strong consistency models like traditional relational databases and cannot have all strong ACID properties. So they focus on the Basically Available, Soft state and Eventually (BASE) principal. This principal states that a system can continue working as usual in case of a failure due to the distributed nature of NoSQL databases and even though there is no guarantee of consistency at any given point of time, data will eventually be consistent at some point in time.

Table 1 summarizes the main differences between ACID and BASE

<table>
<thead>
<tr>
<th>ACID</th>
<th>BASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>The availability of the system usually depends on the availability of the physical machine on which the system is installed (single point of failure).</td>
<td>Ensuring availability in cases where one or more nodes come off. Ensure the availability of data of all nodes.</td>
</tr>
<tr>
<td>System data are stable and can only be changed after a user interaction.</td>
<td>The system data may be changed without user interaction (e.g., data with expiry date).</td>
</tr>
<tr>
<td>A multipart transaction is considered successful when all parts of which consists succeed. If any of them fails, the transaction is entirely rejected.</td>
<td>Most NoSQL systems do not support multipart transactions. Those that argue however, ensure that the multipart transactions have the status of isolation.</td>
</tr>
<tr>
<td>The system guarantees the consistency of the data at any point of time. This means that system data is incorrect, according to rules such constraints, cascades, triggers etc.</td>
<td>The system cannot guarantee continuous data consistency. Conversely, it cannot guarantee that given enough time, the system turns out to be consistent (eventual consistency).</td>
</tr>
</tbody>
</table>

Table 1. ACID Vs BASE

3 NEED FOR NoSQL SOLUTIONS
With the continuous development of cloud computing and it’s descendant the Internet of Things (IoT) various types of applications have emerged, which made database technology more demanding, especially in the aspects of (Fan, 2010, Han et al., 2011b):

1. High concurrent of reading and writing with low latency
   Traditional databases needed to meet the needs of high concurrent reading and writing with low latency at the same time.
2. Efficient big data storage and access requirements
Large applications, such as search engines, need databases to meet efficient data storage while being able to respond to the needs of millions of traffic.

3. High scalability and high availability

With the number of concurrent requests and data increasing databases need to be able to be easily expanded and upgraded in order to ensure rapid uninterrupted service. The fundamental advantage of NoSQL databases is that they allow scaling up to large datasets without any changes in the overall structure of data model or architecture. Hardware requirements and costs can grow linearly as storage requirements grow. Cost-effective scaling up is made possible and high initial investment in hardware requirements are avoided (Lakshman and Malik, 2010).

4. Lower management and operational costs

The increasingly mount of data led to increased database costs including hardware, software and operating costs. Therefore, there is a need to lower the costs for big data storage.

Relational databases have occupied a great portion of the data storage area but when facing the aforementioned requirements some limitations exist:

1. Slow reading and writing

The logic which relational databases are build are prone to bring up deadlocks and other concurrency issues as the data size increases.

2. Limited capacity

Existing relational databases cannot support big data

3. Expansion difficulties

Multi-table correlation mechanism which exists in relational database, became the major factor of database scalability (Han et al., 2011a).

Modern e-health projects like Pincloud have certain needs that noSQL solutions seem to cover. Table 2 summarizes the main requirements of a healthcare system and how noSQL database address these requirements

<table>
<thead>
<tr>
<th>Requirement</th>
<th>noSQL addressing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increasingly healthcare data over time</td>
<td>Horizontal scalability which allows to easily and automatically scale</td>
</tr>
<tr>
<td>Heterogeneous data (e.g. free-text, images etc)</td>
<td>noSQL databases offer data models which allow unstructured or semi structured data to be easily stored</td>
</tr>
<tr>
<td>Data should always be available since it is a major requirements for the continuity of healthcare services.</td>
<td>Eventual consistency (as described in section 2.3)</td>
</tr>
<tr>
<td>Access to data from multiple locations which requires high-performance system.</td>
<td>Comparisons between noSQL and relational databases indicate that noSQL databases offer higher performance.</td>
</tr>
</tbody>
</table>

Table 2. E-health requirements compared and how noSQL databases address them

4 CLASSIFICATION AND COMPARISON

NoSQL databases have some inadequacies such as they do not support SQL which is industry standard, lacking of transactions, reports and other additional features, not mature enough for most of the NoSQL database products were created in recent years and so on. Each product
complies with the different data models and CAP theorem. Therefore, we will introduce the mainstream noSQL database data models, and classify NoSQL according to CAP theorem.

Table 3 briefly presents the data modes, use cases of them and finally products based on each data model.

<table>
<thead>
<tr>
<th>Data Model</th>
<th>Use Cases</th>
<th>Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key-Value</td>
<td>It is preferred for simple nature data that can be represented in a key -</td>
<td></td>
</tr>
<tr>
<td></td>
<td>value form/session management, shopping cart, etc.).</td>
<td>Memcached, membase, Voldemort, Redis</td>
</tr>
<tr>
<td>Document</td>
<td>Preferred for data that follow the form of documents (documents, prescriptions, referrals, etc.).</td>
<td>MongoDB, CouchDB, Couchbase, RethinkDB, Terrastore, Elasticsearch</td>
</tr>
<tr>
<td>Oriented</td>
<td>It is preferred for data which have high correlation among them and</td>
<td>Neo4j, FlockDB, OrientDB, AllegroGraph, GraphDB</td>
</tr>
<tr>
<td></td>
<td>the correlation must be imprinted (social networks).</td>
<td></td>
</tr>
<tr>
<td>Column</td>
<td>It is preferred for very large sized data tables, which are usually</td>
<td>BigTable, Hbase, Hypertable, Cassandra</td>
</tr>
<tr>
<td>Oriented</td>
<td>conferred questions about a few of the available columns (data warehouses, CRM, etc.).</td>
<td></td>
</tr>
<tr>
<td>Object</td>
<td>It is preferred for applications that store data relating to physical or</td>
<td>VelocityDB, Versand, Objectivity, Starcounter, Perst, HSS Database</td>
</tr>
<tr>
<td>Oriented</td>
<td>virtual objects (CAD, CAM, video games, etc.).</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Data models, use cases and products

Based on the CAP theorem and different concerns of noSQL database a classification would be (Hurst, 2010):

1. Concerns about consistency and availability (CA)
2. Concerns about consistency and partition tolerance (CP)
3. Concerns about availability and partition tolerance (AP)

4.1 Qualitative Comparison

For e-health projects, as mentioned before, the main goal is to ensure big data storage and good query performance. To this end document based databases are preferred. Typical document databases are MongoDB and CouchDB.

Apache CouchDB is a flexible, fault-tolerant database. It’s main features are:

1. Supports data formats such as JSON and AtomPub
2. It provides REST-style API.
3. Ensure data consistency by complying with ACID properties.
4. Provides a P2P-based distributed database solution that supports bidirectional replication.

However, it also has some limitations, such as only providing an interface based on HTTP REST, concurrent read and write performance is not ideal and so on.

MongoDB stands between relational and non-relational databases. It’s main features are:

1. It is non-relational database, which features the richest and most like the relational database.
2. Supports complex data types: MongoDB supports bjson data structures to store complex data types.
3. Powerful query language: it allows most of functions like query in single-table of relational databases, and also supports indices.
4. High-speed access to mass data: when the data exceeds 50GB, MongoDB access speed is 10 times than MySQL.
5 CONCLUSIONS

In this paper firstly the background of NoSQL was described. In a second moment the need for NoSQL databases was presented and a contradiction between NoSQL and RDBS was conducted. Finally NoSQL databases were classified based on the CAP theorem.

Based on the above, when deciding to use NoSQL, consideration must be taken to ACID transactions demand, cost etc. For an e-health project specifically options like (1) Capacity, (2) Performance and (3) Reliability.

For these reasons, after objectively analysing the strengths and weaknesses of current mainstream document-based NoSQL databases, MongoDB was selected for the PINCLOUD project.

ACKNOWLEDGMENT

The research leading to the results presented in this paper has received funding from the European Union and the Greek National Strategic Reference Framework Programme (NSRF 2007-2013), Project PinCloud under grant agreement number “11SYN_6_1013_TPE”.

References

FAN, K. 2010. Suvey on Nosql Programme, pp.76-78.
ISO2004 Ts 18308 Health Informatics-Requirements for an Electronic Health Record Architecture.


A CROSS LAYER MANAGEMENT FRAMEWORK FOR ACHIEVING ADDED VALUE IoT SERVICES

Achilleas Marinakis, Dept. Of Electrical and Computer Engineering, NTUA, Iroon Polytechniou 9, Zografou 15773, Athens, Greece, achmarin@mail.ntua.gr

Spyridon V. Gogouvitis, Dept. Of Electrical and Computer Engineering, NTUA, Iroon Polytechniou 9, Zografou 15773, Athens, Greece, spyros@ntua.gr

Dimosthenis Kyriazis, Dept. Of Electrical and Computer Engineering, NTUA, Iroon Polytechniou 9, Zografou 15773, Athens, Greece, dimos@mail.ntua.gr

George Kousiouris, Dept. Of Electrical and Computer Engineering, NTUA, Iroon Polytechniou 9, Zografou 15773, Athens, Greece, gkousiou@mail.ntua.gr

Panagiotis Bourelos, Dept. Of Electrical and Computer Engineering, NTUA, Iroon Polytechniou 9, Zografou 15773, Athens, Greece, mpourelo@mail.ntua.gr

Orfeas Voutyras, Dept. Of Electrical and Computer Engineering, NTUA, Iroon Polytechniou 9, Zografou 15773, Athens, Greece, orfeasvoutiras@gmail.com

Theodora Varvarigou, Dept. Of Electrical and Computer Engineering, NTUA, Iroon Polytechniou 9, Zografou 15773, Athens, Greece, dora@telecom.ntua.gr

Abstract

In a world of multi-stakeholder information and assets provision on top of millions of real-time interacting and communicating Things, COSMOS project aims at enhancing the sustainability of smart city applications by allowing IoT based systems to reach their full potential. COSMOS will enable Things to evolve and act in a more autonomous way, becoming more reliable and smarter. Things will be able to learn based on others experiences, while situational knowledge acquisition and analysis will make them aware of conditions and events potentially affecting their behaviour, in both proactive and reactive way. COSMOS will enable them to create and defuse knowledge by promoting decentralized techniques in the context of social IoT. Things will be semantically described through a social-related ontology, meaning that they will have social characteristics mainly based on others Things feedback while communicating with each other. COSMOS also exploits data and information management mechanisms to handle the exponentially increasing “born digital” data. Management decisions and runtime adaptability will be based on Things security, trust, administrative aspects, location, relationships, information, and contextual properties. Capabilities of IoT systems will be enhanced with aspects from Cloud computing, data management strategies and intelligent systems.

Keywords: Social Internet of Things, IoT data management, Intelligent Things, COSMOS EU project

1 INTRODUCTION

The main goal of COSMOS [COSMOS project] is to provide the mechanisms needed to make Things, in the IoT domain, smarter while also to increase the interoperability of various hardware and software solutions by generalizing the use of semantics and ontologies. Smarter Things will behave more autonomously using cognition loop and situation awareness, not focussing only on current situation but also on expected future ones, using prediction based on machine learning and Complex Event Processing (CEP) techniques. They will be also empowered with social capabilities, like experience acquisition and sharing, based on strong trust and reputation mechanisms. COSMOS aims therefore to
provide the developers with the means, in the form of platform and functional components, to integrate multitudes of different data sources, as well as to use the data processing capabilities, storage, information retrieval, CEP and other COSMOS functionalities supporting autonomous and social behaviours as outlined above.

2 TECHNICAL CHALLENGES AND OBJECTIVES

COSMOS faces the challenge of providing an execution platform / environment with built-in privacy and security approaches, spanning from the application to the storage and the devices levels, while addressing pertinent issues related to data management and optimization of Things in terms of reliability and adaptability to situations. In particular, COSMOS intends to:

- Complement and extend Things semantics to capture dimensions of their social behaviour: building on top of existing social media technologies, COSMOS aims to describe Things using social characteristics in order to achieve enhanced services like discovery, recommendation and knowledge sharing between them. These characteristics will describe the relationships between Things as well as the type and the quality of their interactions. COSMOS will build its own trust and reputation model whose purpose is to analyse the behaviour of Things within a collaborative social network and therefore to detect those which may act in a malicious way and exclude them;

- Extract knowledge from unreliable and incomplete data: in the world of IoT, devices and sensors are deployed or used in varying conditions and different situations. Mostly they are deployed in remote places and are connected using less reliable wireless links. In order to prolong their battery life, data provided by these devices may be sporadic and less reliable. These data themselves are of no value until they are processed intelligently to extract high-level knowledge which can be used to make decisions. Thus, COSMOS aims to provide the tools to analyze real-time raw data, in order to predict future states, combining different machine learning and CEP techniques;

- Enable Things to act in an autonomous and decentralised manner, based on others Things experiences: Things create dynamic networks that can have millions of nodes, often bounded by their location, and which operate under different administrative domains, which may have different rules. The latter highlights that the centralized management of such networks is inefficient. By exploiting the social characteristics of Things, often associated with the corresponding ones for their owners, COSMOS aims at providing decentralized management and coordination decisions based on service, interaction, location and reputation-oriented principles. Such principles are core in social networking technologies and their applicability in the services domain will enhance current management and coordination mechanisms enabling them to deal with a big number of Things. COSMOS will enable Things to make decisions, find and execute plans, based on their own or others Things knowledge, which have similar characteristics in terms of location, type, domain, goals or physical resources;

- Enhance end-to-end security and privacy in and across networks of Things: security, privacy and trust aspects are central in IoT given that besides the sensitivity on the data, IoT applications may influence the physical environment through the deployed sensors and actuators. COSMOS aims at enhancing security, privacy and trust within and across different layers of IoT applications; service, device and data. Regarding the service aspects, social media technologies will be extended to identify reputation and trust patters within a network of Things. On the devices level hardware coded security and privacy approaches will be provided, while on the data level emphasis will be on storage-related approaches. Given that devices may be utilized in different applications / contexts, cross-application security and privacy mechanisms will be developed by considering different applications as tenants. COSMOS will provide tools for data filtering giving the end-users the ability to protect their personal information;

- Provide scalable data and information management exploring the interplay between storage and computation: for an IoT platform such as COSMOS, there are several key phases in the information lifecycle. Massive amounts of data need to be ingested and analyzed in real time. Data may need to be stored persistently in order to enable historical analysis on them and also to be archived over time. The IoT domain presents many challenges in the domain of information
and data lifecycle management. The IoT domain requires large scale data management at low cost. Data will be generated by a large number of devices and will need to be ingested into the system reliably in real time. Moreover, incoming data need to be analyzed in real time and in a way that enables reacting to events detected by the analysis. In addition many kinds of analysis can only be done with data collected over a period of time. Therefore data need to be collected persistently in order to support search and analysis on historical data. In order to support low cost, a scale out architecture using commodity hardware components is necessary. In addition, new data will continually be born into the system and storing all information is costly. Therefore data reduction and archiving techniques are needed in order to reduce the cost of storing the data.

3 STATE OF THE ART ANALYSIS

This chapter provides an analysis of the State of the Art (SotA) for the different technical domains, raised from the challenges mentioned above and its purpose is to explain how COSMOS, with respect to the current SotA, will go beyond it.

- Things semantics: ontologies provide a means to describe sensors, their environment and the measurements they take. COSMOS will enhance existing ontologies by providing a means to describe Things’ behavioural and social aspects as well as formalize the way they accumulate experiences. Baseline ontologies will be reviewed including the W3C Sensor Ontology, e.g. SSN-XG [Michael Compton et al, 2012], as well as the loosely related mobile service Ontology defined in IST-SPICE [SPICE project];

- Complex Event Processing: CEP has originated from the work on an object-oriented language designed for system architectures prototyping with powerful event oriented semantics. Although businesses have been doing for decades near real-time event processing in a form or another [Leavitt, 2009], CEP as seen in current software solutions has been introduced rather recently. The term itself has been popularized by D.C. Luckham in his book “The Power of Events: An Introduction to Complex Event Processing in Distributed Systems” shifted from the analysis of hardware discrete-event systems towards distributed software architectures. Within COSMOS, CEP methodologies will be enhanced and optimized so as to become an enabler for the implementation of next generation applications by providing increased situational awareness and reliability levels;

- Decentralized and autonomous Things management: solutions exist for the management of IoT based infrastructures. The management of Things in COSMOS will be based on social-interaction principles that will use real time analytic concepts for automatic adaptability of the formulated networks. Moreover Things in COSMOS will be highly autonomous using experiences of others to gain heightened situational awareness. Things descriptions and management functions, i.e. self-management and learning supported by cognitive technologies, have been/are being defined in the iCore project [iCore project], COSMOS will build on the publicly available mechanisms. Tools for experience sharing will be exploited based on the outcomes of SOCIOS project [SOCIOS project];

- Security, privacy and trust profiling of Things: security in specifications like ZigBee [ZigBee Security, 2009] is based on a trust centre and key pairs. The Zig-Bee standard has a security architecture based on an online trust centre that is in charge of handling the security relationships within a ZigBee network. This is a strategy that the research was focused over the past three decades in traditional distributed systems, mobile ad-hoc networks and applications. A variety of key agreement and privacy protection protocols that are tailored to IoT scenarios have been introduced in the literature. Security, privacy and trust in COSMOS will be treated holistically in order to provide an end-to-end solution. To this end Things in COSMOS will be profiled and linked with the corresponding administrative domains and context parameters in order to enhance IoT governance frameworks with such properties. COSMOS will also develop adaptive selection approaches based on the aforementioned profiles to manage the uncertainty and volatility of the sources. Privelets is a concept introduced by COSMOS to prevent information inference and preserve privacy;
Information and data lifecycle management: cloud storage systems allow for the definition of metadata describing the stored data. There are numerous storage providers today in the cloud like the one offered by Amazon, in the form of the Amazon S3 services [Amazon], or OpenStack [OpenStack] which is an open source cloud computing software framework originally based on Rackspace Cloud Files [Rackspace Cloud Files]. VISION Cloud project [VISION Cloud project] introduces an infrastructure for reliable and effective delivery of data-intensive storage services, facilitating the convergence of ICT, media, and telecommunications. COSMOS will enhance VISION Cloud project data model by targeting specifically the IoT domain, using its rich metadata to model Things and their relationships. Existing metadata support in VISION Cloud and OpenStack Swift will be used as baseline technologies.

4 HIGH LEVEL ARCHITECTURE AND FUNCTIONAL VIEW OF COSMOS

In this chapter we provide a Functional Decomposition of the COSMOS architecture according to the generic Functional View (FV) provided by the Architectural Reference Model (ARM) in the IoT-A project [IoT-A project]. COSMOS functionalities can be grouped into the following groups:

- Security, providing functionality such as key management and authentication;
- Data Management, providing functionality such as persisting data on the cloud storage;
- Things Management, providing functionality such as discovery of Things based on semantic queries;
- Analysis, providing functionality such as prediction and situational awareness.

The high level view of COSMOS is depicted in Figure 1. A colour code is used to group the components. Worth noting that, from a deployment point of view, the COSMOS framework will include services both at a platform level as well as within Things. Therefore, COSMOS will also require Things to provide a set of common services and interfaces, such as Experience Sharing, called COSMOS-services. These services will form the basis for the decentralized and autonomic properties envisioned by the project. One of the key design choices that have been made is that the different components of COSMOS will offer a REST-full HTTP interface. The choice of HTTP REST has been made as it is widely adopted in the IoT and cloud domains, enables interoperability and allows for easy development. For the same reason JSON has been chosen as the COSMOS data format, even though other options were explored. It should also be noted that there are multiple actors relating to COSMOS, such as Things, applications and human end-users. All of these actors, may interact with some of the functionality offered by COSMOS. For example, a human actor may access the Data Storage, retrieving for instance a data object, the same way that a Thing may access it.

![Figure 1. COSMOS Functional View](image-url)
5 SOLUTIONS

COSMOS functionalities presented above, aim to address the challenges described in chapter 2. Specifically, the Social Monitoring component alongside with Social Analysis are responsible for capturing, measuring and analysing the social behaviour of Things based on their relationships and interactions with other ones. All this information is stored in a social-related ontology that is introduced in COSMOS and uses three types of descriptors; relationships between Things e.g. friendship, social indexes e.g. trust index and social metrics e.g. knowledge sharing. Event Detection and Prediction components analyse the raw data generated by the Things in order to extract valuable knowledge which can be used to avoid undesired future situations. Planner and Experience Sharing components, which run inside the devices, enhance the autonomous behaviour of Things by enabling them to react to events which occur, exploiting their own knowledge or asking similar Things, considered as friends, for help. COSMOS hardware security approach ensures that the access to the cloud storage data is denied to unauthorised users, whereas the so-called Privelents component, running at the Things side, enables the human users, owners of the Things, to keep their private data safe. Metadata Search component is used in order to efficiently retrieve the cloud storage data according to their metadata, which is a functionality of high importance in the IoT domain which deals with large amounts of data.

6 CONCLUSIONS AND FUTURE WORK

COSMOS project builds a framework for making Things more human-like, enabling social behaviour aspects, discovering and retrieving suitable solutions from their friends, while monitoring and ranking reliability and reputation of the contributors. It also provides device side hardware-based security enablers and flexible software-based privacy filters as well as intelligence extraction mechanisms. Next steps include to design and deploy reusable business applications exploiting all the COSMOS capabilities for diverse needs, improving efficiency and cost reduction and thus offering a positive economic impact to the enterprises.

ACKNOWLEDGMENTS

The research leading to these results is partially supported by the European Community’s Seventh Framework Programme under grant agreement n 609043, in the context of the COSMOS Project.

REFERENCES

COSMOS project; http://www.iot-cosmos.eu/

SPICE project; http://www.ist-spice.org/
iCore project; http://www.iot-icore.eu/
SOCIOS project; http://www.sociosproject.eu/
Amazon; http://aws.amazon.com/s3/
OpenStack; http://wiki.openstack.org/
Rackspace Cloud Files; http://www.rackspace.com/cloud/cloud_hosting_products/files/
VISION Cloud project; http://www.visioncloud.eu/
IoT-A project; http://www.iot-a.eu/
A DESIGN APPROACH TOWARDS MVCC PROVISION BY A DOCUMENT DATA STORE FOR ENABLING TRANSACTIONAL SEMANTICS

Pavlos Kranas, Dept. Of Electrical and Computer Engineering, NTUA, Iroon Polytechniou 9, Zografou 15773, Athens, Greece pkranas@mail.ntua.gr
Sotiris Stamokostas, Dept. Of Electrical and Computer Engineering, NTUA, Iroon Polytechniou 9, Zografou 15773, Athens, Greece sotirisl@mail.ntua.gr
Vrettos Moulos, Dept. Of Electrical and Computer Engineering, NTUA, Iroon Polytechniou 9, Zografou 15773, Athens, Greece vrettos@mail.ntua.gr

Abstract

Recent challenges in cloud applications have made the use of NoSQL data stores more popular than ever. These data stores however lack of the important guarantees that traditional relational databases offer: ACID properties and transactional semantics. Implementing ACID transactions has been a longstanding challenge for NoSQL systems. Because these systems are based on a sharded architecture, transactions necessarily require coordination across multiple servers.

In this work, we propose an extended version of the much popular MongoDB data store, which provides transactional semantics and ensure ACID compatibility, while on the same time, maintain all its non-functional characteristics. This is achieved via its integration with an external conflict resolution service, which is responsible for identifying conflicts between modified documents of concurrent transactions and provide ACID properties. MongoDB’s server side core however remains identically with the native version, thus providing the high performance and scalability properties that makes it so popular.

Moreover, having identifying that our proposition cannot be exploited by distributed cloud applications that need to span a transaction across different nodes, we propose an alternative solution that can be used instead, which can be scaled across layers so as to ensure the required non-functional requirements, as throughput and latency.

Keywords: MongoDB, Transactions, Snapshot Isolation, MVCC.

1 INTRODUCTION

The emergence of the cloud ecosystem brought numerous business to migrate their solutions into this new paradigm, with a great amount of diversity regarding their requirements for data management. Common applications mostly used to rely on online transactional processing (OLTP) workloads, where classic relational database management systems (RDBMS) were widely adopted. These databases can be accessed via the well-known SQL language, and ensure ACID database properties during the execution of concurrent transactions, thus providing transactional semantics. However, cloud applications needs are enriched with additional requirements, such as high availability and the ability to elastically scale in numerous instances in order to cope with variable workloads. However they do not require such strong level of high data consistency. Additionally, the emergence of IoT technologies leads to the explosion of big data, where strong consistency is not a first-class citizen. Big data on the other hand requires a database to be able to scale accordingly in order to ensure low latency and high throughput under increased workloads, while cloud applications often require high availability.
In order to cope with this demand, a new category of data store systems have been emerged during the last years, widely known as NoSQL data base systems [1][2]. Apart from their much simplifier data model, one of the main differences, among others [3], is the lack of support of transactional semantics. NoSQL databases have been built from the start to be partition tolerant (the P from Eric Brewer’s CAP theorem) and mostly focuses on availability (the A), rather than consistency (the C) that is provided eventually. Being partition tolerant, a NoSQL database can be scaled up to numerous amounts of nodes, thus being able to provide very low latency and serve increased throughput. The ensurance of ACID compatibility would require the transactions to span across different nodes, thus introducing an additional overhead which NoSQL data stores are not willing to pay.

The obvious drawback of using a NoSQL datastore is the lack of transactional semantics. For the most key-value stores, this is an acceptable trade-off, as the majority of data that are being stored are no critical, and corresponds to big data produced by IoT sensors or other monitoring data, where only analytical operations are being performed and no often modifications are expected. Document data stores, however, provide a much more complex data structure, thus allowing the application developer to exploit the latter for storing the application’s data model. This leads records to be accessed under OLTP workloads, where frequent data modifications occur which brings the requirement for the data access layer to provide transactional semantics. As NoSQL datastores cannot ensure ACID properties, this is usually handled at the application level. The application developer has to implement all the complexity already provided by traditional databases, which can make the code much more error prone and difficult to be maintained. As the popularity of the document stores increases, and among others MongoDB is widely used by cloud applications, the need for document data stores to provide transactional semantic is getting more popular.

In this paper, we present the design of an extended MongoDB data store, that can be integrated with an external transactional management service, in order to delegate to the latter the detection of conflicts among concurrent transactions making MongoDB ACID compatible, while on the same time, maintaining all its non-functional attributes that natively provides, such as its high availability, its rich query interface and its ability to be partitioned in numerous shards. In order to achieve this, we extend its data store client so as to be integrated with an external conflict resolver, in order to provide transactional semantics, but we keep the server-side identically to MongoDB’s official version, in order to continue to provide its native functionalities. We rely on the Snapshot Isolation (SI) paradigm which requires multiple versions of data items to be stored in data database. This additionally raises the requirement for the data store to implement the MultiVersion Concurrency Control (MVCC) mechanism.

The remainder of the paper is organized as follows: section 2 offers some overview of key concepts, section 3 briefly describes relevant work towards the provision of transactional semantics to NoSQL data stores, section 4 presents our design approach and section 5 offers a brief description for a prototype that can support transactions across multiple shards.

2 TRANSACTIONAL SEMANTICS AND MONGODB DATA STORE

The ACID properties in database systems are the following:

- **Atomicity**: Either *all* the operations transaction occur, or *nothing* occurs
- **Consistency**: The database must be consistent after the transition from the state it was before the beginning of a transaction to the state the transaction ended.
- **Isolation**: The transaction must behave as if it is the only operation being performed upon the database.
- **Durability**: Upon completion of the transaction, the operation must not be reversed and upon a rollback of a pending transaction, the database must return to its previous state.

Transactions in traditional RDBMS provide full ACID properties usually by making use of a two phase commit (2PC) protocol using shared and exclusive locks on the database records. The RDBMS
distributes locks in the datastore node in order to prevent concurrent transactions to access records that are already read or modified, according to the selected isolation level [22]. Each isolation level can lead to specific anomalies due to the concurrent execution of transactions. The higher the isolation level is, the less the types of anomalies that can appear. This can be achieved by increasing the number of locks that are distributed during a transaction, with the obvious performance trade-off. This approach can ultimately result to increased delays under heavy write intensive loads, where a transaction must wait for a lock to be released, which finally would ruin the performance characteristics of a NoSQL datastore.

2.1 Snapshot Isolation

Instead of using locks for preventing access on documents that are in conflict during concurrent operations, the Snapshot Isolation (SI) approach can be adopted. It relies on MVCC and timestamp ordering and it is usually lock-free. MVCC demands that every document in the datastore is additionally marked with a monotonically increased number. When an update operation occurs, a new version of the same data time is created, with an incremented timestamp, instead of updating the already existed item. Transactions uses two timestamps for performing CRUD (create, read, update delete) operations: A transaction is assigned with a start timestamp upon initialization. Each transaction has access only to a specific synchronous snapshot of the data. When a read operation occurs, it reads all records from the datastore that satisfy the select query. Those versions that are marked with a timestamp are the latest before the transaction’s start timestamp. When a transaction modifies records, it creates new versions and adds them to its private writeset. A transaction can only access its private versions, so as to read-its-own-writes and all publicly visible data, which have been already committed according to its timestamp. Thus formulating the snapshot that a database can access, giving the impression that is isolated. When a transaction tries to commit, and no conflicts have been identified so far, it assigns a commit timestamp. Then, it sets the private writeset versions to the value of this committed timestamp, and finishes.

Using the Snapshot Isolation paradigm, the aforementioned database anomalies are not feasible, due to the fact that each transaction reads only the latest updated versions that existed before the transaction began. Writes that overlap reads cannot appear, as they will have a timestamp whose value is greater than the transaction’s current. This means that read-write and write-read conflicts cannot occur and a conflict resolution engine must only keep track for write-write conflicts. Read operations cannot block, thus is giving a performance advantage over locking strategies that can be used for ensuring ACID properties.

2.2 MongoDB support for ACID properties

Most NoSQL databases often relaxes the level of support of the traditional ACID properties, in order to achieve better performance, higher availability and more flexibility in storage. MongoDB natively does provide some basic transactional capabilities though.

Atomicity in MongoDB is provided in the level of a single document. This means that a transaction that affects multiple documents in the same collection (i.e. a massive update based on a predicate criteria) is not atomic. Transactions that perform multiple write operations on documents that are stored in different collections have also the chance to partially fail, breaking down the ‘A’ property of ACID-ity. However, operations that perform multiple updates to subdocuments that are all included into the same parent -document, are atomic. This means that operations that must be atomic can be designed to fit in the same document, under its children. However, this introduces a significant programming complexity, due to the fact that it is the application developer responsible for retrieving the appropriate documents, rather than the data store engine itself.

Ensuring consistency is one of the biggest challenges in distributed databases. Due to the CAP theorem, there’s a trade-off between consistency, availability and partition-tolerance. Error! Reference source not found. That provides the definition of 1-copy-serializability, which ensure that all transactions performing operations in a replicated database, will produce the same result in each replica as if they were performing serially. In MongoDB eventual consistency is only guaranteed. In a
replicated set, one server will be the primary one, and the others will be the slaves. Every time, only the primary node is receiving write operations, and then MongoDB guarantees that the changes will be flushed to the slaves eventually. Regarding the read operations, an application provider must choose between reading explicitly from the primary node, or distribute the traffic in all nodes, but sacrificing the level of consistency.

**Isolation** is supported only in single documents by default. Updates on multiple documents will allow other operations to interleave with these updates. If these interleaved operations contain writes, the update operation may produce unexpected results. MongoDB can provide isolation when updating multiple documents only if it is explicitly indicated to do so. However, this does not provide transaction atomicity; neither is possible when MongoDB is configured as a sharded cluster. By default, the database’s model is similar to the transaction auto-commit used in the relational databases. Every performed operation is immediately visible to other concurrent transactions.

Regarding the **durability** of a transaction, MongoDB uses a write-ahead-log, called ‘journal’ where every operation is logged. The journal is committed every 100 milliseconds by default, with the ability for the user to change this parameter. This means that in the rarely event of a system failure, changes to the database that have been reported to the journal after the latest commit will be lost. However, there is the provision for preventing a transaction to commit until the journal is flushed to the database and the latter is fully synchronized.

### 3 Current Approaches

A significant effort from the industrial and academic community has been made in order to incorporate the benefits from the two separated worlds: the RDBMS that offer transactional semantics and the NoSQL data stores that offer scalability and partition tolerance properties, but sacrifices the level of commitment to the ACID properties. Authors of [4],[5] propose a solution based on distributed transactional memory for managing concurrent access to the distributed data grid, implemented by Red Hat’s Infinispan. [8]They implement a two-phase commit protocol over distributed nodes that coordinate in order to ensure ACID properties over a cloud storage service. Authors of [11] introduce a key group abstraction to allow efficient transactional access to an underlying data store. [16] That system relies on a cloud storage service for storing data in a cloud ecosystem and present a solution for guaranteeing ACID properties of a traditional relational database on top. Furthermore another approach [10] implements a transactional management layer that can scale appropriately to serve different workloads. Warp [9] makes use of the concept of linear transactions for providing transactional semantics for key-value datastores. Finally, Authors of [6] and [15] ensure ACID properties to key-value stores.

Most of the aforementioned solutions are designed to offer transactional semantics either only to key-value datastores, or to an underlying storage service layer. Their common characteristic is that operations to the datastores take into account only the key of a specific entry. MongoDB however is categorized as a document data store, thus allowing queries on the various fields of a document, which makes this type of datastore incompatible with those solutions. TokuMX [17] provides an updated version of MongoDB, offering versioned documents so that each transaction can access only specific versions, according to the Snapshot Isolation paradigm. The MVCC ACID-compliant mechanism is based on its fractal tree technology that is also used for MySQL. However, multi-statement transactions are not ensured in sharded deployments. MongoDB’s recently updated release on the other hand, can now be integrated with the WiredTiger storage engine [18][25]. The latter supports transactional semantics with lock-free algorithms using optimistic concurrency control without a centralized lock manager, giving MongoDB MVCC functionality and the ability to execute multi-statement transactions. However, in its documentation [19], it is stated that the phantom anomaly is allowed which is contradictory for a lock-free transactional management implementation that relies on MVCC and snapshot isolation. Furthermore, a recent report from Symas [20] also highlighted that that *writers in this storage engine block readers*, thus concluding that it is not valid with a MVCC design.
4 IMPLEMENTATION OVERVIEW

Mostly inspired by the fact that nowadays there are a numerous implementations that provide transactional semantics on top of key-value datastores, our work is focused primarily on providing MVCC functionality to native MongoDB datastores across multiple shards, thus enable the latter to be integrated with an external transactional service that relies on the Snapshot Isolation approach. As described in the relevant section, most of these implementations are designed to be used by key-value datastores, or storage engines that access data by their identifier. Towards this direction, each transaction in our prototype that performs update operations will have to provide the modified documents’ identifiers to the transactional service. The latter can identify possible write-write conflicts on concurrent transactions that try to modify the same records, if the keys of the modified documents are provided. As a result, the external service will be used as an external conflict resolver, which will notify our prototype in case of conflicts so as for the latter to take appropriate actions in order to ensure the ACID properties. Finally, our prototype will exploit external services for appropriate database recovery upon failures [21], thus ensuring the ‘D’ property of ACID. The focus of this paper is to highlight the overview design principles of our prototype that enables it to be integrated with already delivered services. This section gives a high level architecture of the proposed implementation and justifies the major implementation decisions that have been made.

4.1 Architecture Overview

The main objective of the prototype is to implement all necessary functionalities that MongoDB natively offers, while additionally ensuring ACID-ity when concurrent transactions co-exist. This will allow an application developer to deliver solutions that can exploit MongoDB’s native capabilities, enhanced with ACID transactional management, offered by an external conflict resolution engine. The bird-eye-view of this architecture is shown in Figure 1. From this diagram, the following main components can be identified:

- **Application**: This is the application that will make use of our proposed solution, integrated with an external conflict resolution engine.
- **MongoMVCC-Driver**: The datastore client that will be utilized by the application in order to access the MongoDB.
- **External Conflict Resolver**: An external service that can be used for the identification of write-write conflicts of concurrent transactions.
- **MongoDB**: The native MongoDB server.
- **MongoDB-Ext**: This component contains all required extensions of the MongoDB native server that ensures the integration of MongoDB with the external conflict resolver and implements all additional functionalities.
Figure 1: Execution of a Transaction in our extended MongoDB

The general description of the workflow that takes place during the execution of a transaction is as follows:

- The application needs to create a new transaction. It firstly must communicate with the external conflict resolver, through a local client implementation that is included in the MongoMVCC-Driver, in order to initialize the transaction.
- The external conflict resolver returns back the transaction’s context, and most significantly, its start timestamp.
- At this point, the application is capable of interacting with MongoDB, using its MVCC capabilities that provides the corresponding versioned data items, according to the transaction’s start timestamp. Every update operation (insert, update, delete) keeps a private version of the affected data item in the private context of the transaction, managed by the MongoMVCC-Driver.
- When the application is ready to commit, then it informs external conflict resolver so that the latter can firstly perform all necessary actions in order to prepare the commit phase.
- If no write conflicts are identified until this phase, then the external conflict resolver must update the transaction context with the commit timestamp and inform the application to prepare to commit.
- Before committing, MongoMVCC-Driver has to pass the transaction’s private write-set to the logger service, in order for the latter to make it durable so that proper recoverability will be enabled in case of potential failures. Our prototype will not rely on its native mechanism, as the latter introduces small time intervals where the data cannot be considered durable and can lead to write losses [12]. Consequently, MongoDB will provide the logger service with the full private write-set (all updated versions of data items in the context of a transaction) so that the latter can ensure the durability property.
- Once the durability is granted, in order to finalize the commit, the MongoMVCC-Driver flushes all changes to the native MongoDB datastore. MongoDB-Ext registers all pending transactions, so that it can enforce recoverable actions, in the case of a potential failure.
- If failure occurs during the finalization of the commit phase, then MongoDB-Ext identifies it and requests the write-set of the failed transaction from the logger service.
- The logger service provides the requested write-set and MongoDB-Ext proceeds with the necessary undo/redo actions, in order to recover.
The aforementioned diagram hides many implementation details regarding our proposed prototype. For instance, MongoMVCC-Driver interacts asynchronously with the external conflict resolver every time that it needs to check for a write-write conflict in the scope of a pending and open transaction. This could lead to the abortion of a transaction, due to a write-write conflict from another concurrent transaction. The details of the implementation are described in the following subsections.

4.2 Multi-Versioning and Version Labelling

Multi-versioning is a first-class citizen in a data store that needs to be integrated with a conflict resolution engine that relies on the Snapshot Isolation paradigm, which requires that all transactions can see different versions of data items, according to their start timestamp. MVCC is not natively supported by MongoDB. Our approach is mainly inspired by the work that has been done in PostgreSQL, as described in [13][24]. As a result, every versioned document that is stored in the database additionally contains the following metadata:

- **id**: is the MongoDB’s required unique key of every document stored in the data store. It is a 20 bytes concatenation of unique `dataID` and `commitTmstmp`, which allows versions of the same data item to be stored next to each other.

- **dataID**: is a 12 bytes data item identifier, of type ‘ObjectId’, provided by MongoDB’s native libraries that auto generates this value in a distributed manner. It ensures uniqueness of the key across the various instances.

- **commitTmstmp**: This is an 8 bytes long timestamp, provided by the external conflict resolver, upon commit.

- **nextCommitTmstmp**: This is an 8 bytes long timestamp. If the document is at its latest version, then this field must be null. Otherwise, it points to the next commit timestamp of the next updated version.

- **isDeletedFlag**: This flag indicates whether or not an item is deleted.

MongoDB supports all basic CRUD operations. Every one of those is implemented by our implementation as follows:

- **Read**: Read operations are provided by a predicate condition that is being applied by MongoDB on the data store. Documents that satisfy this condition are being selected and retrieved. In our prototype read operations additionally require a start timestamp. Using the transaction’s timestamp and documents’ metadata, it is always ensured that only the latest undeleted version that satisfies the search condition predicate will be retrieved. MongoDB’s internal query mechanism is responsible for selecting the proper documents in a range scan, without having to introduce an additional performance overhead by additionally iterating through the result data set, in order to retrieve the version that corresponds to the current timestamp.

- **Insert**: Upon inserting a new data item into MongoDB, the `commitTmstmp` gets the value of the transaction’s commit timestamp, while the `nextCommitTmstmp` has no value.

- **Update**: When a data item is updated, the MongoMVCC-Driver has to direct MongoDB to firstly update the value of its last version’s `nextCommitTmstmp` to the transaction’s commit timestamp and then additionally to insert a new version of this data item, as in the insert operation.

- **Delete**: When a data item is deleted, MongoMVCC-Driver has to invoke all actions described in the update operation and additionally set the isDeleted flag for the newly created version to true.

From the aforementioned description of MongoDB’s supported CRUD operations, it can be perceived that the update and delete operations are implemented by two atomic operations in the database: an update of the previous version and an insert of the new version. This creates one issue of atomicity and
one of isolation with respect to the ACID semantics that must be ensured by the data store. Proper atomicity guarantees could be ruined if a concurrent transaction is able to access a version where its nextCommitTmp value has already been updated, but the insertion of the new version has failed. To overcome this situation, MongoDB-Ext registers and monitors all pending transactions, forcing an undo-redo action, by requesting the whole write-set from the logger service, in case of a transaction that its commit phase fails.

4.3 Transactional Management

MongoDB’s native release does not ensure ACID transactional management. Likewise, our prototype will not provide ACID transactional semantics either. This property will be ensured by integrating the latter with an external conflict resolution service. MongoMVCC-Driver consolidates this service in order to identify possible conflicts among concurrent transactions. As it is already described, a newly opened transaction is assigned with a start timestamp. During a read operation, every transaction selects data by additionally including the start timestamp, so that MongoDB can scan for the appropriate versions. During the commit stage, the external conflict resolver should assign a new commit timestamp that will characterize the newly inserted version of the data items. For the proper management of transactions, the following decisions have been made.

4.3.1 Snapshot Reads and Private Versions Management

A major requirement from the snapshot isolation criterion is that each transaction handles its private versions so that transactions never interfere with other’s private versions and ensure that each reads its own writes. When a transaction requests to insert/update/delete a data item, the respective operation’s result is automatically added into its private write-set. The write-sets are implicitly handled at client-side, so each transaction has its own collection of write-sets stored in its context. By doing that, only the owner of a transaction’s write-set has access to this data. At this point, nothing is permanently stored in the data store, being all kept in-memory by the application until the commit phase.

When an update/delete operation takes place, a new version of a data item is introduced, which require for the MongoMVCC-Driver to check with the external conflict resolver for potential write-write conflicts. A conflict might occur due to an on-going transaction, initialized by the MongoMVCC-Driver, which has already request for an update on the same data item. MongoMVCC-Driver invokes a ‘hasConflict’ method on its local transaction client, for every data item that tries to perform an update. This method expects the identifier of the corresponding datastore and the key of the data item to be checked for conflicts, in an array of bytes, thus making it totally transparent to the conflict resolver about the content of the key. If a write-write conflict is identified by the external conflict resolver, an exception can be thrown, which will force the transaction to rollback. However, when an insert operation takes place, MongoMVCC-Driver does not have to check for conflicts, since the newly inserted data item can never be accessed by other concurrent transactions. This is so as it is only visible on the scope of the transaction that created this item, thus unique identifiers guarantees that no equal keys will be generated.

Regarding the snapshot reads, MongoMVCC-Driver submits a read query, by enriching it with the transaction’s start timestamp. MongoDB validates the query and returns an iterator for accessing the results. MongoDB stores only publicly visible data items, ensuring that a transaction will never access other transaction’s private set. However, an open transaction might have already updated some documents, and thus, some results of the returned set might be outdated, in the transaction’s scope. Due to that, the return of the results is performed in two steps: The firstly examines the transaction’s private write-set. It uses its own expression validation implementation, to check if a data item stored in the data store validates the query’s predicate conditions. If so, then it is returned back to the application. If not, it is skipped and the iterator continues to the next item. When the iterator finishes investigating the private write-set, it then iterates through the MongoDB’s return result set. For every item included in this set, it checks whether or not its unique key is already contained in the transaction’s private write-set. If not, then it is returned back to the application. If it is already included, it is skipped and the iterator continues with the next item of the result set. This is due to the fact that this item will have already been examined by the cursor at the first stage, which implies that
this item is outdated in the scope of this transaction. This mechanism ensures that each transaction will access only visible items, and only its last updated data.

4.3.2 Indexing support

MongoDB provides by default a B-Tree index on the field that corresponds to the document’s private key. Moreover, it lets the application developer to define additional indexes on other fields, in order to perform performance effective scans. In current state we relied on the MongoDB’s native index mechanism, by using a compound-index on the ‘dataID’ and ‘commitTmstmp’ additional metadata fields. Upon creation of a new collection MongoMVCC-Driver forces MongoDB to create this index by default.

4.3.3 Finalizing Commit

When the application commits an open transaction, it invokes the local transactional client to commit. MongoMVCC-Driver firstly provides the full write-set of the transaction’s private versions, in order to make it durable. The integrated MongoDB will not rely on its native mechanism for recovery in cases of failure. Instead, it will request the write-set of the failed transactions from the logger service. At this point, MongoMVCC-Driver generates an array of bytes for all data items that are stored in the private write-set. Once the write-set has been successfully committed to MongoDB, then the external conflict resolver is notified so that it can make the newly inserted versions publicly visible.

4.3.4 Logging and Recovery mechanism

The design principle regarding the logging and recovery mechanism is implemented as follows: The integrated MongoDB will not rely on its native mechanism for logging and recovery, as this introduces a critical configurable time interval that could lead to data loss and inconsistency. MongoDB submits all its write-set to an intermediate journal file, in order to avoid multiple accesses to the storage, which would in turn introduce an additional performance bottleneck. The journal, which is used as a write-ahead-log (WAL), is being permanently stored in the disk storage at every period of this time interval. This makes the modified data during the last period vulnerable to data loss, in case of a failure. Due to this, the integrated MongoDB will rely on an external logger service mechanism and will request from the latter the write-set to re-apply in case of transaction failures.

5 TRANSACTIONS ACROSS NODES

As it is shown in the overall architecture of our prototype delivery, the management of the transactional context and the implementation of all required functionality that delivers the MVCC paradigm is implemented in the client side of the data store; the mongoDB’s driver to the data store. This is efficient enough for the majority of the applications; however it does not fit the needs of a modern distributed application which is deployed in a cloud ecosystem.

The aforementioned claim can be further enlighten with the following figure:
Figure 2: Current Implementation does not allow transactions spanning across different application nodes

In this case, each application makes use of the extended datastore client (MongoMVCC Client) that ensures transactional semantics by delegating the write-write conflicts to an external scalable transactional manager, and implements the required MVCC functionality by itself. From this figure, it is shown that each datastore client stores internally and in-memory the transactional context of each transaction: the start timestamp, obtained by the external conflict resolver through the LTMClient, and the transaction’s private write-set, thus all modified and uncommitted data of a transaction. MongoDB is used only for storing the committed changes and accessed only through the delivered prototype. This scenario works adequately for the App_1, which is deployed in a single node using a single Java Virtual Machine (JVM). However, App_2 is a distributed application, deployed on multiple nodes running on separate JVMs. Our motivation is to enable App_2’s different nodes to share the same transactional context across different nodes. Our current implementation provides strong constraints for implementing this requirement, as the transactional context is only stored in-memory and lies in each JVM; modified data stored in-memory in the JVM of Node_1 will never be accessible from a read operation invoked by Node_2. Different nodes cannot share transactional context and transactions cannot span across different nodes.

5.1 A Modified and Proposed Architecture

In order to fulfill the aforementioned requirements, we changed our proposed architecture by introducing a novel layer for managing the transactional context and offer transactional semantics to MongoDB. Our modified architecture is shown in the following figure:
In figure 3 our novel layer enables distributed applications to share their transactional context across different nodes, and also allows a multi-node datastore to offer transactional semantics across different shards. All transactional content is now stored in our novel layer, making the datastore client completely stateless, and thus making it possible to share transactions between different nodes. The datastore client offers to the application developer a façade to the underlying novel layer. The application developer can have the impression that uses the datastore as before, with all the functionality operating the application layer. In reality, every invocation of a datastore’s functionality submits a new job to be executed by our novel layer, in a worker-consumer design pattern. The novel layer maintains a local cache of the newly created objects, which are associated with a transaction, so different application nodes can access the same instance of the intermediate layer to operate.

The architecture is designed to be fully scalable. The datastore client that is deployed in the application layer is now as scalable as the application itself is. MongoDB can be deployed in a multi-sharded configuration, and scales up with the same characteristics as it was designed to be, thus maintaining its high performance and scalability requirements. The additional assurance of transactional semantics is delegated to our novel layer, so MongoDB maintains its properties. Finally, the intermediate layer is also designed to be scalable, by using a distributed process coordinator and let invocations concerning the same transactions to be forward to the associate intermediate node that handles the transaction’s context.

6 CONCLUSIONS

In this paper we have presented our proposition for the design of an extended version of the popular MongoDB database which enables it to become ACID compatible by providing transactional
Due to the fact that MongoDB is widely used for cloud applications, mostly for its high performance and scalability combined with its rich query interface, we preserved those characteristics and we extended for providing MVCC functionality. MVCC provision made possible for MongoDB to be further integrated with an external conflict resolver service which can guarantee the isolation property. Moreover our extended interface enables an application developer to bracket multiple statements of a transaction, thus making MongoDB to guarantee the atomicity property. Its integration with a logger service strengthens its durability further.

Although our proposed solution covers a wide area of application requirements for strong consistency and transactional semantics, distributed applications that need to share transactions among different nodes cannot benefit. We additionally proposed an updated architecture with an intermediate layer that can scale independently, which will be a middleware between MongoDB’s server-side and the data store client. This layer will cache its transaction’s content during its lifecycle.

Acknowledgements

This work was funded by project CoherentPaaS: A Coherent and Rich PaaS with a Common Programming Model (FP7-611068).

7 REFERENCES


[14]. 2011 IEEE/IFIP 41st International Conference on (pp. 176-181). IEEE.
[19]. WiredTiger's support for transactions, http://source.wiredtiger.com/2.5.0/transactions.html
IMPLEMENTING FAILURE DETECTION AND REPORTING IN A LIBVIRT CLOUD THROUGH AD-HOC MECHANISMS.

Vasileios Anagnostopoulos, School of Electrical and Computer Engineer, National Technical University of Athens, Greece, vanag@mail.ntua.gr

Nikos Vlastaras, School of Electrical and Computer Engineer, National Technical University of Athens, Greece, nikosvlast@mail.ntua.gr

Dimosthenis Kyriazis, School of Electrical and Computer Engineer, National Technical University of Athens, Greece, dimos@mail.ntua.gr

Theodora Varvarigou, School of Electrical and Computer Engineer, National Technical University of Athens, Greece, dora@mail.ntua.gr

Abstract

Robust fault detection and reporting of hosts in a cloud is a difficult undertaking. Most problems originate from the distributed nature of the system. In this work, we aim to re-use existing approaches and adapt them to fit in an integrated solution. Specifically, we combine consensus over a gossip messaging layer in order to tackle the scalability problem. We present the architecture of the engineered solution along with its benefits. We also describe its implementation in the context of EU FP7 project Orbit.

Keywords: Distributed systems, Fault Tolerance, Cloud Computing.

1. INTRODUCTION

Cloud environments have seen a tremendous growth the last years. The on-demand provisioning of infrastructure associated with zero maintenance costs and cost amortization over a provider’s infrastructure is an attractive cost effective solution for enterprises that enables them to focus more on their core business. Infrastructure is now provided as a service and there is no need for a separate department within an organization in order to provision computer environments. This is especially true in the web app hosting business or in businesses that provide web services. Infrastructure is provided as a Virtual Machine (VM) with prescribed hardware characteristics. These characteristics in reality correspond to hardware resources assigned from a pool, via the virtualization software to the VM in an attempt to provide an equivalent computer environment to a typical real metal setup.

The biggest benefit of migration to a cloud infrastructure is automation. Resource assignment is automated. Moreover one can automatically re-assign resources to the VM on demand. VM provisioning is also automated. This minimizes the time to increase or reduce computational resources according to demand. Monitoring is also automated. However failure of underlying hardware is a reality and happens independently of a real metal setup or in a VM provisioning cloud server. While in the former case only one environment is affected, in the latter case many customers are affected by the failure. Service Level Agreements penalize the time a VM is unavailable. The Infrastructure as a Service (IaaS) providers try to find methods to reduce down time. While typically many server applications are developed in a way to mitigate this problem (ZEN LOAD BALLANCER), (APACHE CASSANDRA), (AMAZON DYNAMO DB) among others, typically an IaaS provider tries to minimize the downtime in order to increase the stability and performance of the application while provide an additional revenue stream. Moreover, competition from rival IaaS providers, makes the need for fast fail over pressing.
2. THE PROBLEM AND ITS CURRENT APPROACHES

Having a VM failure is a typical problem in cloud deployments. The phenomenon is not unique to them since it is present in typical on premises deployments of software applications that utilize network connections. Power supply failures, buffering and message loss can create a node failure or present themselves as node being non reachable. As early as Lamport (1978) the research community came up with an increasing amount of tools to address the problem. One usually takes this into account when creating networked software and designs the software accordingly. Approaches like (Karger et. al. 1997), Lamport (1998) and more recently Ongaro and Ousterhout (2014) increase the sophistication of bypassing the effects of failures. The typical core of these approaches is a log of actions that one has to replicate across backup nodes. In the meantime the scientific advancements have found their way into software projects. Replicated filesystems like GLUSTERFS and LEOFS, robust software computational environments like APACHE HADOOP or more recently APACHE SPARK, fault tolerant document databases like MONGO and RIAK, fault tolerant graph databases like (TITAN GRAPH DATABASE) and fault tolerant configuration systems like APACHE ZOOKEEPER or more recently LOGCABIN among others. Using these components which encapsulate the research the designer of the system can create a fault tolerant application where a number of replicas run in the compute nodes.

The advent of cloud computing has introduced two additional features into the picture. The first is scale. The number of nodes is overwhelming and constantly changing. The other new feature is transparency. Transparency means that a VM can fail but the effects of this failure is not noticeable to the customer. However this is hardly a reality since typically the effects are either delays or spontaneous change of VM state to a previous snapshot. While the user can achieve fault tolerance at his/her layer one can achieve faster fault tolerance at the hypervisor level, at the level that is responsible for the management of VMs. Typical applications are video streaming or content provisioning like document management systems. This kind of fault tolerance is achieved via VM state migration. One has two VMs one acting as primary that provides the service and the other one acts as a secondary that synchronizes its state with the primary. This can be full state snapshots or deltas that are sent to the secondary and applied incrementally. Hybrid approaches are also possible. When the primary fails, the service provisioning is done by the secondary. When the secondary fails, a new secondary is used to keep in sync with the primary. When and how much state is migrated is an area of research and a major problem though not the focus of this work.

A related major problem is the detection of the failure that initiates the fail-over from the primary to the secondary or the initiation of a new secondary. This is the main focus of this paper. Typically one employs an active strategy like a heartbeat in (Stelling et. al. 1999) or a passive one like a ping in Spurgeon (2000) to detect that the VM is down from an agent. However both solutions are far from robust in a distributed setup. Network delays, lost packets or agent failures make the solution unreliable. In order to increase the reliability of the solution, one uses multiple agents through a majority voting procedure. But it solves none of the aforementioned problems. In reality it makes them seem worse. While the essence of the approach is correct as the software fault tolerance experience has shown the approach to increase fault tolerance in this way is more complex.

Infection style gossip protocols like SERFDOM are a possible way to solve the problem at the hypervisor layer. Gossip style protocols are better suited due to scale and agent behaviour while still suffering from the problem of reporting the failure to a centralized entity. In this work we take a hybrid approach combining a classical consensus approach over a gossip protocol to solve the reporting problem. Gossip protocols are an attractive solution since overlay networks are the key abstraction which is particularly the case for cloud networks. Their functionality is agent-like.
Figure 1 Global emergent behavior from local interactions in gossip protocols.

Function is always implemented as a distributed algorithm that communicates over the overlay network. These networks can emerge or can be created over a suitable infrastructure, offering transparency and even encryption when it is necessary. Gossip-like phenomena are commonplace. For example, human gossip, epidemics (virus spreading, etc), computer epidemics (malicious agents: worms, viruses), forest fires and diffusion.

Gossip protocols have extremely simple local algorithmic description which makes them amenable to analysis and understandable. They also offer global emergent behaviour. This also leads many times to robustness. Gossip protocols are used in a variety of applications. In distributed data mining they are used for clustering (Van Resse et. al 2003) and collaborative filtering (Baraglia et. al. 2010) among others. They are also used in wireless sensor networks (Zanaj et. al. 2007), (Ahyoung and Ra 2010). In our case we will use the sensor network behaviour of the protocol over a classical IP network. The main features of the approach remain the same.

3. THE ORBIT PROJECT

Mitigating the effects of downtime requires significant investment with meticulous planning to appropriately address each type of common downtime cause. Moreover, while unrecoverable software faults can either be addressed via application-specific long-term improvements or via generic watchdog solutions which restart unresponsive services, other causes such as unplanned hardware faults or planned maintenance can be addressed by server replication. While server replication provides zero downtime capabilities (i.e. recovery is immediate) - a gap exists in current offerings - as only either expensive hardware-level or application-specific solutions exist.

The EU FP7 ORBIT project ORBIT attempts to address this gap by introducing a new paradigm of virtualized resource consolidation in which memory and I/O resources used by a guest Virtual Machine (VM) are provided by multiple external hosts instead of limited to a single physical server. By combining the features of this novel virtualized resource consolidation paradigm with existing VM Fault Tolerance (FT) active-passive state synchronization solutions, ORBIT is able to provide unprecedented robustness capabilities. ORBIT is thus able to address a wide range of scenarios from single-host FT up to entire-site Metropolitan Area Network (MAN) based Disaster Recovery. Moreover, ORBIT’s novel architecture is a perfect match for cloud-wide deployments, thus
complimenting existing tools available for SMEs and service providers. ORBIT aims to provide an application agnostic fault-tolerance solution for cloud infrastructures that make it possible for the first time to migrate critical enterprise workloads to the cloud without compromising on the availability and performance of the system. ORBIT eliminates the complexity of deploying and managing fault-tolerance solutions at the application level and completely eliminates the effort cloud customers previously invested to deal with unreliable cloud platforms. By enabling cloud platforms with such fault tolerance capability the ORBIT technology accelerates the development and deployment of cloud computing and Internet services.

4. ORBIT FAULT DETECTION ARCHITECTURE

In Orbit we use the OPENSTACK software to maintain an IaaS landscape which relies on the LIBVIRT portability layer. Libvirt is a library that provides a uniform interface across various hypervisor and virtualization technologies. On the project testbench we rely on the libvirt + qemu combination. This requirement poses some constraints in the way our fault detection stack is structured. Specifically we need an interface to the local libvirtd demo to receive events for failure of VMs. Apart from this technicality the problem is the same as the one outlined in section 2. We need to detect host failures across the VM deployment landscape and report them to a central authority reliably.

The approach is based on using various software components to achieve our fault detection capabilities. The next schematic shows our architecture and the communication flow. The storage layer is typically a replicated document database. A notable feature is that after initialization, at normal operation, the Leader Layer takes messages only from the underlying Gossip Layer. There is no direct communication between hosts and members at the Leader Layer. Even members of the Leader Layer, use the Gossip Layer to communicate messages.

![Diagram of ORBIT fault tolerant architecture](image)

Figure 2 The layers of Orbit fault tolerant architecture.

Hosts form a Gossip Layer for gossiping failures, additions or removals. Some of the hosts form Authority Islands (AIs) which are leader election sets of a predefined member number. The main motivation is for scalability reasons. Handling of various events cannot be done in all nodes but they are filtered per AI. AI membership is determined at node initialization either via prior configuration or via an automatic mechanism. The automatic mechanism is out of the scope of this work for simplicity. A storage layer is used for assignment of a predefined number of AIs to hosts. Moreover the storage layer is used for count maintenance of AIs. When members fail or leave, new ones are inserted in order to maintain a pre-defined number of members. Finally the storage layer is used to maintain an epoch count per host in order to avoid the problem of stray notifications due to the asynchronous communication assignment. The communication with storage is done from the host at
the initialization phase before entering the cluster. Moreover the AI assignment is responsibility of the host which is determined through a query to the storage layer. The communication after this phase is accomplished through the gossip layer as it is shown in the next figure. Multiple AIs can report that the same host is down upon failure. This is where the epochs come into play. When a host crashes and comes back into normal operation a stray failure notification (because of eventual not bounded delivery) can be problematic. Epochs are a typical method to suppress past messages that can otherwise cause illegal state transitions.

As we have already noted, the leader of an AI attempts to maintain a constant number of members. While failures happen, the leader contacts the storage layer to find out candidates. This is done through a least loaded discipline. A selection of candidates is communicated through the gossip layer in order to notify the candidates of membership. Having become member of an AI, a host increases his load in the storage layer. An interesting aspect of the architecture is the AI operation. In typical protocols like Raft (Ongaro and Ousterhout 2014) only the leader the tasks arrive only at the leader. In our case, due to the usage of a gossip protocol this is no longer possible. In order to maintain the compatibility with Raft, when a non-leader member has a message, it forwards it to the leader. There can be possible optimizations at this operation but compatibility and code reuse have higher priority for us.

For the Gossip Layer, we used the SERFDOM implementation of a modification of the SWIM protocol described in (Abhinandan et. al. 2002). It is a system written in the language GOLANG by Google which allows one to run an agent in each host which acts as a swarm member. Already the Serf protocol operates for 10.000 node deployments without issues. Consequently it is a scalable solution though further research is necessary. Serf executes external programs in order to convey information from the swarm to external programs. External programs run a client and supply arguments to disseminate information to the cloud. This is particularly suitable to our case since we use only an external program, the AI agent which bi-directionally communicates with the Serf agent in order to provide the Leader Layer.

5. CONCLUDING REMARKS & FUTURE WORK

In this work we demonstrated the fault detection architecture of ORBIT. Using this architecture we develop a robust fault detection system for hosts running a hypervisor to offer IaaS. We used a Gossip Protocol, Serf, as a messaging medium, upon which we build a Leader Layer that attempts to reuse existing consensus research but adapt it to the unique characteristics of the solution. We use a storage layer to maintain the size of AIs constant, to maintain epoch counts of hosts and finally to discover AIs to assign to a new host. We believe that with some minor optimizations we can achieve good
performance and scalability. We tried to re-use existing code in an attempt to not re-invent the wheel. The vast list of distributed components needed much attention since many systems did not work as advertised or were not really fault-tolerant.

An ongoing effort to finalize the implementation is currently underway. We aim to prove the effectiveness of the solution using common metrics. Since there is no single task initiator but multiple we have to make modifications. In the future, optimizations will be introduced to take into account the specific characteristics of the environment. There is a possibility that we can re-engineer the Leader Layer to a new consensus implementation adapted to gossip messaging. Automatic AI membership and AI formation is another issue we have to tackle at a later iteration in order to reduce administrative effort. Finally a proof is necessary to ensure that the design achieves its goal. As practice has shown, it can be a major undertaking in the realm of distributed systems.

ACKNOWLEDGMENTS

This work is funded under European Union’s Seventh Framework Programme (FP7/2007-2013) under: ORBIT– Business Continuity as a Service, 2013-2016, EU grant agreement 609828

REFERENCES

GLUSTERFS, [Online], Available from :http://www.gluster.org/ [Accessed 05/15/2015]
LEOFS, [Online], Available from : http://leo-project.net/leofs/ [Accessed 05/15/2015]
MONGODB, [Online], Available from : http://mongodb.org/ [Accessed 05/15/2015]
LOGCABIN, [Online], Available from : https://github.com/logcabin/logcabin [Accessed 05/15/2015]
APACHE ZOOKEEPER, [Online], Available from : https://zookeeper.apache.org/ [Accessed 05/15/2015]
SERFDOM, [Online], Available from : https://serfdom.io/ [Accessed 05/15/2015]


OPENSTACK, [Online], Available from : https://www.openstack.org/ [Accessed 05/15/2015]
LIBVIRT, [Online], Available from : http://libvirt.org/ [Accessed 05/15/2015]


GOLANG, [Online], Available from : http://golang.org/ [Accessed 05/15/2015]
CREATING A EUROPEAN eHEALTH SPACE FOR CROSS-BORDER ePRESCRIPTION AND PATIENT SUMMARY SERVICES

Dimitrios G. Katehakis, Institute of Computer Science, Foundation for Research and Technology – Hellas, Greece, katehaki@ics.forth.gr
George Pangalos, Informatics and Information Security Laboratory, Aristotle University of Thessaloniki, Greece, pangalos@auth.gr
Andriana Prentza, Department of Digital Systems, University of Piraeus, Greece, aprentza@unipi.gr

Abstract

During the past few years, a lot of work has been done in establishing the necessary tools for providing cross-border Information and Communications Technologies (ICT) solutions for public services in Europe in domains of priority, including eHealth. Several Large Scale Pilot projects (LSPs), aiming towards the delivery of electronic cross-border services, have marked a first step in this direction. They have contributed towards ensuring interoperability, by providing new services and improving overall efficiency and effectiveness in a complex environment, making it easier for national companies to do business abroad and helping citizens when they cross borders, whether for tourism, pension or to work. This paper focuses on ongoing work in the Electronic Simple European Networked Services (e-SENS) Large Scale Pilot (LSP) project in the eHealth domain for improving those services through the integration of existing generic Building Blocks (BBs), outcomes of previous or current LSPs. It also aims to support reliable and secure exchange of medical data in a cross-border setting in order to support evolving interoperability, legal and security requirements. More specifically, the implementation of the cross-border ePrescription and Patient Summary (eP/PS) use cases, in line with European Directive 2011/24/EU on patients’ rights in cross-border healthcare, is examined. The need for consolidating the existing outcomes of non-health specific BBs is also outlined, together with related issues that need to be resolved for improving technical certainty and making it easier to use healthcare services abroad in cases of emergency.

Keywords: eHealth, Cross-border Public Services, Electronic Health Records, Interoperability, ICT Solutions, Patient Summary, ePrescription.

1 INTRODUCTION

Although several public eServices are available at national level, this is not always the case across borders. In order to help their development, a number of LSPs have been developed and run under the the Information and Communication Technologies Policy Support Programme (ICT-PSP) of the Competitiveness and Innovation Framework Programme (CIP) of the European Commission (ICT-PSP, 2015 and CIP, 2015), in five main areas: eID, eProcurement, eBusiness, eHealth and eJustice, to engage public authorities, service providers and research centres across the European Union (EU). LSPs pilot a number of solutions, or BBs, that enable cross-border digital services in those policy areas. Each such block consists of a number of components (common code) and uses a number of standards and specifications. They all also share a key characteristic: they are intended to be taken up as part of online services which make these online services cross-border enabled.

Four such LSPs have been completed so far (Digital Agenda, 2015):

- epSOS (European Patients – Smart Open Services) which was related to the exchange of clinical information, with initial focus on both Patient Summary (PS) and ePrescription/ eDispensation (eP/ eD) solutions (epSOS, 2015).
• PEPPOL (Pan-European Public Procurement Online) which has implemented technology standards for European governmental public electronic procurement (PEPPOL, 2015).
• STORK (Secure identity across borders linked) which has developed a European eID interoperability platform that allows European citizens to log in to public services of other member states using the eID technology of their home country (STORK, 2015).
• SPOCS (Building the next generation Points of Single Contact) which has used the natural person eID solution developed by STORK as well as the Virtual Company Dossier (VCD) concept of PEPPOL for document containers and has generalized it to package company information for transmission to Points of Single Contact (PSC) in other countries (SPOCS, 2015).

Three other LSPs are also still currently running:
• STORK 2.0 that extends the scope of STORK to mandates and representation (e.g. of legal entities) and advances from eGovernment to private sector applications (STORK2.0, 2015).
• e-CODEX (e-Justice Communication via Online Data Exchange) that builds on and makes necessary changes to deliverables from SPOCS and the other pilots to fulfil its objectives for easy and secure access to legal information and procedures in other EU Member States (e-CODEX, 2015).
• e-SENS (Electronic Simple European Networked Services) aiming to consolidate and solidify the work done, to industrialise the solutions and to extend their potential to more and different domains (e-SENS TA, 2013). e-SENS focuses strongly on core BBs such as eID, e-Documents, e-Delivery, semantics and e-Signatures across the different LSP domains. These BBs aim to provide the foundation for the platform of “core services” for the eGovernment, cross-border, digital infrastructure foreseen in Regulation (EU) No 1316/2013 for establishing the Connecting Europe Facility (CEF) (REGULATION No 1316/2013, 2015).

This work puts focus on the ongoing activities for implementing e-SENS BBs in the eHealth domain, to facilitate cross-border eP/PS services, in order to improve efficiency, cost-effectiveness, safety and confidence. The overall legal framework for the eHealth pilot for eP/PS within the project is largely regulated by Directive 2011/24/EU on the application of patients’ rights in cross-border healthcare (DIRECTIVE 2011/24/EU, 2015).

Section 2 (The Use Case Scenario for Cross-border ePrescription and Patient Summary Services) provides a brief description of the eP/PS use case scenarios for cross-border services under consideration. Section 3 (Background Initiatives) describes specific relationship with prior LSPs and other domain initiatives. Section 4 (System Architecture and Use of e-SENS Building Blocks) illustrates how the already developed architecture of epSOS is being adapted to support the incorporation of cross-domain BBs. Section 5 (The e-SENS Pilot Implementation) presents the e-SENS pilot implementation in the eHealth domain and specifically for Greece. Finally, Section 6 (Discussion) concludes by presenting prospective issues and the open potential.

2 THE USE CASE SCENARIO FOR CROSS-BORDER ePRESCRIPTION AND PATIENT SUMMARY SERVICES

The specific use case describes how to support cross-border care for eP/PS in line with Directive 2011/24/EU on patients’ rights in cross-border healthcare (e-SENS D5.4, 2015). Even if the use cases for eP and PS are different, their European background and motivation are similar. In consequence, these two use cases are introduced in a common perspective, although each one has a distinctive process description.

2.1 Process Description for Patient Summary

In the PS use case, the patient is a visitor to the country of care, for example someone on holiday, or attending a business meeting, or one that lives in one country but works in another. The health professional may have some information available from previous encounters, in which case the patient may have a patient record locally stored and possibly also a PS in the country of affiliation. Both sources of information could be consulted and updated by the health professional.
The following ‘actors’ compose the e-SENS use case for PS:

- A patient/citizen who is seeking for healthcare treatment abroad.
- A healthcare professional or provider who is providing healthcare treatment. The healthcare professional is in the need to access remote patient electronic health record using the national infrastructure.
- Two National Contact Points (NCPs) and the epSOS Central Services (CS) for National Contact Point (NCP) configuration and terminology handling.
- Providers of trusted sources including national registries of citizens, patients and health professionals.

Also, the following conditions must be met before the PS use case can start:

- A patient/citizen requesting a healthcare professional for medical assistance abroad (Country B).
- A PS must exist in the patient/citizen’s country of affiliation (Country A).
- The healthcare professional is a person legally authorised in Country B to provide healthcare and is identified and authenticated in Country B.
- A mechanism to validate the identity of the patient at the Point of Care (PoC) has to be available.
- The healthcare professional at Country B knows the identity of Country A.
- A health professional must be related to at least one Healthcare Professional Organization (HPO) or to a health authority.
- The patient/citizen must provide consent (previously given or during the encounter) to the healthcare professional before health data is exchanged.
- There is a chain of trust between system actors in this process.
- The health professional must be able to access the “communication layout” that handles the PS in the European countries.

The use case begins when a healthcare professional in Country B receives a request for healthcare assistance from a patient/citizen from Country A. The flow of events is as follows:

- Patient is identified.
- The health professional requests the validation of the identity of the patient.
- The request is conveyed to the patient’s country of affiliation (Country A).
- Country A provides the (positive or negative) patient’s identification confirmation.
- Country A provides the patient’s identity and consent confirmation to the health professional.
- Once the identity of the patient is validated, the patient consent is verified.
- Once the identity of the patient is validated, the healthcare professional of Country B requests for the PS of Country A.
- If the PS exists, Country A provides the PS of Country A to the health professional.
- The PS of the patient/citizen seeking for healthcare treatment abroad is displayed to the health professional.

If all the above pre-conditions are met, then the healthcare professional of Country B can have access to the PS of the patient in Country A. If the identity of the patient cannot be properly validated in Country A, then Country A informs Country B and subsequently the healthcare professional of the identification failure. If the PS of the patient does not exist or cannot be retrieved from Country A, then Country A informs Country B and subsequently the healthcare professional of the failure.

It is considered essential that related requirements be included in bilateral or multi-lateral agreements between partnering states (MS/ACs) in order to maintain convergence. For real patient’s health data to be exchanged there are also strong binding legal requirements, like for example the ones specified in the epSOS Legal Framework Agreements.

2.2 Process Description for ePrescription/ eDispensation

In the ePrescription (eP) case the patient context is similar to the PS case: The patient is visiting the country of care. If a prescribed medical product is not available abroad, the attending pharmacist may, depending on the circumstances, dispense a different brand or package size of a comparable and suita-
ble product to the patient. In case of a product being dispensed, the eDispensation (eD) document is returned to the country of affiliation, to allow the update of the corresponding ePrescription.

The following ‘actors’ compose the e-SENS use case for eP:
- A patient/citizen who is seeking for having a medical product dispensed abroad.
- A pharmacist who is on duty to dispense the prescribed medical product. The pharmacist is in the need to fetch the remote ePrescription document record using the national infrastructure, and to submit the corresponding eD, once the medicine is dispensed.
- Two NCPs and the epSOS CS for NCP configuration and terminology handling.
- Providers of trusted sources including national registries of citizens, patients and health professionals.

Also, the following conditions must be met before the eP/eD use case can start:
- A patient/citizen requesting a pharmacist for having a medical product dispensed abroad (Country B).
- A valid for dispensation ePrescription document must exist in the patient/citizen’s country of affiliation (Country A).
- The pharmacist is a person legally authorised in Country B to dispense medical product and is identified and authenticated in Country B.
- A mechanism to validate the identity of the patient at the pharmacy has to be available.
- The pharmacist at Country B knows the identity of Country A.
- The patient/citizen must provide consent (previously given or during the encounter) to the healthcare professional before health data is exchanged.
- There is a chain of trust between system actors in this process.
- The pharmacist must be able to access the “communication layer” that handles the ePrescription documents in the European countries.

The use case begins when a pharmacist in Country B receives a request for having prescribed a medical product from a patient/citizen from Country A, who owns an ePrescription. The flow of events is as follows:
- Patient is identified.
- The pharmacist requests the validation of the identity of the patient.
- The request is conveyed to the patient’s country of affiliation (Country A).
- Country A provides the (positive or negative) patient’s identification confirmation.
- Country A provides the patient’s identity and consent confirmation to the pharmacist.
- Once the identity of the patient is validated, the patient consent is verified.
- Once the identity of the patient is validated, the pharmacist of Country B requests for the list of valid ePrescription documents of Country A.
- If the ePrescriptions (ePs) exist, Country A provides the list of ePs of Country A to the pharmacist.
- The pharmacist selects the requested eP, accesses to the eP of the patient/citizen seeking for having the medical product dispensed abroad.
- The pharmacist dispenses the medical product.
- The pharmacist generates eD.
- The eD document is transmitted using the NCP to Country A.

If all the above pre-conditions are met, then the pharmacist of Country B can have access to the eP of the patient in Country A. If the identity of the patient cannot be properly validated in Country A, then Country A informs Country B and subsequently the pharmacist of the identification failure. If the eP of the patient does not exist or cannot be retrieved from Country, then Country A informs Country B and subsequently the pharmacist of the failure.

As in the previous use case, it is considered essential that requirements be included in bilateral or multi-lateral agreements between partnering MS/ACs in order to maintain convergence. Also, as in the PS use case, legal requirements are again crucial to assure the usage of real patient data. There is also a
need to assure compliance with the ePrescription EU guidelines adopted in 2014 (GUIDELINES ON eP, 2015).

3 BACKGROUND INITIATIVES

The following subsections present key background work and initiatives related to the deployment of cross-border eP/PS services to support mobility and facilitate access to care across countries in the EU.

3.1 Smart Open Services for European Patients (epSOS)

The epSOS LSP (2008-2014) concentrated on proving that a reliable and secure exchange of medical data in a cross-border setting is actually possible and feasible. By the end of the project, up to 19 epSOS Participating Nations (PNs) launched their epSOS pilots.

The epSOS architecture has focused on the exchange of medical data in two primary use case settings: PS and eP/eD. Ancillary value added services were implemented and piloted on top of the two primary use cases: Patient Access, Healthcare Encounter Report, and Medication-related Overview. Both, the primary use cases as well as the added-value services primarily feature clinical challenges. However, substantial parts of the non-functional requirements, in particular regarding data protection, confidentiality, and information-security aspects, had to be specifically approached by the eHealth domain, since none of the – at the time available- solutions provided effective answers to the challenges faced by the eHealth domain that could fit for the purpose of cross-border care services.

The fundamental epSOS architecture is depicted in Figure 1.

![Figure 1. Fundamental epSOS architecture.](image)

The central portion of Figure 1 signifies the heart of the epSOS architecture: the NCPs of the countries of affiliation (NCP-A) and care (NCP-B). Those systems have been deployed by each active participating nation and served as the primary contact to each Member State (MS). Their duties and responsibilities are described in the epSOS Legal Sustainability Recommendations. The NCPs are furthermore anchor points for cross-border interoperability as their exposed interfaces are commonly agreed upon by all PNs. The NCPs also serve as trust anchors, brokers (although exclusively trusting their own actors) and bootstrapper (through member state agreements an existing regulation) as well as serving as specific entry/exit points for applicable legislation and jurisdiction. The right-hand side shows the country of affiliation or country A, representing the primary data location of the patient as well as the provision points for the epSOS business services, such as eP/PS. These services and the connected national infrastructure of country A provide the information on patients and can unambiguously identify them.

While fully achieving its primary goal, as well as establishing a new eHealth ecosystem throughout the European MS/ACs, it became apparent that several supporting functionalities and services might be
improved by already existing or currently developed tools and means, that may be used on different domains, provided they can adequately deliver an assertive solution for healthcare use cases and scenarios.

3.2 Secure Identity across Borders Linked (STORK)

The STORK LSP (2009-2012) provided an interoperability framework for eID natural person authentication in online processes (including a limited set of attributes often coming with eID tokens, like name, date of birth, or address). STORK 2.0 (2012-2015) extends by representation and mandates, and an enriched set of attributes through attribute providers. The interoperability framework is based on the Security Assertion Markup Language SAML 2.0 (OASIS SAML, 2015).

STORK is meant to be sector-independent. The high-level STORK and STORK 2.0 process is that authentication is always delegated to the citizens’ infrastructure (either a Pan European Proxy Server – PEPS – component of the MS infrastructure or a Virtual Identity Provider – V-IDP – decentralized software component). The two deployment models “centralized PEPS”/“decentralized V-IDP” are an MS decision. For the Service Provider, authentication requests get routed through Country B components (again V-IDP or PEPS depending on the MS deployment choice). Ad-hoc collaboration between the epSOS and the STORK LSP, called STepS (STORK meets epSOS subproject of STORK), revealed a significant potential of both initiatives complementing each other alongside with a very beneficial side effect of implicitly consolidating the solutions towards common basic infrastructure for shared tasks. The new specifications of STORK 2.0 are now allowing a realistic re-use of concepts and blocks.

STORK can augment existing epSOS patient identification, as the patient’s eID tokens can provide a high level of assurance of patient’s unique identification. Traits can be augmented through STORK and STORK 2.0 attribute provision.

Operating STORK at a PoC is challenging, if for example eID tokens have requirements on the IT environment, like card readers, drivers, etc. Mobile eID can play a major role to overcome that. While STORK itself is agnostic to the actual eID credential, the ubiquitous nature of mobile phones together with its zero-footprint characteristic (not imposing requirements on the computing environment other than e.g. a browser), may allow use at a PoC.

3.3 Shaping the Future of Electronic Identity (FutureID)

The FutureID project (2012-2015) (FutureID, 2015) is an Integration project partially funded under the ICT theme of the Cooperation Programme of the 7th Framework Programme of the European Commission. It builds a comprehensive, flexible, privacy-aware and ubiquitously usable identity management infrastructure for Europe, which integrates existing eID technology and trust infrastructures, emerging federated identity management services and modern credential technologies to provide a user-centric system for the trustworthy and accountable management of identity claims.

The FutureID infrastructure provides significant benefits to all stakeholders involved in the eID value chain. Users will benefit from the availability of a ubiquitously usable open source eID client that is capable of running on arbitrary desktop PCs, tablets and modern smart phones. FutureID allows application and service providers to easily integrate their existing services with the FutureID infrastructure, providing them with the benefits from the strong security offered by eIDs without requiring them to make substantial investments.

This is expected to enable service providers to offer this technology to users as an alternative to username/password based systems, providing them with a choice for a more trustworthy, usable and innovative technology. For existing and emerging trust service providers and card issuers FutureID is expected to provide an integrative framework, which eases using their authentication and signature related products across Europe and beyond. To demonstrate the applicability of the developed technologies and the feasibility of the overall approach FutureID develops two pilot applications and is open for additional application services who want to use the innovative FutureID technology.
3.4 Other initiatives

A number of other initiatives have also been implemented recently. A European e-Health governance structure has been established at the political level (by Article 14 of the Directive 2011/24/EU, through the eHealth Network – eHN, (eHN, 2015)); at the strategic level (via the eHealth Governance Initiative – eHGI (eHGI, 2015) – a follow-up to the Calliope Thematic Network (CALLIOPE, 2015)); and at the operational level through various projects (epSOS, NETC@RDS/ENED, etc. (NETC@ARDS and ENED, 2015)).

The implemented LSPs have already proven that providing cross-border services can be made simpler. In numerous domains, technical BBs have been developed and piloted that enable seamless cross-border services. The underlying technology to support cross-border eP/PS usually relies heavily on CS for publishing and processing cross-border configuration information. The Expanding Health Data Interoperability Services – EXPAND – (EXPAND, 2015) initiative is the guardian of several such epSOS assets as well as assets from other European project that have ended. In that scope, the EXPAND Thematic Network provides governance and support whenever an in progress project aims to fulfil its goals by building on top of those assets. Another major objective of EXPAND is to handover to CEF (CEF, 2015) a set of mature eHealth assets that could be used as baseline for the CEF eHealth Digital Service Infrastructure (DSI). EXPAND also operates as a steering committee for eHealth use case pilots (like PS or eP), assuring the correct alignment with epSOS requirements and recommendations, as well as the foreseen directives for the CEF eHealth DSI.

It is noteworthy that during epSOS lifetime at least two proofs of concept have been implemented: the FETNCP and the OpenNCP (OpenNCP, 2015). All the countries planning to pilot e-SENS have adopted the OpenNCP implementation. e-SENS pilots also consider OpenNCP as the foundation for the pilots’ implementation and operation. The e-SENS pilots also encourage a deep integration of innovative e-SENS BBs as supporting technology of the OpenNCP. They also commit to driving the evolution of the OpenNCP regarding maturity, applicability, and innovation, while no constraints are put in any other NCP implementation.

4 System Architecture and Use of e-SENS Building Blocks

Figure 2 outlines the proposed eHealth infrastructure by e-SENS which extends the existing epSOS based architecture with supplemental components provided by e-SENS, as well as the systems topology by defining the interrelations and orchestration of the supplements. Some of the newly integrated components are highlighted in orange and only feature their most common integration means (such as a web service or user agent).
Figure 2. The e-SENS global eP/ PS use case architecture.

The STORK back-end module is depicted as a component within the national infrastructure of the respective countries. The functionality and provision of eID functionality of the FutureID client is immediately supported by the existing services of STORK. Any STORK-intrinsic interactions between the countries are performed over the existing STORK backbone and therefore they are not depicted within the diagram. The Transformation Manager (TM) deals with the transcoding and translation of information embedded in the clinical documents. The Entity Registries (ER) provide directory services towards the stakeholders, such as identity/property information about a health professional or patient. The ER includes the specific registers, such as the patient or health professional registry as well as the meta directory services that combines and provides the services of multiple local registries for a common data access. The Abstraction Layers are pieces of system integration facilitators that bridge the gap between legacy backend systems and the e-SENS eHealth solution. The environments delimit the regulatory protected realms of the stakeholders with “Trusted HPO” environments benefitting from special legal protection (for instance professional discretion and confiscation protection). The Data Sources (DS) accommodate the clinical data repositories of a country. The FutureID Client is a component designed to operate within the User Agent (such as a Web Browser) at the PoC or the citizen’s IT. Its primary functionality is to extend the capabilities of formerly incompliant IT towards the application of advanced eID, trust, and security functionality directly within the realm of the user.

4.1 Common interaction patterns

The eP/ PS use case is based on two generic interaction patterns. The primary interaction for exchanging a PS is the “Request of Data” pattern of Figure 3, in which a healthcare professional within country B is requesting a singular currently active instance of a clinical document, such as a single PS from country A.
However, whenever a larger/ selective number of documents or instances are requested or different versions of medical data are available that still preserve clinical value, the rather simplistic “Request of Data” interaction pattern of Figure 3 is unable to accommodate this request efficiently and effectively. Therefore, a second interaction pattern is established that enables the healthcare professional to selectively request a subset or collection of medical documents: “Request Overview and Pick Details” (Figure 4). Using this pattern, the healthcare professional in Country B is firstly requesting an overview about all available medical documents about a particular patient and is then able to selectively retrieve the currently relevant. This interaction pattern usually applies more to ePs as those traditionally are provided as multiple atomic clinical documents.

The full context of the operations can be consolidated into the “Application Architecture” Interaction Pattern is depicted in Figure 5.
4.2 Use of e-SENS Building Blocks

It is envisioned that the usage of the cross-domains e-SENS BBs will provide interoperable and stable solutions that will improve existing epSOS services. Such BBs could be the cross-border electronic identification of patients, the reduction of the managerial and administrative operation overhead, the appropriate capture of digital evidences of exchange, the update of the rather dated security safeguards, the cross-border encoding and transporting security context, etc. Each one of these, linked with the injection of an e-SENS Building Block (BB) supporting technology to further facilitate the secure exchange of eP/PS, is discussed briefly in the following sub-sections.

4.2.1 eID for patient identification

Although the original epSOS components responsible with the processing of manual and electronic identification have been designed to be replaced by more appropriate and robust means, a post-alignment of the subsequent healthcare standards, profiles, and interaction patterns is advisable. While this alignment primarily serves the technical domain as well as carrying towards a more robust provision of data protection aspects, some notable side effects with benefits for the piloting nations are anticipated:

- In cases in which the “global” patient identifier of a particular patient is returned immediately or is automatically matched to the national equivalent through the eID means, the epSOS-internal patient identification workflow may be completely skipped.
- The regulatory burden of a positive and correct patient identification and unambiguous linkage of data is currently an organizational burden of the treating healthcare professional who is required to confirm the identity material as presented by the patient as well as the integrity of the link between that material and the patient referenced in the returned medical data.
- The proper provision and application of the highly diverging identification means is a fundamental prerequisite of any successful and meaningful exchange of medical data. However, this operational burden is currently absorbed by the treating healthcare professional, despite their unfamiliarity with the various national means of identification. The eID may relief the healthcare professional of this burden and consequently remove a significant obstacle towards user acceptance.
- In addition to pure delivery of eID, most token carrier and national eID means support advanced security safeguards, including the generation of cryptographic session/transaction keys or pseudonyms. Those may be applied to the healthcare transactions to raise the overall confidentiality as well as putting the data subject in a position to effectively exercise the rights granted under the respective national and pan-European legislation.
The means for establishing a robust patient identification within epSOS is based on several technological and organisational prerequisites originally designed and specified in order to accommodate national specialties, unavailability of suitable technology, and the former absence of pan-European procedures to identify and authenticate patients in a cross-border scenario. The e-SENS eID BB is set to overcome the inefficiency and merely fundamental robustness of the original epSOS process by establishing the means to operate purely electronic identification for not only identifying but ideally authenticating patients for the clinical workflows, while preserving a full compatibility with the existing epSOS technology.

The new e-SENS electronic identification process, as supporting technology for the e-SENS eP/PS use case, is based on the generic interaction pattern for patient identification and authentication. The patient is identified by electronic means against an eID Provider (STORK/FutureID) returning a unique eID for the patient itself. The healthcare professional's software reuses this identifier to: either obtain the sectorial eHealth patient identifier by performing the corresponding Integrating the Healthcare Enterprise (IHE) Cross-Community Patient Discovery (XCPD) transaction (IHE XCPD, 2015), or, immediately applies the obtained eID directly for the medical data request if the obtained patient identifier already qualifies as a sectorial eHealth patient identifier.

Once the patient is univocally identified (through a traditional XCPD workflow) or authenticated (through an e-SENS eID workflow) in the remote country, healthcare professional obtains an epSOS based Treatment Relationship Confirmation Assertion TRC(A) from the NCP-B. Using the Identity Assertion (IdA) and the TRC(A), any epSOS transaction can then be performed, as depicted in Figure 6.

![Figure 6: Generic e-SENS eID flow of events.](image_url)

Medical data must only be disclosed or shared after the patient was identified and authenticated with sufficient accuracy (with respect to Country A demands). Each Country B data consumer as the intended recipient of medical data must identify and authenticate the patient with sufficient accuracy. A
Country B data producer must identify and authenticate the patient with sufficient accuracy before releasing medical information about that patient to a country A data consumer. On successful identification, country A must issue a unique patient identifier that can be used for further transactions on the patient’s medical data. Country A may restrict the usability of this identifier to a certain time span or to a certain requestor.

Patient identification is of highest priority. This is why the EU through the eIDAS Regulation (eIDAS, 2015) aims at boosting the user convenience, trust and confidence, while keeping pace with technological developments, promoting innovation and stimulating competition. Following the formal adoption of the Regulation, related delegated/implementing acts will be developed. This will be accompanied by the necessary policy, standardisation and communication activities at the EU and International levels to ensure understanding and a positive environment for the acceptance and wide uptake of the new legislative framework.

4.2.2 Metadata Locator Service for end point detection

epSOS is using CS for addressing this issue. A capability lookup can provide metadata about the communication partner’s interoperability capabilities on all levels defined in the European Interoperability Framework (EIF) for legal, organizational, process, semantic and technical interoperability levels. The metadata can be used to dynamically set interoperability parameters and ambitions between the sender and receiver.

Central Configuration services although not very highly prioritized, are considered to be background infrastructure and a priority when looking at infrastructure redundancy with view to CEF adoption. Therefore the adoption of the e-SENS capability lookup BB lies within its scope. E-SENS will use the Simple Metadata Publisher (SMP) developed by PEPPOL and generalized and standardized by OASIS (OASIS SMP, 2015).

The sender can retrieve the information necessary for setting up an interoperability process. The Service Metadata Publisher stores the interoperability metadata, which enables routing of documents received from a sender to the correct recipient. SMP service metadata is a combination of information on the end entity recipient (its identifier, supported business documents and processes in which it accepts those documents) and the gateway (metadata which includes technical configuration information on the receiving endpoint, such as the transport protocol and its address). Every community participant is registered in only one SMP registry.

4.2.3 Non-repudiation for patient access to audit trails

Non-repudiation services are also necessary to generate, collect, maintain, make available and validate evidence concerning a claimed event or action in order to resolve disputes about the occurrence or non-occurrence of the event or action. While patient access to audit trail is considered to be of high priority for collecting end-to-end evidence chains, epSOS does not provide non-repudiation information in the infrastructure and non-repudiation of origin and receipt can be manually obtained by (un)signed audit trails.

The epSOS means on Audit Trail and Non-Repudiation have been established with the scope and needs of an LSP in mind. That is:

- avoidance of any immediate implementation burden for the piloting MS/ACs,
- isolation from existing national solutions including non-exposed national infrastructure,
- data source for evaluation purposes, namely the epSOS Automatic data Collector, and
- strong separation between the concerns of the MS involved in the medical exchange (epSOS Country A is protecting the concerns of its assigned patients, epSOS Country B is protecting its health professionals and treatment context).

Non-repudiation aspects in a real-life four-corner model (production system) are not a trivial task. The International Standards Organization and International Electrotechnical Commission (ISO/ IEC) 13888-3 standard (ISO/ IEC, 2015) defines four types of non-repudiation tokens, namely non-repudiation of origin, of receipt, of delivery, and of submission. These tokens (or evidence) are not
used in the same way for all the sectors. In fact, non-repudiation of delivery and submission are defined where delivery agents are used (e.g., store-and-forward message exchange pattern).

Content of non-repudiation tokens is defined to be sector-specific, and not to be defined project-wide. epSOS transactions are defined for NCP to NCP communications only, in a synchronous fashion, thus requiring mandatory non-repudiation of receipt tokens (namely, the audit trails) and optional non-repudiation of origin (digital signatures). The new use case aims at enhancing the epSOS approach with a more formal account of Evidence, thus enabling the epSOS LSP to have European Telecommunications Standards Institute (ETSI) Registered Electronic Mail (REM) evidences (ETSI REM, 2015), guaranteeing the sustainability of the other project's evidence emitter, even re-using the same software. Adding non-repudiation is also expected to improve security.

4.2.4 eSignatures for authenticating arbitrary artefacts

epSOS documents are not digitally signed and therefore more advanced electronic signature facilities are required that exceed the capabilities of the technical systems provided by epSOS. Consequently, the consolidated BB of e-SENS regarding eSignatures that combines functionality of STORK, FutureID as well as the current regulatory reality set forth by eIDAS is improving the original capabilities.

The e-SENS eSignature BB can provide: (i) assertion and authenticated attribute signatures, (ii) time stamp signatures for non-repudiation, and (iii) optional document signatures as currently assumed required by some piloting nations.

Not signed artefacts are considered to be of mixed prioritization. It has therefore been suggested to remain out of scope from the initial e-SENS piloting plans, since it requires use case extension and has an IHE dependency (since this BB behaviour is not yet accepted by IHE).

4.2.5 Trust Establishment for end-to-end security and security relaxation

Trust establishment is a key task, both during bootstrapping and operational stages. In epSOS trust establishment is implemented by using Trust-service Status Lists (TSLs) and NCP-service Status Lists (NSLs) containing remote certificates chains used to validate security means (e.g., validating SAML assertions, mutual authentication on TLS channels). During the epSOS operations, the Security Expert Group (SEG) had to aprove some ‘relaxation’ and amendments to the original epSOS security specifications, mainly due to the impossibility to find suitable certification authorities able to issue the required certificates.

The e-SENS Trust Services Solution Architectural Template (SAT) aims at providing a specification for cross-border and cross-sector trust establishment and certificate layouts following strictly the eIDAS regulation. Once it is finalized, its findings will enable the eHealth domain to overcome the abovementioned relaxations and align to the eIDAS. End-2-end security and security relaxation is considered to be of mixed-to-low prioritization.

5 THE E-SENS PILOT IMPLEMENTATION

5.1 The eHealth pilot implementation plan

The e-SENS eHealth pilot implementation will support cross border care for Patient Summaries and ePrescriptions, in line with Directive 2011/24/ EU on patients’ rights in cross-border healthcare. The first set of e-SENS pilots is already under implementation and cross-border piloting is expected to become operational in June 2015.

In the PS case, the patient is a visitor to the country of treatment, for example someone on holiday, one attending a business meeting, or one that lives in one country but works in another. The health professional may have some information available from previous encounters, in which case the patient may have a patient record locally stored and possibly also a PS in the country of origin. Both sources of
information could be consulted and updated by the health professional. In the eP case the patient context is similar to the PS case, e.g., the patient is visiting the country of treatment. If a prescribed medical product is not available abroad, the attending pharmacist may, depending on the circumstances, dispense a different brand or package size of a comparable product to the patient. In case of a product being dispensed, the eD document is returned to the country of affiliation, to allow the update of the corresponding ePrescription.

Even though most of the MS piloting in e-SENS have already piloted PS and eP services during epSOS, the process described assumes that the plan is agnostic about previous experiences in piloting these services, in order to allow new MS/ACs to come also on board.

Figure 7: e-SENS eP/ PS Summary Pilot Plan.

The e-SENS pilot plan is organized in four distinct phases, each of one having a clear purpose and outcome. An overview of the e-SENS eP/ PS Summary Pilot Plan is shown in Figure 7.

(i) Phase 1: Baselining. Its purpose is to provide a clear picture of the pilot implementation plan in order to assure the unbiased comprehension by all involved, the effort it demands from each single stakeholder, understand the changes and their impact on assets, as well as to understand the de-
dependencies and risks. Its outcome is an agreement between all stakeholders in providing in time all
the needed means foreseen in the Pilot Plan.

(ii) Phase 2: Restart Piloting. Its purpose is to re-establish the baseline conditions (e.g. necessary CS)
for MS to Pilot the PS and eP use cases and to enhance the NCP reference implementation accord-
ing to the e-SENS requirements. Its objective is to restart PS and eP pilots in the new e-SENS en-
vironment and specifications.

(iii) Phase 3: CS refactoring. Its purpose is to refactor the CS architecture and operation paradigm to
the e-SENS specifications, including Trust Services. Its objective is to deploy the new architecture
and operation paradigm for configuration services, based on specifications for cross-border and
cross-sector trust establishment and certificate layouts according to the eIDAS regulation.

(iv) Phase 4: Patient eID. Its purpose is to implement an enhanced Patient identification scheme based
on Electronic tokens and improve the liability and the user friendliness of the current (manual)
process. The objective is to release for MS adoption, a new version of the NCP reference imple-
mentation that combines two methods of patient identification: a manual one (as it was on
epSOS), and an electronic one (according to the e-SENS eID BB).

A revised epSOS testing strategy will also be applied to e-SENS eHealth eP/ PS pilots before they go
live (and it covers legal, security, semantic, organisational and technical aspects).

5.2 The national pilot plan for Greece

The piloting in e-SENS is the next step for Greece to foster European wide cross-border eHealth ser-
vices and a logical next step to the epSOS pilot services. The Greek pilot is included in the first set of
pilots to become operational in June 2015. The first use-case to be implemented will be cross-border
eP.

Greece is a country with a high influx of tourists throughout the year. The opportunity to dispense
electronic prescriptions and access patient summaries from other countries in a Greek pharmacy and
health care facilities respectively is a great advantage. The priorities for piloting in e-SENS in Greece
are determined at the political level by the priorities of the state and the readiness of the national pilot
partners to support such priorities. Within e-SENS, Greece will implement the eP/ PS use cases.

Greece has already piloted in the framework of epSOS ePrescription as country of treatment for the
patients (country B). National cross border initiatives are focusing on expanding current services to
services as country of affiliation for eP and also initially as country B for PS. In anticipation of the
latter, the epSOS national implementation team has already implemented and tested the epSOS MTC
which is necessary for the semantic transformation of the PS. However, the needed legal and organiza-
tional framework for electronic health records, currently in process of development, will need to be
secured before Greece can expand into the eP/ PS use case beyond pre-production. It is also under-
stood that the e-SENS eHealth pilot will take place initially with test data only. E-SENS extensions to
be piloted need to be able to follow existing current situation in Greece, especially in the eID domain
where end to end security via smart card technology for example is not supported. As such a STORK
based eID approach seems to be more in-line with future developments.

Greece has implemented the epSOS open NCP and will maintain the NCP with any further extensions
whether delivered in e-SENS or in other projects (such as EXPAND). It is anticipated that the current-
ly expressed political commitment will also result in sustainable operation of the NCP under the legal
agreements to be established within the Subgroup. The provision of the current cross border pilot ser-
vices and the future extensions will take place within the framework of existing European regulations.
Both eP and electronic patient records are regulated by national legislation.

Greece is expected be able to go live at pre-production with simulated data by the end of 2015. It is
also expected to be ready to go operational immediately after all domain and national pre-conditions
are met. It is desirable that action with actual users is taken in advance of deployment, possibly within
CEF.
Once the solution has been tested and validated, it will run in pre-production environment. This phase includes the following actions:

- Installation of pre-production testing environment
- VPN Connections
- Certificate Management
- Training of pilot participants
- Management and monitoring of the running/operation phase of the pilot
- Pilot environment maintenance – improvement of pilot implementation
- Helpdesk support (1st level)
- Assessment and evaluation of the pilot at national level

The national PS service was foreseen to be launched in 2015; it is therefore likely that Greece may participate in e-SENS with a full PS-country A, B service. It is however desirable that action with actual users is taken in advance of deployment, possibly within CEF. Greece is both a highly touristic destination and has also a highly digitized health sector. Embedding e-SENS/epSOS functionalities into the local apps is likely to increase doctors’ buy-in and active collaboration.

6 Discussion

This paper has focused on the ongoing work in the e-SENS project for improving ICT based cross-border solutions for public services in Europe in domains of priority, including eHealth. This is implemented mainly through the integration of existing generic BBs that support reliable and secure exchange of medical data in cross-border settings. An evolving architecture, to accommodate cross-domain BBs within the continuously evolving European eHealth space and to support cross-border services for eP/PS has been presented.

It is anticipated that the eHealth domain will greatly benefit from mitigating non-domain concerns, such as eID, trust anchors, trust bootstrapping, crypto-management, and baseline infrastructure security towards other domains that are authoritatively responsible of providing those exact measures. Even so, deep assessment is necessary in order to deeply understand the legal, organizational, semantic and technical interoperability framework that has been established in the last fifteen years in Europe. Such assessments may provide evidence that, what in principle are non-domain concerns (e.g. baseline infrastructure), may in fact be tightly tied to domain requirements or pre-conditions (e.g. metadata profiling).

A core foundation and a set of basic principles are also necessary to realise interoperability between MS/ACs. Although a lot of work has been done in this respect, there is still a need for consolidating the existing BBs and testing technical and legal issues. The European Interoperability Framework (EIF) (EIF, 2015) and the current LSPs already marked a first step in this direction. The next step is now required to unlock the potential of cross-border services and define the standards to enable cross-border services.

Specifications that are being developed in e-SENS are expected to contribute to the implementation of the European Interoperability Framework (EIF) for basic cross-border public services in Europe. In this way, e-SENS prepares the ground for the future Digital Service Infrastructures under the Connecting Europe Framework program (CEF).

Acknowledgements

Work reported in this paper is partially supported by the European Commission and specifically by the Information and Communication Technologies Policy Support Programme (ICT-PSP) of the Competi-
tiveness and Innovation Framework Programme (CIP) under the e-SENS project (“Electronic Simple European Networked Services”, contract number 325211, http://www.esens.eu). The authors would like to explicitly acknowledge significant contributions from Sören Bittins (FOKUS, Germany), Massimiliano Masi (FMH, Austria), Marcello Melgara (LISPA, Italy), Licínio Kustra Mano (SPMS, Portugal), and Francois Wisniewski (LIST, Luxembourg). Any opinions, results, conclusions, and recommendations expressed in this material are those of the authors and do not necessarily reflect the views of e-SENS or the European Commission.

References


e-SENS D5.4. e-SENS Deliverable D5.4: Second-wave Update of Plans and Status of Domain and National Pilots

NETC@RDS. A step towards the electronic EHIC. [cited 24 April 2015]. http://www.netcards-project.com/
OpenNCP Community Home. [cited 24 April 2015]. https://openncp.atlassian.net/wiki/display/ncp/OpenNCP+Community+Home
ELECTRONIC CROSS-BORDER ACCESS TO LEGAL MEANS AND PROCEDURES IN EUROPE - THE GREEK eCODEX PILOT

George Pangalos, Informatics Laboratory, Faculty of Technology, Aristotle University of Thessaloniki, 54006, Greece, pangalos@auth.gr

Ioannis Salmatzidis, Director, IT Center of Aristotle University of Thessaloniki, jsal@it.auth.gr

Ioannis Pagkalos, Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, ipagkalo@auth.gr

Abstract

European integration has resulted in a fast increasing number of legal procedures containing cross-border effects which require better cooperation between the national judicial systems and more efficient and faster procedures for the users. ICT technologies can help make cross-border judicial procedures more transparent, efficient and economic, both in civil and criminal matters. e-CODEX is an important project of the EU in the domain of e-Justice that aims to provide to citizens, enterprises and legal professionals an easier access to justice in cross border procedures, as well as make cross-border collaboration of courts and authorities easier and more efficient. It develops the required infrastructure and the organizational, procedural and legal environment necessary and also conducts a number of real life cross-border pilots. One of the first such pilots to become operational is that of the European Payment Order (EPO), in which Greece also participates. In this paper we briefly present the services of cross-border access of citizens and legal professionals to legal means in Europe provided by e-CODEX and also the Greek e-CODEX pilot of European Payment Order, which has already launched operation in the framework of this major e-Justice project.

Keywords: eJustice, crossborder IT services, automated legal procedures, elaw

1. NEED AND OBJECTIVES

The European integration and the resulting high mobility of European citizens, procedures and businesses, has resulted in a steadily increasing number of legal procedures containing cross-border effects. These procedures require better and faster cooperation between the different national judicial systems involved. Different forms of communication to cope with the continuous growth in data exchange, beyond the traditional exclusively manual ones, are also necessary. ICT (Information and Communication Technology) can help make judicial procedures more transparent, efficient and economic. It can also help citizens, companies, administrations and legal practitioners get facilitated access to justice. This results in both smoother access to information and the ability to process cross-border cases more efficiently.

Several initiatives have already been undertaken by the European Union to facilitate the above. One of the most important ones is the e-Justice initiative [1]. e-Justice aims to improve access to justice and to facilitate cross border judicial proceedings through the use of information and
communication technology and EU-wide interoperability. It also targets the information deficit and language barriers. Its potential audience includes citizens, businesses, legal practitioners and the judiciary. It also strives to develop the European e-Justice Portal, a one-stop (electronic) shop for justice information in the EU [1].

e-Codex (e-Justice Communication via Online Data Exchange) is one of the most important Large Scale Projects of the EU in the domain of e-Justice that aims to provide to European citizens, enterprises and legal professionals an easier access to justice in cross-border procedures and to make cross-border collaboration of courts and authorities easier and more efficient by creating interoperability between the existing national ICT solutions [2].

The main objective of e-CODEX is to enable access to justice systems across Europe and provide an easier (digital) way to execute cross-border procedures and exchange legal information between EU-countries, replacing bureaucratic paperwork, no matter the differences between the EU countries. More specifically, e-CODEX aims to provide an easy and secure access to legal information and procedures in other EU Member States for businesses and citizens, greater cross-border effectiveness of legal processes through common standards and greater interoperability of information systems. It also aims to improve efficiency of cross-border judicial processes through standards and solutions that ease and facilitate the cross-border case-handling activities [2].

The e-CODEX project started in 2012, it is funded through the ICT Policy Support Program of the EU and has a duration of 50 months. Twenty five EU and EU-associated countries participate in the project (Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Estonia, France, Germany, Greece, Hungary, Italy, Ireland, Jersey, Lithuania, Malta, Netherlands, Norway, Poland, Portugal, Romania, Spain, Sweden, Turkey, United Kingdom, CCBE and CNUE). Greece participates in e-CODEX through the Greek Ministry of Justice, which has mandated the overall responsibility for the national participation and implementation to the Informatics Laboratory of the Aristotle University of Thessaloniki.

Beyond the development of the required infrastructure and the organizational, procedural and legal environment necessary for providing the e-CODEX services (organized around seven work packages), e-CODEX also runs a number of real life pilots (use cases). One of the first such pilots to become operational is that of the European Payment Order (EPO), in which Greece also participates. In the following chapters we will briefly present e-CODEX objectives and structure (mostly based on [2]) and the methodology used and experience gained from the implementation of the first Greek EPO pilot.

2. TECHNICAL APPROACH AND IMPLEMENTATION STRUCTURE

As noted already, as the European Union evolves, it requires different forms of communication to cope with the continuous growth in data exchange caused by increasing commercial activities. Exclusively manual processing does not provide the responsiveness that a modern society requires. There is therefore an urgent need to automate legal procedures and to make cross-border judicial procedures more transparent, efficient and economic both in civil and criminal matters. Information and Communication Technologies can significantly help in this direction since they allow the use of e-Services and the interconnection of the national infrastructure that Member States have already established. e-CODEX aims to automate legal procedures without re-inventing the wheel, thus making the system more mature.

In this context, electronic transport of data and documents is a key piece of the solution. Any functionality developed for a cross-border e-Justice service will necessarily mean transport of
information from one country to another, including communication between the e-Justice Portal and some national solution. Because there is a focus on security and availability for the cross-border e-Justice service, e-CODEX coordinates and establishes an appropriate, efficient and secure e-Delivery solution (figure 1).

![Figure 1: The e-CODEX e-Delivery solution and inter-connection approach [2]](image)

The e-CODEX e-Delivery solution and inter-connection necessitates that interoperability in several technical and semantic aspects is guaranteed. Thus, e-CODEX cross-border e-Justice services are based on a decentralized approach consisting of (i) an e-CODEX e-Delivery platform, (II) an e-CODEX Gateway, (iii) an e-CODEX National Connector, and (iv) a National System (service provider) (figure 2). This approach makes it easier to integrate existing national solutions into a new cross-border e-Justice service [2].

The National Connector is responsible for all semantic local mapping and guarantees the ability of the national systems to communicate with the e-CODEX gateway. It is usually linked to a National System which is, in turn, used by the courts. The e-CODEX Gateway establishes a secure and standardized connection with any other Gateway on either the national or portal side (figure 2).

![Figure 2: The overall information flow supported by e-CODEX [2]](image)

Communication flows from the National System, to the National Connector, to the National e-CODEX Gateway and then, respectively, to the foreign country’s e-CODEX Gateway, their National Connector and their National System. The e-Justice Portal of the EU is the only e-CODEX component that communicates directly with the gateway without a connector (since there is no necessity to transform documents to a national standard). Depending on the service to be supported, the bi-directional communication could be from the e-Justice portal to the
courts, from court to court, from court to the secure mailbox in the e-Justice portal, etc. The overall information flow supported by e-CODEX is also depicted in figure 2.

3. THE E-CODEX BUILDING BLOCKS

A basic characteristic of e-CODEX is that the project develops common building blocks that can be used in, or, between Member States to support cross-border operation of processes in the justice field. Such solutions have been developed in different areas, ranging from safe transportation to identity and document standards, and are used in several different e-CODEX pilots. The developed solutions also enable a safe environment for all kinds of different user-groups to access a wide range of legal services across Europe. More specifically, e-CODEX is currently using the following main building blocks (components) [2]:

a. e-Delivery: The e-Delivery building block is at the center of the e-CODEX architecture and is responsible for securely transporting information between member states. The e-Delivery system includes for every participating country the gateway and a basic framework for the national connector. The connector is then customized by each participating country to fit its specific needs. The e-CODEX e-delivery / e-transport building block is a reusable connection solution based on the ebMS 3.0 standard (based on the Holodeck b2b messaging software) and the ETSI REM, and ISO, OASIS and ETSI standards. It is essentially a content agnostic, plug and play cross-border connection solution, that could also be used as the basis of other inter-European and world-wide projects.

b. e-Signature: The e-Signature building block is part of the national e-CODEX connector and helps to sign documents and generate the so-called ‘trust-ok’ token. It also checks the validity of incoming signed documents and, thus, helps keep the cross-border transmission of documents secure. It also provides connection with the national e-identity frameworks (existing national solutions). The implementation and verification of e-Signatures is based on the so-called DSS-Tool [2].

c. e-Identification: The e-Id building block makes it possible for EU citizens and legal professionals to access e-CODEX services and the e-Justice portal by authenticating themselves through the use of their national identities. A role authentication system for lawyers called “Find a lawyer 2” is also being considered to be connected to e-CODEX.

d. e-Document: The e-Document building block deals with document conversion and semantics. Documents need to be converted to conform to national standards. The document conversion is done through the use of XML schemas and mappings. The XML data accompanying the documents are automatically converted to fit the national case management system. This conversion again takes place on the way back to transform data to the EU standard. This makes it possible for national systems to stay independent and still participate in the cross-border exchange of data. All of this takes place in the national connector.

4. THE E-CODEX PILOTS

Beyond the development of the necessary infrastructure and the organizational, procedural and legal environment necessary for providing the e-CODEX services, e-CODEX also runs a number of real life pilots (use cases). During the piloting phase, which started in 2013, real life scenarios are being tested. Upon completion of the pilot phase, an evaluation will also be done and further adaptations will be undertaken. Pilots will be easily adoptable by countries wishing to join the use case at a later stage, provided that they fulfill the necessary technical and legal requirements. These pilots also respond to some of the weaknesses that have been identified in regards to current practices, by providing a possibility to reduce delays, in the interest of both the judicial authorities
and the requested person, and ease the collection of statistical data.

Five major legal procedures have been identified by e-CODEX so far as use cases for piloting:

- The European Payment Order (EPO), based on Regulation (EC) No 1896/2006 of the European Parliament;
- The Small Claims (SC), based on Regulation (EC) No 861/2007 of the European Parliament;
- The European Arrest Warrant (EAW), based on (EC) Council Framework Decision 2002/584/JHA of 13 June 2002;

5. THE EUROPEAN ORDER FOR PAYMENT PILOT

The swift and efficient cross-border recovery of outstanding debts is of prime importance not only for EU citizens but also for companies, as late payments often constitute a major reason for insolvency, threatening the survival of many small, medium-sized and even large businesses and resulting in numerous job losses. EU has taken the initiative to simplify and speed up the recovery of uncontested monetary claims in cross-border cases by creating a harmonized European order for payment procedure (EPO) [10]. EPO procedures are applicable in several cases. For example, sales contracts, rental agreements, contracts of service (related to transport, hotels, restaurants, etc), subscription agreements (newspapers, magazines, etc), insurance contracts, out-of-court settlements, membership fees, e.t.c.. It must be noted that many such small claims don’t reach courts, since the expected financial outcome doesn’t worth the effort (accounting for 63% of all such cases EU-wide today) [1,2].

Cross-border communication in this area used to be so far mainly paper-based. The EPO e-CODEX pilot implements the necessary technical infrastructure and interfaces for secure electronic cross-border submission of business documents, based on the European order for payment procedure. The e-CODEX pilot enables companies, institutions and legal professionals (e.g. lawyers), to electronically file EPO cases to the competent court in another piloting Member State, by connecting to the respective national filing systems via e-CODEX.

EU citizens are therefore supported by a new functionality, through which they are able to fill directly from their desk the application-form for a European order for payment and submit this application and the accompanying documents directly in electronic format to the competent court in any other Member State participating in the pilot. All of the above result in speedier access to efficient justice in cross-border money claims. Courts also become more efficient in handling more cases with less burdensome paper effort.

The e-CODEX implementation of the EPO cross-border pilot is based on the EPO workflow, as described in current EC regulations (1896/2006 and 936/2012) [10]. According to the EU regulations, a claimant fills in the claim, signs it with his electronic signature and sends it to the competent court. The court considers the applicability of claim and decides to issue an EPO, or rejects it, or requests further information, or assesses that only a part of the claim meets requirements. The defendant then either accepts or contests this decision. Finally, if there is no reaction by the defendant, the court declares EPO enforceable and processes it.

The main stakeholders of the e-CODEX EPO pilot are the legal professionals and companies
which need to submit EPO claims using electronic interfaces, the EU citizens which are enhanced to support electronic communication via e-CODEX, and the Courts which are connected via their national electronic filing systems and their national back-office applications for court case management. Other stakeholders of the e-CODEX EPO pilot include the Justice Ministries of the participating Member States, which are responsible for the national filing systems and the national court case management systems, as well as the EU Commission, which is responsible for running the European e-Justice Portal.

Some EU Member States already allow the electronic filing of EPO cases, especially for key customers of justice that produce the main case load for courts in civil proceedings (e.g. lawyers, banks, insurance companies and social security institutions). For example the national filing system "EGVP" (Elektronisches Gerichts und Verwaltungspostfach) in Germany, and the "ERV" (Elektronischer Rechtsverkehr) system in Austria. However, currently these national filing systems can only reach participants in their own Member State, while the e-CODEX EPO pilot aims to provide interfaces for cross-border communication between EU member states.

6. THE GREEK EUROPEAN PAYMENT ORDER PILOT

Each country intending to participate in the different use cases selected for the e-CODEX piloting phase has to meet a number of organizational and technical requirements. For each one of the use cases a country participates in, a corresponding national case management system has to be available (or be implemented). The participants have also to make sure that their national technical, organizational and legal framework provisions allow for the respective electronic submission and processing of documents to a court.

The key players for the implementation of the first e-CODEX EPO pilot in Greece have been the Greek Ministry of Justice (overall political responsibility and guidance), the Aristotle University of Thessaloniki (national coordinator and technology provider, responsible for setting up the e-CODEX Gateway and the National Connector and for implementing the national EPO Case Management System - CMS), and the Athens Court of First Instance (CMS host and also responsible for involving stakeholders during the piloting stage - mainly lawyers of the Greek Bar Associations). Just prior to the EPO pilot launch (beginning of 2014), the electronic submission of documents to Greek courts was also made available to the lawyers of Attica.

Significant time and effort was spent during the design and implementation phase of the Greek EPO pilot on discussing with potential users, mainly lawyers of the Greek Bar Associations. Also on assessing the current conventional practices of the paper-based procedure and on organizational issues regarding the potential and limitations for the application of EPO procedure in Greece and their exploitation by local lawyers. Those include:

- The detailed organizational analysis of the Greek Payment Order procedure
- The detailed study of the e-Codex requirements regarding e-Signatures and e-ID and the assessment of their compatibility with the Greek eID government portal "Ermis" [3]
- The analysis of the European Payment Order and its actual implementation at national level by the Greek courts and lawyers
- The organization of the involvement of stakeholders: the Athens Court of First Instance (Judges and Court Administration), the Athens Bar Association and more recently the Thessaloniki Bar Association
- The setting up of the necessary technical transporting infrastructure (Gateway, National Connector, CMS, etc.)
- The development and Implementation of a suitable security policy and the necessary procedures foreseen by the ‘Circle of Trust’ agreement signed among all EU partners
participating in the pilot
. The development of the national EPO Case Management System
. The promotion of dissemination activities for local stakeholders, etc.

The involvement of stakeholders (mainly lawyers, court clerks and Judges) from the early stages of the exercise has been crucial in order to gather information on organizational issues regarding the actual courts’ application of the European procedure and the actual use of the procedure by lawyers, their attitude against the new procedure and the issues that they may encounter. The Greek – specific problems encountered include the scarce use of the EPO procedure by Greek lawyers, the relatively scarce number of EPO claims that courts receive, the difficulties for lawyers in identifying the competent courts in another country and the payment of fees. Issues and difficulties that other future piloting countries may also encounter.

The Greek EPO pilot (figure 4) is already operational. Greek Lawyers can today submit and process easily, from their offices, using a fully automated, fast and easy electronic procedure, EPO cases directly to competent European courts (figure 5). Lawyers from other piloting European countries can also do the same. For the time being, cases from abroad can only be submitted to the Athens Court of First Instance (which however covers approximately 60% of all cases at national level). A detailed users’ manual (handbook) for the Greek e-CODEX EPO pilot and an instructions manual for the Greek users are also available (figure 6) [8,9]
To deal with a manageable technological and organizational complexity is key to the success of piloting. For this purpose, Greece decided to follow a gradual piloting approach. Initially piloting is restricted to lawyers. It will be open to citizens with digital signatures at a later stage. This has nothing to do with technological limitations, but rather with the speculation that the electronic procedure may be too complicated for the average user at this stage, especially when it comes to understanding the legal terms and procedures, the retrieval of the competent court and the payment of fees.

Finally, in order to ensure that a recently installed system (e-CODEX gateway and connector) works correctly, a peer with an already working system must send a variable number of test messages. This can prove difficult to set up, since it is difficult to predict exactly when these messages are needed. For this purpose, we are also currently setting up an automated testing tool that can be easily adapted to send test messages to future piloting countries. The testing platform may acquire all relevant roles (sender and receiver) and support not only EPO XML schemas, but also schemas regarding the rest of the use cases, such as small claims, criminal law use cases, etc.
The implementation and use of the Greek EPO pilot so far has shown that its introduction has several advantages. It helps to speed up and makes easier the processing of EPO cases that include parties domiciled in different member states by removing physical distances and transforming them in an electronic system. It is also important to note that no special expertise or infrastructure is required for installing and using the system (only a computer, an internet connection and a digital signature is needed). The new system also supports a rapid debt recovery, by reducing delays often occurring in today’s civil procedures, and thus helps build a better economic environment all over Europe. It also provides a direct and secure communication with courts and also provides for the acknowledgement of exchanged forms through proofs of receipt. It also helps reduce costs of cross border communication by running the entire process electronically (instead of registered mail). Finally it helps eliminate language barriers since filling of documents is made in own language (court accepted languages are only used for some small parts of the form). It should finally be noted that the Greek lawyers that have used the system so far are very positive on its usefulness and simplicity of use.

7. CONCLUSIONS

The fast increasing number of legal procedures containing cross-border effects in Europe requires better cooperation between different national judicial systems and simpler and more efficient procedures for the users (citizens, lawyers, companies, etc). Information and Communication technologies can help make those procedures more transparent, efficient and economic, in civil and criminal matters alike. e-CODEX aims to provide to citizens, enterprises and legal professionals an easier access to justice in cross border procedures and to make cross border collaboration of courts and authorities easier and more efficient. The project has already developed the required infrastructure and the organizational, procedural and legal environment necessary. It also runs a number of real life cross-border pilots. One of the first ones to become operational is the European Payment Order (EPO) one. In this paper we briefly presented the services of cross-border access of citizens and
legal professionals to legal means in Europe provided by e-CODEX as well as the Greek e-CODEX pilot of European Payment Order, which has already started operating in the framework of this major project. Today, through this pilot, Greek lawyers can easily submit and process EPO cases directly to the competent European courts from their offices, using a fully automated, simple and fast procedure. Lawyers from other piloting European countries can also do the same. All they need is a computer, an Internet connection and a digital signature. It was, therefore, shown that modern Information and Communication Technologies can help improve efficiency of cross-border judicial processes, through solutions that ease and facilitate the cross-border case-handling activities.

8. REFERENCES

2. The e-CODEX website, http://www.e-codex.eu
8. The users handbook for the e-CODEX EPO pilot, https://www.jol.nrw.de/bscw/bscw.cgi


Management and Organizational Issues in Information Systems

Chairs: Dr. Ahmad Ghoneim and Dr. Euripidis N. Loukis
Table of Contents

**Track – Management and Organisational Issues in Information Systems**

4. [PID-71] Ozgun Imre, “LEARNING BY NEGOTIATION – IMPLEMENTING A JOURNAL MANAGEMENT SYSTEM”
5. [PID-103] Hamed Sarbazhosseini, “ORGANISATIONAL ISSUES INVOLVED IN PROJECT PORTFOLIO MANAGEMENT IN TERMS OF ORGANISATIONAL STATE-TRANSITION APPROACH”
TOWARDS A UNIFIED META-MODEL FOR GOAL ORIENTED MODELLING

Amjad Fayoumi, The University of Nottingham, Nottingham, UK, amjad.fayoumi@nottingham.ac.uk
Evangelia Kavakli, University of the Aegean, Mytilene, GREECE, kavakli@aegean.gr
Pericles Loucopoulos, The University of Manchester, Manchester, UK, pericles.loucopoulos@manchester.ac.uk

Abstract

Goal oriented modelling (GOM) is one of the most prominent and widely accepted techniques in information systems research. Since the early 1990’s, a large number of GOM approaches have been proposed aiming to a better alignment between business strategy and the behaviour of supporting systems. Different GOM approaches focus on different activities in the early stages of system development and propose a variety of strategies for reasoning about goals. A number of researchers have stressed the advantages of integrating different GOM techniques, especially in the context of modern global business environments. This is evidenced in the increasing number of publications in this area. However as each GOM language (even versions of the same language) comes with its own syntactic and semantic singularities, such integration requires a number of complicated transformations which is a major obstacle to model and tool interoperability, and prevent wider adoption by practitioners. In order to provide a unified view of GOM, one needs a common understanding of GOM concepts, their semantics and deployment. To this end, this paper proposes a language independent meta-model based on the analysis of eight GOM languages. Generic concepts were identified and a robust semantic definition among these concepts was built in a unified meta-model. We claim that the unified GOM meta-model could help in a) analysing existing goal models in order to provide insights regarding different goal modelling perspectives b) identify semantic similarities / overlaps between existing GOM techniques c) provide the basis for a reference model for GOM.

Keywords: Goal oriented modelling, goal oriented meta-model, goal oriented language.

1 INTRODUCTION

In general, goal-oriented actions are actions directed towards the realization of some specific state of the world (Castelfranchi and Paglieri, 2007). Inspiring by this way of thinking, the goal oriented modelling approaches in information systems (IS) rooted to three decades back as a requirements elicitation, modelling, analysis and validation technique. That time practitioners realized the need to trace the rationale of IS development which was impossible to capture by other software engineering techniques. Goal-oriented modelling adopts a top-down analysis approach in order to elicit system requirements from the systems environment aiming to develop a valid information system. Goal oriented approaches offer rich semantic and syntax presented either in terms of natural language specifications or graphical notations. Goal oriented approaches and techniques come under different names such as goal-driven engineering, goal modeling (GM), goal oriented requirement engineering (GORE), goal oriented modelling (GOM). The goal concept has emerged from research in Artificial Intelligence (goal-directed autonomous agents) as well as organizational/enterprise modeling (goal-directed organizational behavior). Furthermore, GOM research has its roots in design problem solving and cognitive research that suggest the use of goal-driven processes in many kinds of activities that
humans perform. Hence GOM languages use a vocabulary inherited from these fields e.g. agents, roles, constraints, obstacles, beliefs, expectations, strategies, plans and so on. Supplementary concepts have been added in order to describe the process of goal setting and refinement including scenarios and context, among others.

Since the early 1990’s, goal-oriented modeling has become an essential element of the IS research filed and, to a lesser extent, industrial practice. In particular, for requirements engineering (RE) activities goal models have been used to elicit, represent and analyze a) Organization requirements e.g. strategic goals (Bleistein et al., 2006), risk (Asnar et al. 2011), operation (Santos et al., 2010), organizational change (Kavakli and Loucopoulos, 2006), etc. b) Multiple stakeholders requirements e.g. agents tasks (Lapouchian and Lespérance, 2006), trust (Yu and Liu, 2000), point of view (Kaiya and Saeki, 2004), collaboration (Yu et al., 2011), resources and capabilities (Danesh and Yu, 2014) c) Information systems requirements e.g. system requirements (Lamsweerde, 2001), software functions (Schnabel and Pizka, 2006; Lapouchian et al., 2006), self-adaptive systems (Bryl and Giorgini, 2006; Liaskos et al., 2012), security (Liu et al., 2003), safety (Kelly and Weaver, 2004), general non-functional evaluation (Chung et al., 2000) etc. As a result, different goal modeling approaches have been proposed each having different semantics, concepts and notations due to the fact that researchers have adopted different views on what is actually a goal and what are the boundaries of the goal model in a particular context and for a particular purpose.

In modern business environments, there are increasingly more situations (e.g. joint-venture, multi-level strategies, joint and distributed multi-projects) where a single GOM technique is neither practical nor feasible as project participants may use different modeling languages. Recent research has stressed the need for a holistic unified GOM language, in order to maximize its usage and deployment (Patrício et al. 2011). Such unification requires a clear and robust understanding of the semantics as well as the differences and similarities between different GOM concepts. This will allow the definition of a unified view of GOM languages in systematic manner. This unified view can assist the analysis of existing goal models in order to provide insights regarding different goal modeling perspectives, as well as to identify semantic similarities / overlaps between existing GOM techniques. To this end, this paper adopts an abstraction process that integrates eight well-known GOM techniques’ concepts into a single and unified meta-model, described in section 2. Section 3 presents the unified goal oriented meta-model and discusses the different aspects of GOM. Related work is discussed in section 4. Finally, section 5 concludes the paper highlighting open issues that provide the foundation for further research in the GOM field.

2 THE PROCESS OF BUILDING A UNIFIED META-MODEL

A meta-model is an explicit model of the constructs and rules needed to build specific models within a domain of interest, in this case goal modeling. The intention in this paper is to use metamodeling in order to create a unified meta-model for the purpose “integrating” existing GOM approaches, bringing together existing artifacts suggested in different GOM meta-models. The resulting extensible unified goal oriented meta-model provides a language independent goal oriented ontology. The mainstream GOM approaches which this meta-model is based on are: The Knowledge Acquisition in automated Specification (KAOS) (Dardenne et al. 1993; Objectiver, 2007), the Enterprise Knowledge Definition (EKD) goal meta-model (Loucopoulos et al., 1997; Kavakli and Loucopoulos, 1999), the Business Motivation Model (BMM) (OMG, 2010), the i* framework (Yu et al., 2011), the Goal-Structuring Notation (GSN) (Kelly and Weaver, 2004; Attwood et al., 2011), the Non-Functional Requirements (NFR) framework (Mylopoulos et al., 1992; Chung et al., 2000), the Goal-Based Requirements Analysis Method (GBRAM) (Anton, 1996) and Technie (Borgida et al., 2009).

Figure 1 depicts the integration process in terms of 3 levels: model-level, meta-level and meta-meta level. Different representations of a single goal model in the aforementioned GOM languages are shown at the lowest level of abstraction (model level), together with their meta-level representations as the second level. An integrating meta-meta model is presented at the highest level. The unified GOM meta-model development process includes the following steps: (1) generating individual GOM meta-models, (2) concept mapping, and (3) concept integration.
In the first step, for each GOM approach, its concepts and their interrelations is described as a meta-model using a common foundation language (in this case UML). Generation of the meta-models was based on studying existing models’ descriptions found in the literature. This was not straightforward since most languages do not provide an abstract syntax. Even if a meta-model was provided, in most cases non-standard constructions were used to visualize it, omitting multiplicities, specialization-related constraints and abstract classes. Integrity constraints were only given partially and informally. Thus, special attention was necessary in order to cover all represented concepts and their relations in the developed meta-models.

Although, different GOM meta-models include similar concepts these are captured by different meta-models in different ways, e.g. using different names or different structure. Therefore, it is necessary to perform a mapping between concepts of the different meta-models eliminating any redundancies (step 2). This mapping involves the analysis of GOM concepts based on their definition found in the literature. A review of goal modeling constructs is shown in Table 1. Concepts expressing similar aspects of reality are grouped together.

The third and final step concept integration, refers to the unification of the concepts representing the same aspects of reality to a single general concept at the meta-meta-level. Indeed, we identified a number of overlapping concepts. For instance, assumption and expectation can be read the same; also achievement goal and objective are equal. Maintenance goal and quality goal can be non-functional or soft goal, while undeveloped goal appears as a type of versioning and goal status. Goal types such as hard and soft goals were proposed for modeling goals, other types such as End, Mean and sub-goals presented to describe the operationability of the goal. For instance, strategy and plan can say the same thing, operation and process can be equally reflected in real life scenarios. Also if we look at issues, obstacles, constraints and challenges constructs, we also assume those can all grouped in issues, where issues can be challenges, obstacles, constrains (can also be part of the environment/context, similar to events). Belief, assumption, expectation and claim are cognitive states of either the actor or the analyst, which those also can be overlapping in essence when it comes to industrial practice. Relations in i* such as (make, help, hurt, break positive and negative, unknown) can be summarized in three scales (support, conflict and hinder). We comprehend that most of the GOM approaches focused on
the modelling of the system environment, requirements and constraints. The purpose is to better feed into the software applications design; we locate events into goal execution aspects, where multidirectional events may impact on goal achievement, goal execution and its surrounded context. Finally we also see claim as equal to belief. Table 2 summarizes our findings based on definition analysis of the eight methods concepts.

<table>
<thead>
<tr>
<th>Concept</th>
<th>KAOS</th>
<th>EKOS</th>
<th>BHM</th>
<th>IT* / Tropes</th>
<th>CSN</th>
<th>NFR</th>
<th>GBRAM</th>
<th>Techne</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal achievement</td>
<td>Goal</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Soft goal</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Maintenance goal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Achievement goal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quality goal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Undeveloped goal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vision</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mission</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objectives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strategy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tactics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operation</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Process</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goal ownership</td>
<td>Agent</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Actor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Role</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Position</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actor Boundary</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Context</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domain property</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goal execution</td>
<td>Issue</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Obstacle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Constraint</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Requirement</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Event</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Solution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Resource</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goal formulation</td>
<td>Expectation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Belief</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Claim</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Argument</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assumption</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Option</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scenario</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Preference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Overview of goal oriented modelling constructs

<table>
<thead>
<tr>
<th>Concept</th>
<th>Overlapped concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard Goal</td>
<td>Functional Goal</td>
</tr>
<tr>
<td>Soft Goal</td>
<td>Non-functional Goal</td>
</tr>
<tr>
<td>Maintenance goal</td>
<td>Type of Non-functional Goal or Constraint</td>
</tr>
<tr>
<td>Achievement goal</td>
<td>Objective</td>
</tr>
<tr>
<td>Quality goal</td>
<td>Type of Non-functional Goal</td>
</tr>
<tr>
<td>Undeveloped goal</td>
<td>Claim or goal status</td>
</tr>
<tr>
<td>Plan</td>
<td>Strategy</td>
</tr>
<tr>
<td>Operation</td>
<td>Process</td>
</tr>
<tr>
<td>Requirements</td>
<td>Technical type of goals</td>
</tr>
<tr>
<td>Constraint</td>
<td>Part of the requirements</td>
</tr>
<tr>
<td>Solution</td>
<td>Type of goal to be achieved</td>
</tr>
<tr>
<td>Belief</td>
<td>Claim</td>
</tr>
<tr>
<td>Argument</td>
<td>Belief + assumptions</td>
</tr>
<tr>
<td>Scenario</td>
<td>Part of argument</td>
</tr>
<tr>
<td>Option</td>
<td>Part of argument</td>
</tr>
<tr>
<td>Preference</td>
<td>Part of argument</td>
</tr>
<tr>
<td>Expectation</td>
<td>Type of goal, also it can be an assumption. But we see it in our integrated metamodel as an cognitive element relevant to prediction of unknown</td>
</tr>
</tbody>
</table>

Table 2. Goal oriented modelling overlapped constructs
Based on the previous analysis we have constructed our unified GOM meta-model. It should be noted that meta-model integration relies significantly on the analysts’ experience in finding appropriate logic of ontological and taxonomical relations without losing the expressiveness of the meta-model concepts. Figure 2, presents an overview of the integrated GOM meta-model using UML class diagram. Generalization has been used in order to represent the generic concept for several concepts that represent the same aspect of reality.

![Figure 2. The overview of the integrated GOM meta-model](image)

The main assumption in the proposed integration is that GOM languages express similar concepts. This makes it possible to create a common integrated meta-model. Conceptually, this integrating meta-model represents a union of all the concepts found in the GOM languages. This paper, focus on semantic interoperability only. Additional work is needed in order to address technical interoperability issues. However, we believe that this meta-model is a first step towards the definition of a common goal reference model.

### 3 Analysis of the Meta-Meta-Model Constructs

A shown in Figure 2, the concepts of the unified goal oriented meta-model can be categorized into four different aspects with respect to the type of goal reasoning that they support, namely: goal ownership (who), goal formulation (why), goal achievement (how), goal execution (what), aspects. Figures 3 to 6 present the classification of the general GOM concepts of the integrated meta-model with respect to the four GOM aspects, as well as their inter-aspects relationships (in dark grey). In particular, Figure 3, describes the goal ownership aspect including all concepts related to the organizational entities that are responsible for achieving a goal. Figure 4, depicts the concepts representing the goal formulation aspect (why), including the concepts related to the cognitive process of goal formulation.
Figure 3. Goal ownership aspect (who)

Figure 4. Goal formulation aspect (why)

Figure 5. depicts the concepts involved in the goal achievement aspect. These concepts are relevant to the process of goal analysis and decomposition of high-level goals to operational goals.
Figure 5. Goal achievement aspect (how)

Finally, Figure 6, corresponds to the goal execution aspect (what) representing the concepts internal and external to the organization that influence (cause, trigger or control) the goal execution.

Figure 6. Goal execution aspect (what)
4 RELATED WORK

This work is in line with our previous work reported in Kavakli and Loucopoulos (2005), which reported the analysis of 15 GOM languages along four dimensions: “usage” (what RE activity does goal modelling contribute to?), “subject” (what is the nature of goals?), “representation” (how are goals expressed?) and “development” (how are goal models developed and used?). The result of this analysis indicated the fragmented nature the need for more integration in the field of GOM. In Kavakli (2002) we further proposed a unification of goal meta-models at the “usage” level.

Analysis of relevant literature reported in Horkoff et al. (2014) shows that a decade later the GOM picture has not changed since the authors reach the conclusion that “many approaches are narrowly focused, with most approaches focusing only on a few stages of the software lifecycle, not often providing an end-to-end solution”.

A number of recent approaches have dealt with the horizontal transformation between GOM languages. For example Matulevičius et al. (2007) describes a comparison of KAOS and i* (GRL) using UEML as the foundation ontology. Rather than providing a unified GOM model their aim is to identify semantic discrepancies of the two approaches. Patricio et al. (2011) focus on syntactical mapping between the two approaches mentioned previously. Similar to our work, Nwokeji et al. (2013) define a consolidated intentional modeling language using a Model Based Software Engineering (MBSE) language integration technique; however the scope of this work were limited to two goal modelling languages (once again KAOS and i*). The work presented in this paper presents a holistic approach in terms of a) the number of GOM techniques integrated b) the semantic mapping between GOM constructs c) the clarification of GOM syntax using UML class diagrams and d) the classification of the GOM meta-model constructs with respect to different aspects of goal modeling.

5 CONCLUSION AND FUTURE WORK

This paper proposes an integrated goal oriented meta-model that is language-independent. The meta-model was developed through abstraction based on eight well-known GOM approaches.

The value of a unified, language independent, GOM meta-model is that it provides a unique semantic specification of goal oriented concepts and their relationships thus eliminating invalid interpretations by experts in different domains. As such it can act as a reference between multiple GOM techniques of the same project.

In addition, this work contributes to the clarification of the GOM syntax using UML class diagrams. Furthermore, the mapping process has revealed ontological redundancies of GOM concepts meaning that two concepts have the same or overlapping semantics i.e., they refer to the same things.

Furthermore, analysis of the integrated meta-model constructs may reveal further issues such as incomplete domain coverage, when a language does not convey information on a certain aspect of the application domain. Indeed, the analysis of the integrated meta-model constructs indicates that current GOM approaches focus on goal setting and execution and do not address goal evaluation an important aspect of goal formulation. Aligned to goal evaluation is goal adaptation, also not dealt with in current GOM approaches, whereby goals are supposed to be steady and there can only be predefined alternative plans of goal execution. However, evaluation of the results of goal execution might in turn trigger the adaptation of existing goals or the formulation of new goals. Incorporating goal dynamics requires the definition of appropriate goal states and a set of operations for moving between states.

We claim that modeling of goal dynamics is more suitable for today’s open service oriented systems characterized by the heterogeneity and autonomy of the participating agents. Furthermore, it could better fit the requirements of intelligent, context aware systems.
References


A QUALITATIVE STUDY OF THE SIGNIFICANT
MOTIVATORS FOR SHOPPING ONLINE AMONG
UNDERGRADUATE STUDENTS IN THE UK

Mohamed Haffar, Faculty of Management, Bournemouth University, UK, 
mhaffar@bournemouth.ac.uk
Wafi Al-Karaghouri, Brunel Business School, Brunel University London, UK
Wafi.Al-Karaghouri@brunel.ac.uk
Ahmad Ghoneim, Brunel Business School, Brunel University London, UK
Ahmad.Ghoneim@brunel.ac.uk
Sulaf Assi, School of Applied Sciences, Bournemouth University, UK
sassi@bournemouth.ac.uk
Harry Phillips, Faculty of Management, Bournemouth University, UK
7673370@bournemouth.ac.uk
Joseph Wilkinson, Faculty of Management, Bournemouth University, UK
4323012@bournemouth.ac.uk

Abstract
This paper examines the factors of online shopping, which motivate undergraduate students (UG) at a
UK university. A thorough analysis of the relevant literature reveals that consumer's attitude and
intention towards shopping online are not only influenced by TAM's basic constructs but also by
exogenous factors, e.g. consumer traits, situational factors, product characteristics, previous online
shopping experiences, and trust in online shopping. However, there are limited empirical studies on the
drivers/motivators to shop online. In particular, a critical review and analysis of the literature reveal that
a contribution to knowledge would be derived from the literature backed by the investigation of the most
important drivers for shopping online amongst Bournemouth undergraduate students. The study also
explores the relative importance of such factors that encourage students to shop online. To do so, in-
depth interviews with 15 UG students were conducted to collect the primary data to gain an insight into
students’ online shopping intentions and draw conclusions on the factors that influence their behaviour.
The analysis of the data collected indicates that UG students held positive attitudes towards the online
shopping medium. The findings suggest that the most significant motivators for shopping online among
Bournemouth undergraduate students are perceived ease of use and perceived usefulness followed by web
experience, trust and convenience, while enjoyment was found to be of the least relative importance.

Keywords: Internet Marketing, TAM, Undergraduate Students

1. INTRODUCTION

The growth of the internet is considered an extraordinary event (Krishna and Guru 2010) and is
unprecedented in comparison to adoption rates of other information technology tools (Pallab, 1996). A
report published in 2012 identified the UK as Europe’s leader for online shopping reporting 82% of UK
internet users made an online purchase in 2011 (OFNS, 2013). 73% of British people used the internet
every day in 2013, Office for National Statistics – OFNS- (2013). Ecommerce has been suggested as the
“truly revolutionary impact of the Internet Revolution...” (Drucker, 2002, p.3). The number of consumers
using the internet to buy products is on the increase. In 2013, 72% of all UK adults used the internet to
buy products online, up 19% from2008 (OFNS 2013).
Students are a prevalent demographic in using the online shopping medium. A study of American college students revealed that 91% had made a purchase online (Lester et al. 2005). A study by Ahuja et al. (2003) indicated that students had higher future purchase intentions to shop online than nonstudents across a range of products. Higher levels of educate people are associated with an increased likelihood of using the internet to shop as they are more comfortable using non-traditional shopping channels (Burke, 2002), owing to high levels of internet literacy (Li et al., 1999; Johnson, 2005; Rahaman, 2014).

There have been many studies into the adoption of technology in using a variety of different models, such as the technology acceptance model (TAM) (Davis, 1989), the theory of Planned Behaviour (TPH) and the Diffusion of innovation (DoI) theory (Citrin et al., 2000; Vijayasarathy and Jones, 2000; Goldsmith, 2001; Teo, 2001; Chen et al., 2002; Carter and Weerakkody, 2008). The Technology Acceptance Model (TAM) (Davis, 1989 & 1993; Venkatesh and Davis, 2000) can be used to explain how, when presented with a new technology; users/potential users face a number of factors which influence their decision about how and when they will use it. TAM has been proven to be a feasible model for testing a consumer’s acceptance of new technology. However, recently in order to define more specific drivers for attitudes towards online shopping there have been new factors added to TAM’s basic constructs, perceived ease of use and perceived usefulness, in terms of exogenous variables. These exogenous variables improve the viability and predictive nature of TAM and allow it to be applied to online shopping. There is however a lack of empirical studies to test the impact of these factors on a consumer’s intention to shop online, thus presenting an opportunity to further research the impacts of such factors on a consumers intention to shop online (Jin and Osman, 2014).

Whilst indeed studies have used the TAM as a means of understanding consumers usage of certain technologies (Venkatesh and Davis, 2000; Gefen et al., 2003; Al-Azzam, 2014; Khan, 2015). There have been few studies which used TAM as a means of understanding students online shopping behaviour (Bigne et al., 2005; Hsieh and Liao, 2011; Eden et al., 2012; Lim and Ting, 2012; Bankole, and Babalola, 2012). Whilst these provide some insight, there is a significant lack of studies concerning the factors that contribute to online shopping behaviour and in particular amongst students in the Dorset area. This study is significant as it attempts to identify the key factors that influence UK students’ decision making when making online purchases.

2. INTERNET MARKETING

The ever growing emphasis on internet marketing can be seen by the continuous growth of European advertising expenditure which was valued at €24.4 billion in 2012, of which €6.442 billion came from the UK, a 13% increase from 2011 (Fennah, 2012).

Internet retail environments are increasingly playing a part in the relationship between marketers and consumers (Demangeot and Broderick, 2007). The interactive and interpersonal nature of the internet means that online users are able to control content thus, online marketing uses a pull strategy based on consumer response (Peterson et al., 1997; Bigne et al., 2005). This two way communication leads to significant market opportunities for internet marketers (Warrington et al., 2000; Waldo, 2000). Internet marketers are at a significant advantage in that they can share information cheaply and quickly in comparison to traditional mass marketing approaches (Warrington et al., 2000; Waldo, 2000).

Despite the potential advantages, many organisations fail to adopt an internet marketing strategy to utilise their products and services. Wolf (1998) argued that this is due to the fact that many organisations fail to adapt their traditional mass marketing strategies to allow for the differences faced by the online marketer. However, adopting an online marketing and retail strategy may not always be possible due to the nature of the product itself. The unsuitability of the product is one of the main barriers to internet commerce (OECD 2007, p.122.). Furthermore, research by Cales (2000) found that four times as many individuals spent time searching for product information online than actually made an online purchase. This may be due to a higher perceived risk when purchasing a product online as the consumer is unable to inspect the product physically (Chiou and Pan, 2009; Szymanski and Hise, 2000).
3. THEORETICAL EXPLANATIONS OF ONLINE PURCHASING BEHAVIOUR AND THE FACTORS THAT INFLUENCE IT.

Consumers are more likely to make an online purchase due to ease of use, offers quick access, it is easy to make a purchase, make an order cancellation and make returns (Novak et al., 2000; Szymanski and Hise, 2000; Limayen et al., 2000; Abels et al., 1996). Retailer reputation, length of product warranties, variety of products, price and customer service are also important factors for consumers to consider in making their decisions (Limayen et al., 2000; Haubl and Trifts, 2000). Importantly for this study individuals with higher levels of education have been found to be more likely to shop online as they are more comfortable with the technology (Burke 2002). As higher levels of education is being associated with higher levels of internet literacy (Li et al., 1999; Eden, et al., 2012).

Perceived levels of risk in comparison to traditional means that, shopping has been found to be higher when shopping online in terms of privacy, economic and financial risk (Chiou and Pan, 2009; Vijayasarathy and Jones, 2000). However, the aforementioned studies used anon-probability convenience sampling and thus the results may lack generalisability. Whilst some studies used students as participants (Haubl and Trifts, 2000; Lester et al., 2005), in Canada and the USA and thus the results may not be generalised to UK students.

A thorough review of the literature has indicated that many theories have been used frequently when studying the behavioural intention in technological products are of the Technology Acceptance Model (TAM) (e.g. Teo, 2001; Chen et al., 2002; Moon and Kim, 2001; Yulihasri et al., 2011) the Theory of Planned Behaviour (TPB) (Vijayasarathy and Jones 2000; Shim et al., 2001), the Diffusion of Innovation (Citrin et al., 2000; Goldsmith, 2001 and the Theory of Reasoned Action (TRA) (Bezjian-Avery et al., 1998; Chen and Wells, 1999).

The Theory of TRA is a well-established theory, having successful predicted behavioural intention for a wide range of consumer products (Chung and Pysarchik, 2000). However, it is not suitable to online shopping as it fails to take into account external factors that may influence a consumer’s behaviour when making a purchase online (Leelayouthayotin , 2004). Furthermore the Theory of Planned Behaviour is more suited to complex situations where motivations are not fully understood.

4. TECHNOLOGY ACCEPTANCE MODEL (TAM)

TAM was devised by Davis (1989) to extend the TRA. Although the core variables of TAM are tailored to understand the adoption of computer-based technologies in the work place, it has been proven to be a suitable theoretical foundation when measuring the adoption of shopping online (Chen et al. 2002; Goldsmith, 2002; Venkatesh, 2000; O’Cass and Fenech, 2002). Furthermore the variables of TAM, Perceived usefulness (PU) and Perceived ease of use (PEOU) have been proven on multiple occasions to be of both high reliability and validity (Chen et al. 2002; Venkatesh and Davis, 2000). Several authors have also found that these variables provide significant benefits from a reduced set of measures over other existing models (Chau, 1996). The review of literature identified that TAM is the most heavily cited theory in explore adoption of information technologies, and has been found to be suitable in exploring E-commerce adoption (e.g. Teo 2001; Chen et al. 2002; Moon and Kim 2001; Yulihasri et al. 2011).TAM has much empirical support in the adoption of online shopping and is a far more reliable, cost effective model for analysing behavioural intention.

However, TAM has been criticised for failing to take into account the influence of social aspects on technology acceptance (Chen et al. 2002; Moon and Kim 2001). Although PEOU, PU and Enjoyment
have been identified as being fundamental in determining the acceptance and use of information technology in the work place, it is believed that these constructs may not explain the consumer’s behaviour towards newly emerging technologies such as online shopping as the factors contributing to the acceptance of new technologies are likely to vary based on the technology, target users and context (Moon and Kim, 2001). Therefore, due to the specific nature of this research focusing on the phenomenon of online shopping it is necessary to use additional factors beyond the existing constructs of Perceived Ease of Use, Perceived Usefulness and Enjoyment. These factors have been identified as Product characteristics, Web experience and Trust and will be discussed in the following sections.

4.1 Constructs of TAM

4.1.1 Perceived Usefulness

Perceived usefulness (PU) has been found that there is an indirect link between PU and the consumer’s attitude and a strong direct link between usefulness and intent (Davis, 1989; Davis et al.’ 1992; Gefen et al., 2003; Venkatesh et al., 2003). Furthermore, Koufaris (2002) validated PU as a construct that was seen to influence the intention of Internet shoppers. However, Jackson et al. (1997) and Lucas and Spitler (1999) have criticised the link between PU and attitude stating that there is no empirical evidence to support the relation between the two.

In the context of this research the new technology is shopping online and performance refers to the outcome of shopping online. Therefore, PU refers the consumer’s belief that shopping online will enhance the outcome of their shopping experience (Monsuwe et al. 2004). The outcome of the shopping experience depends on the following determinants: shopping performance, shopping productivity and efficiency of accomplishing shopping goals (McCloskey, 2004). This was supported by Childers et al. (2001) who state that consumers with a favourable attitude towards online shopping had used online retailers which enable shoppers to increase shopping productivity, effectiveness and ability thus resulting in online shopping being perceived as useful. In addition to this, Kim et al. (2003) found that online shopping sites that provide functions in order to help consumers in making more educated shopping decisions will be perceived as useful by consumers. Furthermore, consumers perceive online shopping sites that are able to provide services to consumers that are not available through traditional shopping such as the comparison of products as useful, contrary to earlier criticisms this leads to the consumer developing a more favourable attitude toward online shopping (Bisdee, 2007). Therefore the researchers believe the criticisms of Jackson et al. (1997) and Lucas and Spitler (1999) are only applicable to the adoption of technology and not online shopping. Thus, further research is needed to validate PU in the context of online shopping and to explore whether it is seen to be an important driver.

4.1.2 Perceived Ease of Use

Perceived ease of use (PEOU) is defined as, the degree to which a person believes that using a new technology would be free of effort (Davis, 1989). Existing research has found perceived ease of use to have an effect on usage intention both directly and indirectly through its effect on perceived usefulness (Davis et al. 1989; Venkatesh, 2000; Venkatesh and Morris, 2003). Contradictory to these studies many other researchers have found no empirical evidence to support the relation between perceived ease of use and perceived usefulness (Chau and Hu, 2002; Bajaj and Nidumolu, 1998; Hu et al. 1999; Jackson et al. 1997). In addition to this, Chau (1996), Hu et al. (1999) and Lucas and Spitler (1999) found no relation between perceived ease of use and behaviour intention.

In the context of this research PEOU is the consumer’s perception that online shopping will be free from effort, this then refers to the process that leads the consumer to making an online purchase (Venkatesh, 2000; Monsuwe et al., 2004). Contrary to earlier criticisms, PEOU has been said to play a major role in online shopping (Monsuwe et al., 2004; Venkatesh, 2000) as a result the authors believe the
criticisms are only applicable to the adoption of information systems rather than online shopping. However, Gefen and Straub (2000) found that PEOU only has an indirect influence on online shopping behaviour by directly influencing PU thus suggesting that in the adoption of online shopping PEOU isn’t as a significant driver as PU. It is argued that the easier online shopping is the more useful consumers will perceive online shopping to be (Ramayah and Ignatius 2005; Childers et al. 2001; Moon and Kim, 2001). Therefore, further research is needed to determine if PEOU is an important driver of online shopping for students and whether there is a link between PEOU and PU.

4.1.3 Enjoyment

Enjoyment is defined as the extent to which the activity of using the new technology is perceived to provide reinforcement in its own right apart from any performance consequences that may be anticipated (Davis et al. 1992). The role of enjoyment is to enrapture the intrinsic motivation, for which there is a strong body of theoretical and empirical evidence regarding its importance in the use of technology (Davis et al. 1992; Venkatesh and Speier, 1999 & 2000). In the context of this research Enjoyment comes from the emotions stimulated during the online shopping experience (Monsuwe et al. 2004). Childers et al. (2001) found enjoyment to be a consistent and strong predictor of consumer attitude towards online shopping, as a result, if consumers enjoy their experience of online shopping they are more likely to develop a positive attitude toward online shopping and thus more likely to engage in further online shopping. In addition to this, Ramayah and Ignatius (2005) suggested that those who are more computer literate are more likely to gain a sense of enjoyment from shopping online. The researchers will look to test whether enjoyment is a strong predictor of attitude toward online shopping and if it is seen as a driver for online shopping among students.

4.1.4 Exogenous Factors

Product characteristics

Product characteristics is the next external factor, the consumer’s decision of whether to shop online or not is influenced by the type of product or service they intend to purchase (Monsuwe et al., 2004; Dan and Xu, 2011). For the purpose of this research the authors have adapted the work of Jain and Jain (2011) in order to categorise product characteristics into four types depend on what consumers are more likely to buy online, i.e. Commodities, semi-commodities, fulfilment intensive and high ticket/ large size. Commodities are typically standardised and familiar products such as books, DVDs, music and games for example, which are more likely to be purchased online due to the fact that there is no need for physical assistance or pre purchase testing and there is little doubt on the product quality (Grenewal et al., 2002; Reibstein, 1999; Jain and Jain, 2011). Furthermore products of a more sensitive nature such as health and beauty products fall into the semi-commodity category, as they are likely to be bought online due to the level of privacy and anonymity that online shopping offers (Grenewal et al., 2002; Jain and Jain, 2011). On the other hand, it is believed that high ticket value products that require more knowledge or experience when making a purchase such as cars or computers are more likely to be purchased through traditional methods and not online (Elliot and Fowell, 2000). There is however, a lack of research into the type of products which students buy online in the Dorset area. As a result the researchers will explore the role that product characteristics play in motivating student to shop online and if there is any relation between the type of product and intent to shop online.
Web experience

The way the consumer interacts with the website can be controlled in order to positively influence the consumer’s intention to make a purchase online, such as design, events and emotions (Constantinides, 2004). In addition, the web experience in particular is seen as a determinant of the construct perceived ease of use, and the ease of which information can be understood as having an effect on the perceived ease of use (Lederer et al., 2000).

Kotler (2003) states that web experience is a new additional controllable determinant of the consumer’s behaviour when making an online purchase. This driver takes into account elements such as the initial searching for the product/service, quality of information content, the selecting, comparing and evaluating of information in order for the consumer to make an informed decision when making a purchase (Kotler, 2003; Tarafdar and Zhang 2005; Salaun and Flores 2001). Moreover, many studies indicate the importance of web design (Flavin et al. 2006; Huizingh, 2000; Lee and Lin, 2005; Zviram et al. 2006), especially the significance of visual attractiveness (Heijden, 2003) in influencing the consumers purchase intention. Therefore, the researchers will seek to understand the role of web experience in motivating students to shop online. If there is a link between web experience and PEOU, in order to test the importance of web design in determining online purchase intent.

Trust

Online trust is dependent on the consumer’s perception of whether the risks are greater than the perceived benefits of the online transaction (Teo and Liu, 2007). Gefen et al. (2003) found that trust in an online retailer reduces the consumers concerns about the uncertainty, complexity and risk of online shopping, therefore increasing their intentions to purchase. There is much theoretical and empirical support for integrating trust with the existing TAM constructs. This is due to the fact perceived risk and trust of the consumer are seen as having a significant influence on a consumer’s intent to purchase online (Pavlou, 2003; Castillo and Nicolas, 2007; Monsuwe et al. 2004; Schaupp and Belanger 2005). Chircu et al. (2000) believed that trust in an online retailer increases PEOU. If the consumer has trust in the retailer they feel more in control and as a result don’t feel the need to monitor the retailers’ actions or check every detail meaning there is less time and effort involved in making a purchase. However, in a context where trust is low the opposite can occur, the consumer feels less in control and thus the consumer checks every detail meaning process involves more time and effort (Pavlou, 2003). Therefore, we will look to explore if trust is seen to be a driver in online shopping for students and if there is any link between trust, PU and PEOU.

5. METHODOLOGY

An exploratory research strategy was adopted by the researchers as not much is currently published about the factors influencing UK students to make online purchases. Exploratory research has been used in many studies (Yilmaz and Hunt, 2001; Keaveney and Parthasarathy, 2001; Keaveney, 1998; Wang et al.2000; Rodgers and Harris, 2003) and aims to increase a researcher’s understanding of a topic allowing for subsequent research to be conducted with a solid understanding. Thus, adopting an exploratory research strategy was best suited to allow fulfilment of the study aims and objectives.

Due to the exploratory nature and to meet the objectives of the study, we used non-probability purposive sampling. Purposive sampling is a form of non-probability sampling that is mostly associated with qualitative research. Whereby the researchers do not seek participants on a random basis and instead uses a more strategic approach in order to ensure the chosen sample are relevant to the questions being asked (Bryman, 2012; Cooper and Schindler, 2003; Blumberg et al. 2005; Easterby-Smith et al. 2008). Therefore the researchers used non-probability purposive sampling in order to select participants based on the following criteria: that are students of Bournemouth University and shop online. The main critique of purposive sampling is the chance of sampling bias (Wilson, 2010). However, we have overcome this by
having clear criteria that is relevant to the question thus avoiding poor judgement (Wilson, 2010; Zikmund and Babin, 2007).

The researchers decided to conduct semi-structured interviews to collect primary data. Fifteen semi-structured interviews were conducted with students studying various degrees at Bournemouth University. Eight participants were males and seven were females. The researchers developed a series of questions that were used as an interview guide, the questions were based on the constructs of TAM outlined in the literature. This ensured reliability between each interview whilst still allowing the interviewer to ask additional questions should he feel that a question or topic area had not been answered or understood adequately.

Participants were assured that the data collected would remain confidential (Maylor and Blackmon, 2005). In addition permission to record the interviews was obtained (Maylor and Blackmon, 2005). To improve the reliability and validity of the method and reduce possible bias, the interviews were audio-recorded and notes were taken to aid further analysis (Zemliansky and Lowe, 2011). Each interview lasted 30-45 minutes, allowing sufficient time to explore the topic.

6. DATA ANALYSIS AND FINDINGS

For the purposes of this study, a qualitative content analysis was used to analyse the data collected from the interviews. This method is the most popular for the analysis of semi-structured interviews (Jankowicz, 2005), as it creates codes and categories in a systematic manner, to explore large amounts of textual data unobtrusively and interpret and identify trends, meanings and patterns (Adams et al. 2007). To achieve this, all the interviews were then transcribed, coded and carefully analysed. Although content analysis has been criticised for decontextualising information, it is still believed to be the best approach, as it translates long transcripts into attitudes, intentions and emotions to help answer the research question (Grbich, 2011).

6.1 Perceived Usefulness

A thorough review of the previous studies indicates that the PU of online shopping is an important factor that influences student’s online shopping behaviour. Information gathered from the in depth personal interviews highlights that the majority of students believe online shopping is more beneficial than other shopping mediums such as high street shopping. For example, interviewee 13 (P13) stated, “...I believe ecommerce has enabled a significant power shift in the market, whereby consumers now have greater purchasing power than ever before, due to greater range, price comparisons and ability to view and write reviews regarding services, products and customer relations”. However, few participants and in particular female participants only believed this partially, for example p6 stated, “I wouldn’t say it’s more beneficial I like both. I like the convenience of shopping online but I enjoy being able to see clothes in store.”

The findings show perceived usefulness is a key determinant of student’s online shopping behaviour. This proves similar to that of the literature as it identifies and validates perceived usefulness as a factor that influences the intention of Internet shoppers (Koufaris, 2002). This is because students perceive online shopping to improve shopping performance, productivity and efficiency (McCloskey et al. 2004). Findings indicate that participants have strong usefulness perceptions for the online shopping medium. Many participants indicated that online shopping was useful in terms of time saving in comparison to traditional means of shopping and provided the user with more information:

P9: ‘...the time that you save...you can put into product research.’
P6: ‘...Being able to access all the information in front of you on your computer...saves you a lot of time'
This leads to consumers having a positive attitude toward shopping online, as they believe online shopping is more useful and beneficial than traditional methods (Monsuwe et al., 2004). While few respondents did state they felt online shopping didn’t improve their shopping performance due to delivery costs making it more expensive, the majority of the participants indicated that the online shopping medium was useful in allowing them to find the cheapest price. The ability to compare prices on the internet was cited, by some, as the reason behind this:

P1: ‘...It’s plain and simple that online shopping is cheaper...’
P3: ‘...you can compare different retailers’
P12: ‘...there are a lot of price comparison websites. That means that you can go to the retailer that is selling it at the cheapest...’
P7: ‘most things you can actually find cheaper online rather than physically going to find something’
P14: ‘...you can go on the compare sites and find things that you wouldn't necessarily find out if you just picked it up in

It is concluded that that perceived usefulness is an important driver of online shopping for students. Most of the interviewed students perceived that online shopping improves the outcome of their shopping experiences. Furthermore, when asked to rank the relative importance of the factors tested usefulness and their effect on was consistently ranked as one of the most important factors in influencing their online shopping behaviours, Therefore the findings relating to PU support various research (Davis et al. 1989; Koufaris 2002; Hsieh and Liao 2011; Yulihasri et al. 2011; Jiang et al. 2000) which indicated that PU strongly influences an individual’s intention to shop online.

6.2 Perceived Ease of Use (PEOU)

The findings indicate that students at Bournemouth University have strong ease of use perceptions of the online shopping medium. Some participants however, reported that whilst they found online shopping easy, this varied dependent on the website they use:
P11: ‘...it does depends on which website you use.’
P5: ‘It depends on the website but yes generally.’

Furthermore, other participants identified that whilst they found online shopping easy, that this was because of their significant experiences with the internet. For example:
P8: ‘...it's easy for me, I've been doing it a long time so I know what I'm doing.’
P13: ‘I would consider it easy but I have a lot of experience with the internet’
P6: ‘I find it easy to use but that is because I'm familiar with the internet and how it works.’

In addition participants believed that it cost little in terms of effort to become skilled at online shopping. For example:
P1: ‘It takes very little effort...’
P7: ‘...I don’t think there has ever been a point where I’ve found it especially difficult.’
P4: ‘...I grasped it straight away, it was nothing difficult.’
P11: ‘...you pick it up very quickly...’
Furthermore, participants indicated that they found it easy to use the online shopping medium to find the products they are looking for:

P8: ‘...You know which websites are going to be able to provide you with the product. It’s normally quite easy to navigate those websites.’
P12: ‘...I think they make it quite easy. The best thing is when they have got a search engine.’

In addition participant 9 added “It’s easier as you can do it whenever you want to and there are no queues.”

The findings show similar evidence to the existing literature, which perceived ease of use is an important driver of consumer intent when shopping online (Ramayah and Ignatius, 2005). The findings support that there is a link between PEOU and PU as the easier online shopping is; the more useful consumers will perceive online shopping to be (Ramayah and Ignatius 2005; Heijden, 2000; Childers et al. 2001; Moon and Kim, 2001). This is because all participants perceive the process of online shopping as being easier than traditional methods, thus suggesting it is more useful. This result was because online shopping gives the consumer greater control of the process, allowing the consumer to browse and complete the purchase at their convenience. However, contradictory to this some students find the process confusing. This could be overcome by online retailers taking measures to improve the online shopping process by implementing a unified process and improving the web experience.

Contrary to the assertions of Davis (1989, 1993), who suggested that PEOU is of particular importance during the initial period of user experience with a technology, when ranking the relative importance of the factors discussed, PEOU was ranked as the most important factor. Therefore, the findings suggest that PEOU is not limited to significance during the early stages of adoption of the online shopping medium as all participants had significant online shopping experience, yet still ranked, on average, PEOU as the most influential factor.

6.3 Enjoyment
The interviews revealed that the majority of participants do not enjoy the process of shopping online. This was especially apparent of the male participants all of whom said that they did not enjoy the process of shopping online. Participants stated that they use the online shopping medium for convenience as appose to enjoyment. For example:

P9, P7: ‘No, I don’t enjoy it, I do it because it’s convenient.’
P13: ‘I just consider it more of a necessity. It’s the most convenient option and sometimes there is no need to speak to somebody about it if the requirement isn’t complicated.’

Contrary to the findings of various research (Childers et al. 2001; Cheema et al. 2013; Koufaris 2002; Eighmey 1997), whom asserted that enjoyment is a strong predictor of consumer attitudes towards online shopping, the majority of participants indicated that they didn’t enjoy the process of shopping online, yet still demonstrated positive attitudes and made regular online purchases. Enjoyment was found to be, on average, of least significance. Thus the findings of previous studies (Childers et al. 2001; Cheema et al.2013; Koufaris 2002; Eighmey 1997) are challenged by the findings as the majority of participants revealed that they did not enjoy the process of online shopping yet still held positive attitudes and made regular online purchases.
6.4 Trust

The results of this research suggest that trust is an important factor in determining whether participants. Many participants indicated that they felt there was more risk involved in making an online purchase in comparison to traditional shopping supporting the findings of several previous studies (Szymanski and Hise 2000; Limayen et al. 2000; Korgaonkar and Wolin 1999; Vijayasarathy and Jones 2000; Chiou and Pan 2009). For example:

P3, P7: ’It is easier to trust a traditional retailer because you’re there when your purchasing what you want, so you know what you’re getting’
P5: ’...I’d say it’s definitely easier to trust a traditional retailer as you’re able to build a physical relationship with them’

Interviewees said that they would only purchase goods from well-known online retailers and wouldn’t purchase from online retailers whom they didn’t trust.
P.11, “…I do tend to stick to the same sites and ones which I know are genuine websites.” Participant 9 also added: “I always go to the same sites, I know I can trust them and you can get most of what you need from one place these days. However if I cannot find what I need/want I will look for new sites that may offer this.”
P14: ’...when I want to buy a product I look straight away at the well known brands such as Amazon...’
P6: ’...I wouldn’t use them unless I trusted them.’
P9: ’...trusting a company will definitely be a deciding factor as to whether I would purchase online...’

Our results support the findings of Gefen et al. (2003) in that trust increases intention to purchase, as the respondents all said they would go to a familiar site first when looking to make a purchase online. This then also links in with the findings on web experience that building trust in the consumer increases purchase intent.

6.5 Situational Factors

The interviews revealed that situational factors have an influence on participants’ online shopping behaviours. Participants reported that if they had little free time they’d be more likely to shop online, for example:
P3: ’...I don’t always have a lot of free time anyway...I can just quickly come online and it only takes me between five and ten minutes...’
P9: ’...I would be more likely to shop on line if time was an issue...’

The findings from the interviews indicate that situational factors do play a role in participants’ online shopping behaviours. This finding supports existing studies such as that of Venkatesh (2000) who states that the perceived convenience of online shopping has a positive impact on the consumers’ attitude toward online shopping. In line with the findings of Donthuand Garcia (1999) participants revealed that if under ‘time pressure’ they’d be more likely to shop online citing the ability to shop online at any time quickly and conveniently; thus supporting various research (Castells 1997;Wolfinbarger and Gilly 2001; Perea et al. 2004).

However, few participants revealed affiliations with traditional retailers that have led them to shopping traditionally; despite being able to purchase the same products online still have an affiliation with specific traditional retailers. For example:
P 3: ’...I do like to go into somewhere like HMV where you can look through the titles. I actually enjoy that...’
P12: 'There is a local Burton which is about five miles from my house. I'd much prefer to travel the five miles...than purchase from Burton online. I know the people in the store and can go there and try the product out.'
P15: 'When shopping for makeup I prefer to go into House of Frasier....'

Nevertheless, these participants still had strong and positive attitudes towards online shopping, thus supporting the assertions regarding 'attractiveness of alternatives' made by Perea et al. (2004).

6.6 Prior Internet Experience

The findings show that prior web experience is an important factor that influences a student's intention to shopping online in the future.

P11: “I do tend to find it very easy to find what I'm looking for while shopping online. With Features such as search tools and filters which enable me to refine my search I tend to find what I'm looking for quickly.”
P5: “Yes, the abundance of information and goods available online means there is always someone selling what you need.”
P2: ‘...the reason I go back to shop on Asos is because every time I’ve had a positive experience...'
P3: ‘...Amazon and Play.com are often cheaper and you get free delivery and I have often had quick responses from them. That has made me go back to them as its been good for me.’

Respondents indicated that negative online shopping experiences had limited further online purchases with the responsible retailer. For example:
P1: 'I've had examples with Amazon where I have bought university books where it's taken weeks on end to arrive...that negative experience has made me reluctant to use Amazon.'
P14: '...Amazon. I had a year membership with them which ran out...they took it upon themselves to charge fifty pounds...for a further year membership. I will probably never use Amazon again...'
P12: '...I bought something from a clothing website. It didn’t fit right, the fstitching was wrong and it wasn’t a good item. I haven't used it since.'

The findings that our study has supported the previous research study such as Shim et al. (2001) and Eastwick and Lotz (1999) who found a relationship between prior internet experience and online shopping intentions.

In further support of Shim et al. (2001) and Eastlick and Lotz (1999) findings indicated that negative online shopping experiences prevent further purchases with the retailer whom was responsible for the negative experience. However, negative experiences did not prevent the respondents from shopping online, but did encourage them to look for alternative retailers.

Moreover the findings of the research suggest that the initial impression of a website has a significant effect on students purchase intent, as all participants believe the initial impression of a site is important. P 9: “...I’d say I make up my mind on a site within the first few seconds of clicking on it. If I don’t think it looks legit I will look elsewhere but if it looks well designed it builds trust and I’m more likely to have a look at what they are offering.” P10: “…If the website looks dodgy I won’t purchase anything.”

The Findings are consistent to that of Flavin et al. (2006), Huizigh, (2000), Lee and Lin, (2005), Zviram et al. (2006) and Wade and Renata (2011) studies which all found that web design is an important
motivator of online shopping. This was because the students found it easy to find what they want/need when shopping online. Such a result is mainly due to the increase in use of search bars, filters and the quality of content in web design that empowers the consumer and enables them to find what they want quickly. Moreover these findings in particular confirm the association between visual attractiveness of a website and purchase intent (Hejden, 2003), as there is clear correlation between the initial impression of a website and the students intention to proceed to a purchase. In addition to this, good web design is found to build trust and confidence in the consumer where as poor web design builds uncertainty and acts as a deterrent.

6.7 Product Characteristics

The findings indicate high cost items unsuitable for online purchase owing to increased risk and the desire to physically inspect products of high value prior to making a purchase. For example:

P7: ‘...Anything that is quite expensive I wouldn’t purchase online.’
P15: ‘...where you’re going to part with large amounts of money, it’s better to go in store and talk with someone...’
P8: ‘I would say products that are more expensive. You would like to have a physical experience with the product before you purchase it. You are taking more of a risk when purchasing something online for a higher price.’
P12: “As I said I think I buy everything online, I guess the only things I could think of that I wouldn’t are big purchases such as a car or house. However, I’m a student at the end of the day so they are a little out of my price range at the moment!”

The results show that a product characteristic isn’t an important factor in motivating students to shop online shopping. This is because students are open to buying anything online however; the research found students tend to buy commodities due to their low disposable income. The findings differ from that of Monsuwe et al. (2004) they stated that product characteristics is an important factor in influencing consumers online shopping however, the researchers have identified it isn’t among the population of UG students. Whereas, the results are consistent with those of Jain and Jain (2011), Reibstein (1999) and Gernewal et al. (2002) in that the majority of products students purchased online where commodities. However, there is a floor in the findings as the researchers could probed beyond the respondents first answer to make them think about higher ticket items rather than what they associate with purchasing while shopping online. Moreover, the researchers believe that it’s not the product characteristics that motivate students to shop online but instead the choice of products. Product choice was identified as a significant driver for students being mentioned 8 times. The availability of products online is far greater than that of the high street with unlimited access to retailers offering a far wider selection of products. Furthermore this not only applies to the wide range of products but also stock in that stores have a limited amount of stock in comparison to online retailers who typically have far more, meaning the consumer is more likely to find what they need thus increasing consumer purchase intent. This finding is validated by being consistent to the published research which had identified the wider selection of products as being a motivation to shop online (Srinivasan et al. 2002; Szymanski and Hise 2000; Wolfinbarger and Gilly 2001; Yoon 2002). Therefore we concluded that a product characteristic is not a valid influence on student’s online shopping behaviour.

7. CONCLUSIONS

Whilst indeed studies have used the TAM as a means of understanding consumers usage of certain technologies (Gefen et al., 2003; Venkatesh and Davis, 2000). There have been few studies have used
TAM as a means of understanding students online shopping behaviour (Hsieh and Liao 2011; Lim and Ting 2012). TAM is the chosen model that will be used by the researchers to develop a greater understanding of the factors that influence undergraduate students when shopping online. This is due to its empirical standing in the adoption of ecommerce and its reliable nature and cost effectiveness. In addition, exogenous factors were reviewed and added to the traditional constructs of TAM in order to achieve the most thorough means of examining student’s usage of online shopping.

The findings indicate that the factors that influence Bournemouth University students to shop online and their relative importance are as follows: Ease of use, usefulness, trust, situational factors, prior internet experience, product characteristics and enjoyment. It can be concluded that the original constructs of TAM; ease of use and usefulness were perceived to be the first and second most important factors respectively in influencing users’ online shopping behaviours. This indicates the applicability of TAM for exploring students online shopping behaviours in the UK, supporting previous research that has asserted the suitability of TAM in exploring e-commerce (Chen et al., 2002; Moon and Kim 2001; Lederer et al., 2000; Yulihasri et al., 2011; Hsieh and Liao, 2011; Lim and Ting, 2012).

Trust was, on average, ranked as the third most important factor by participants, followed by situational factors. Prior internet experience and product characteristics were ranked joint fifth. Enjoyment was perceived as the least important factor. These findings challenge existing research (Childers et al. 2001; Cheema et al. 2013; Koufaris 2002; Eighmey 1997) that asserted enjoyment to be a strong and consistent predictor of attitudes towards online shopping. Most participants stated that they did not enjoy the process yet still demonstrated positive attitudes and made regular online purchases.

8. MANAGERIAL IMPLICATIONS

The results of this study offer several managerial implications for practitioners and online retailers on how to efficiently plan their investments and align resources in order to further entice students to shop online.

A lack of trust was found to be one of the primary factors that limit students online purchasing. Emphasis should therefore be placed on demonstrating that online retailers can be trusted. The lack of face-to-face contact was found to be one of the reasons that limited participants’ trust of online retailers.

Online retailers and marketers should therefore integrate a 'human touch’ into their websites. This could be achieved by integrating video tutorials and video support; using programmes like Skype, allowing the user to build relationship with an individual as appose to solely relying on faceless interaction.

Guarantees regarding website security and safety should are also important in maximising user trust perceptions. Furthermore, substantial investment in technical support should be made to mitigate potentially negative experiences; which were found to limit student’s future usage of online retailers. Therefore we suggest that the government implement legislation to create a unified process, this would then eradicate any confusion as well as building consumer trust and security when shopping online. Such a change would then have a positive effect on the PEOU of online shopping.

In addition, convenience was identified as a significant motivator of online shopping among students and thus retailers should look to enhance the convenience of online shopping to the consumer and in turn this would increase the level of online shopping they partake in. This could be achieved by improving delivery, which is a key dimension of convenience, reducing the cost, improving delivery times, and speed would all have a significant effect and increase the PU of online shopping.

Also, web developers should look to design sites for that are both aesthetically pleasing yet still functional. The initial impression of a website proves crucial in determining consumers purchase intent by building consumer trust and confidence. However, online shopping is seen to be a necessity by many and as a result the website should still prove to be functional in order to allow the consumer to browse and purchase with ease. This can be achieved through effective use of filters, search bars and quality content.
In addition, consumer reviews are also seen as useful content in aiding the consumer to make a purchase. These changes would then have a positive effect on both the PU and PEOU of online shopping amongst consumers.

Moreover, investment should be spent on maximising the ease of using online retailer websites. This should in turn provide students with time savings and further positive usefulness perceptions of the online shopping medium. Time saving, monetary savings and convenience were elements of usefulness that were frequently cited as important factors in determining whether or not participants shopped online and were consistently found to be of more importance than enjoyment. Thus, online retailers and marketers should prioritise allowing students to make quick and efficient online purchases as appose to maximising user enjoyment.

9. LIMITATIONS AND FUTURE RESEARCH

Despite the useful findings of this empirical study, there are some shortcomings that lead to avenues for future research, as follows:

One limitation is that this study was based on the cross-sectional research design, as opposed to a longitudinal design, which does not allow affirmative causal explanations and restricts the researchers' ability to tackle or refer to the change or development of the phenomenon under study over a period. Owing to rapid technological changes and increasing adoption of the online shopping medium, a future work using longitudinal study into the online shopping behaviours of students would allow changes to be explored over time. This will allow for more precise and clearer picture of the drivers for shopping online amongst student in the UK.

It would be interesting to see the change in consumer behaviour with the growth of e-commerce and changes in technology. For example a study to explore the difference in consumer behaviour and attitudes toward shopping online between generation X whom have always had access to the Internet as a shopping medium and the current sample of generation Y to whom online shopping is a relatively new phenomena. In addition, future research could explore gender differences and their effect on students’ online shopping behaviours enabling recommendations to be made to online retailers and marketers on how best to adapt their online strategies to cater for gender specific markets.

The researchers conducted 15 semi-structured exploratory interviews with UG students as an initial pilot study and supporting method. The findings of our current study can’t be generalised to students throughout the UK. However, this initial stage has helped us to make sure that the research problem exists and the research questions were worth investigating. This is because it is strongly advisable not to rush into detailed surveys before less expensive and more readily available sources of information have been exhausted (Zikmund, 2000). Therefore, quantitative questionnaire survey will be conducted at the subsequent stage, over a larger sample of UG students from different universities to further validate and generalise the findings. The findings obtained will be reported in a future paper. It is hoped that hybrid research methods (quantitative and qualitative) will be used in data collection techniques and analysis procedures either at the same time or consecutively (Saunders et al. 2012; Jogulu &Pansiri, 2011). Triangulation methods provide more viewpoints and perspective as well as deeper and broader information on the phenomenon being studied (Cooper & Schindler, 2003; Creswell, 2009; Zikmund, 2003) and to increase the quality, validity and reliability of the findings and decrease the degree of bias (Bouma, 1996).
REFERENCES


BUILDING AN INNOVATIVE SOLUTION FOR WILDFIRE PREVENTION AND MANAGEMENT: THE “AEGIS” PLATFORM

Nikos Athanasis, email: thanasis@geo.aegean.gr, Department of Geography-University of the Aegean
Christos Kalloniatis, email: chkallon@aegean.gr, Department of Cultural Technology and Communication-University of the Aegean
Michail Vaitis, email: vaitis@aegean.gr, Department of Geography-University of the Aegean
Christos Vasilakos, email: chvas@aegean.gr, Department of Geography-University of the Aegean
Kostas Kalabokidis, email: kalabokidis@aegean.gr, Department of Geography-University of the Aegean

Abstract

The Web wildfire prevention and management information system AEGIS is currently under development aiming to reduce potential human, environmental and property losses. AEGIS leverages advanced capabilities of Geographic Information Systems (GIS) (like parallel processing in the Cloud and utilization of artificial neural networks) through a user-friendly web interface, without the need for extensive training on commercial or complicated GIS applications. This work describes the analysis of the requirements procedure followed for the development of AEGIS. Emphasis is given on how different user groups can be identified and how their opinion can be taken into account to enhance the usefulness of the proposed system. The methodology described can be used as a framework in the development of forest fire management systems.

Keywords: wildfires, Web-GIS, requirements analysis

1. INTRODUCTION

In Greece, wildfires constitute the most devastating natural disasters along with floods and earthquakes. The current operational wildfire danger rating system is based on a qualitative semi-empirical approach. In addition, fire simulation during a wildfire event are not easily utilized by the fire departments that are responsible for the confrontation of each event. On the other hand, a wildfire management system including fire danger assessment and fire simulations cannot rely on methodologies that have been developed per se for other countries, because they do not take into consideration the local particularities of the Greek climate, vegetation, topography and human geography.

Within this context, the AEGIS system was designed to appear as a cost effective, easy-to-use forest fire management system, independent of commercial software for the end-users. The main functionalities of the system are a fire danger rating system and a fire behavior modelling scheme. The structure of the algorithms will be based on parallel computer processing techniques to ensure both scalability and efficiency of the calculations.

The scope of this paper is to present the requirements analysis of the AEGIS system. Research experience in the utilization of wildfire prevention systems (Grasso and Singh, 2009; Barber et al., 2010; Chuvieco, 2004; Davies et al., 2009; Fiorucci et al., 2008; Lee et al., 2002) as well as in the development of web-based GIS platforms for forest fire control (Kalabokidis et al., 2012), ontology-based geo-portals for wildfires (Kalabokidis et al., 2010; Kalabokidis et al., 2011) and the development of wildfire applications in the Cloud (Kalabokidis et al., 2014) is a valuable asset towards the requirement analysis of the platform. Users have the ability, without the requirement of knowing the handling of complicated fire management systems, to integrate fire science models and decision
support planning modules and to utilize the capabilities of the system. Artificial neural networks and innovative geospatial tools are utilized, while advanced propagation algorithms are used for fire behavior modeling and mapping including burning consequences.

Furthermore, the goal-modeling representation was used for conducting the analysis of the functional and non-functional requirements of the system under development in order to conduct the mapping and proper analysis of the main organizational and functional goals into respective sub-goals as well as system processes. Finally, all use cases along with the respective user groups are identified and presented with UML diagrams.

Based on the importance of requirements analysis stage and the knowledge IT professionals should acquire for the system to realize all organizational and functional goals, the paper is structured on a respective manner following the analysis process. Specifically, in Section 2 the AEGIS platform is presented, while the goal-process model of the AEGIS platform is presented in Section 3. The goals and the respective sub-goals are derived from the aforementioned description for the system under development. In Section 4, the end-user requirements analysis is presented. Specifically, for verifying the usefulness of the proposed system from key user groups a questionnaire was decided for taking the input on the requirements of the system. In Section 5, the respective use cases are presented. All user categories along with the group of services that they can access are presented using UML diagrams. In Section 6, the architectural components of the system are presented. Finally, Section 7 concludes the paper and raises points for future development.

2. **THE AEGIS PLATFORM**

The AEGIS application is designed and implemented as a web-based platform and provides access to fire prediction data (risk and behavior), as well as additional information such as socioeconomic activities, roads, land uses, water tanks locations, patrol routes, fleet tracking, satellite images, vegetation types, terrain and weather data. All functionalities provided by AEGIS are accessible to the local fire agencies and authorities of civil protection through an appropriate graphical user interface. Without any special knowledge about fire risk and fire propagation modeling, end-users are able to utilize a wide range of maps directly in the web browsers of their PC’s, portable computers, Global Positioning Systems (GPS) or smartphones; i.e. with no other specialized tools required to be installed in their devices. Thus, with this approach a valuable assistance and a decision tool is offered to the local authorities responsible for wildfire management to extract useful information towards the design of an effective operational wildfire prevention and management plan. All available output results are visualized by utilizing the web-GIS design tools that provide powerful mapping and geo-processing functionalities free-of-charge.

With the AEGIS innovative and advanced programming tools, firefighting personnel, emergency crews and other authorities can design an operational plan to encompass the forest fire. Fire management professionals are able to locate vehicles of the Fire Service and other resources online and in real-time. Fire patrol aircrafts using GPS tracking and communications have the ability to send coordinates for each item to the system, depicting them on an electronic map; and detection cameras could augment these data by transmitting images of high risk areas into the AEGIS fire system. The AEGIS platform enables end-users to query the databases and get answers immediately, locate points of interest in high-resolution satellite images and connect their portable computers or GPS devices to download information.

The AEGIS application is able to offer services beyond simple coordination of emergency activities. Remote automatic weather stations and a weather forecasting system based on the SKIRON/RAMS weather model provides crucial data needed for fire prevention and early warning. Geographical representation of the fire risk potential and identification of high-risk areas at different local regions is being provided daily, based on parallel computer processing techniques; to resourcefully improve the current fire risk estimation methodology used by the Greek General Secretarial for Civil Protection and come up with the first ever quantitative Greek Fire Danger Rating System. By using and testing the innovative proposed fire behavior algorithms, maps are produced on demand and real-time to graphically represent the spread and intensity of a forest fire at different times and places, including burn probabilities and fire effects. All these information are calculated with advanced parallel processing computer techniques, in this proposed state-of-the-art project in terms of human, technical and research resources.
The developed system is applied in seven (7) different study areas with high-hazard, high-value and high-use wildland-urban environments. Each area covers a mix of different conditions either in socioeconomic situations (i.e. rural/urban and interface areas, changes in population size/density etc.) or in environmental factors (i.e. weather, vegetation, topography etc.). By applying the results and outcomes of this research in the areas, knowledge is gained and tools are developed that may allow us to apply the system to the rest of Greece with minimal effort and resources in the future.

3. ANALYSIS OF GOALS AND PROCESSES

Initially, the analysis of the organizational goals and the corresponding processes that must be followed is conducted. The purpose of goal-processes analysis is to find the specific objectives that must satisfy the system to be built. The result of this analysis is a UML diagram that conceptually represents the relationships between objectives and processes.

The analysis starts with the most generic goal and concludes in concrete objectives, the implementation of which leads to achieving the overall objectives set out in the more generic levels. Each final target is associated with one or more processes. The processes are those that determine the developing system services and provide a guide for the development team, because the implementation of processes means the achievement of the objectives set.

The mapping of the goals (G) and processes (P) follows the reverse tree format where the highest node is the root of the tree (Figure 1). The analysis using the target-process model is necessary to precede any other analysis, because it provides:

- A clear imprint of the objectives of developing the system.
- A mapping of relationships between objectives.
- An analysis of the objectives in specific ones that contribute to easier and faster implementation of the system.
- The confirmation that all requirements have been taken into consideration.

The diagram in Figure 1 shows the model of the goals and processes revealed by the analysis of the system. The model describes the objectives derived from functional and non-functional requirements. Functional requirements describe the basic needs that must be satisfied by the system and from which all services of the system are derived for each user group. Non-functional requirements describe mainly rules that should be applied to all services of the system and usually are about the safety of users and data, the maintenance of the system’s availability and data integrity. In general, non-functional requirements lead to “horizontal” implementation of services for the whole system and not individual services. The processes presented facilitate the easier transition from the theoretical objectives in the implementation of the services offered by the system.

As stated above for completing the goal-process model the following tasks were conducted:

- An analysis of the system needs as being expressed by the description presented in the introduction. The aforementioned description was shaped with the system’s stakeholders and the team of analysts.
- Interviews with the major end-users the respective system is intended to be used from.
- Identification of non-functional requirements based on previews experience of the designer team as well as the respective literature described in the introduction.

The result of the analysis is presented in Figure 1. All decomposition links between parent (higher level) and child (lower level) goals are “AND” decomposition links meaning that the realization of the parent goal requires the successful realization of all child goals. The proposed services that will be implemented through the identified processes are matched with the end goals in Table 1.
Figure 1  The Goal – Process diagram
### Table 1: Analysis of services, goals and processes of the AEGIS platform

<table>
<thead>
<tr>
<th>Service</th>
<th>Goal (G)</th>
<th>Process (P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visualization of fire risk maps (Vasilakos et al., 2007)</td>
<td>G1.1</td>
<td>P1, P7</td>
</tr>
<tr>
<td>Visualization of burn probability maps (Ager et al., 2011)</td>
<td>G1.2</td>
<td>P2, P3, P7</td>
</tr>
<tr>
<td>Visualization of real time weather data</td>
<td>G2.1</td>
<td>P3</td>
</tr>
<tr>
<td>Visualization of forecast weather data</td>
<td>G2.1</td>
<td>P2, P7</td>
</tr>
<tr>
<td>Access to fire management data, i.e.:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Base maps / satellite maps</td>
<td>G1.3</td>
<td>P1</td>
</tr>
<tr>
<td>• road network</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• water sources</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• evacuation sites</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• cover types</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• high risk areas</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exploitation of Google Earth (KML) data</td>
<td>G1.3</td>
<td></td>
</tr>
<tr>
<td>Online map creation</td>
<td>G2.3</td>
<td>P6</td>
</tr>
<tr>
<td>Map printing</td>
<td>G2.3</td>
<td></td>
</tr>
<tr>
<td>Access to historical fire risk data</td>
<td>G2.2</td>
<td>P4</td>
</tr>
<tr>
<td>Access to historical weather data</td>
<td>G2.2</td>
<td></td>
</tr>
<tr>
<td>Access to historical burn probability maps</td>
<td>G2.2</td>
<td></td>
</tr>
<tr>
<td>Routing</td>
<td>G2.3</td>
<td>P5, P11</td>
</tr>
<tr>
<td>Finding the closest routes to water sources</td>
<td>G2.3</td>
<td>P5, P11</td>
</tr>
<tr>
<td>Calculate drive times from a specific location</td>
<td>G2.3</td>
<td>P5, P11</td>
</tr>
<tr>
<td>Location tracking of fire vehicles on duty</td>
<td>G2.3</td>
<td>P6, P11</td>
</tr>
<tr>
<td>Visualization of new fire spots</td>
<td>G2.3</td>
<td>P6, P11</td>
</tr>
<tr>
<td>Access to web cameras</td>
<td>G2.3</td>
<td>P6, P11</td>
</tr>
<tr>
<td>Access to the provided information through mobile apps</td>
<td>G5</td>
<td>P12</td>
</tr>
<tr>
<td>Fire behavior simulation</td>
<td>G1.2, G1.3, G2.1, G2.2</td>
<td>P1, P2, P3, P11</td>
</tr>
</tbody>
</table>

4. **User Requirements Identification**

To confirm the usefulness and necessity of the aforementioned analysis, a questionnaire with the proposed services was compiled. The questionnaire was answered by different groups of potential key users to consider the priorities of the functionalities that will be developed. Questions were either qualitative (e.g. “How important do you consider the calculation of the risk?”) or quantitative (e.g. “What is the preferred time frame for the creation of the fire risk maps?”). The full questionnaire is presented in Appendix I. The analysis of the user’s replies to the questionnaire in order:

- To confirm that the proposed services cover the requirements of the potential users of the system.
- To emerge the individual characteristics of the services that will help the developers to better
customize these services.

The questionnaire was drawn up with the help of the free, web-based software Google docs and emailed in a large sample of the end users. Particular emphasis was given to the heterogeneity of the sample; i.e. users should have true connection with the subject of the work, but they should come from different organizations. The questionnaire was answered by 34 people who work in research organizations, higher education institutions, fire and forest services, business organizations and firefighting volunteers.

The services offered were grouped into functional and non-functional requirements. Functional requirements describe the proposed operations that should be running by the system. Non-functional requirements include general characteristics of the system. Users were asked to answer how important would be if the system was built according the following characteristics:

- Efficiency; where users were asked about the necessity each of the services to be available in low response time.
- Availability; where users were asked about the necessity each of the services to be available at specific time periods.
- Usability; where users were asked about the necessity each of the services to have an operation guide and assistance for their use.
- Security; where users were asked about the necessity each of the services to be accessible only through credentials.

The analysis of the results from the questionnaire answered by the groups of potential users is shown in Table 2. The specific results assist the designers and developers on building a system that will satisfy the user's needs in a more holistic manner.

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preferred time frame for the creation of the fire risk maps</td>
<td>72 hours</td>
</tr>
<tr>
<td>Risk maps to be provided by the application within the preferred time frame</td>
<td>1 map / 6 hours</td>
</tr>
<tr>
<td>Preferred maximum time duration of fire simulations</td>
<td>12 hours</td>
</tr>
<tr>
<td>Preferred time frame of accessing historical fires</td>
<td>1 decade</td>
</tr>
<tr>
<td>Fires interested to have access to burned / reforested areas</td>
<td>&gt; 100 ha</td>
</tr>
<tr>
<td>Preferred measurement units of wind speed</td>
<td>km/h and Beaufort</td>
</tr>
<tr>
<td>Types of base maps that should be incorporated</td>
<td>Satellite images and annotation</td>
</tr>
<tr>
<td>Preferable period in which the provided services will be available</td>
<td>Throughout the year for all provided services except of risk assessment and burn probability services that are necessary from May until end of September.</td>
</tr>
</tbody>
</table>

Table 2  User requirements through questionnaire analysis

5. USE CASES

After the analysis of the results from the questionnaire, different use cases of the system are modelled by using UML use case diagrams. These diagrams contribute to the emergence of the services that each group of end users has access. As a consequence, the services can be developed by assigning the appropriate access rights to the corresponding user groups. Three groups of users were identified:
1. Basic user group: In this group belong non-registered users that will be able to utilize only a subset of the services provided. These services include any non-critical functionality; e.g., visualization of any map (except fire behavior simulations maps), accessing real-time weather data and searching for historical weather data.

2. Privileged user group: This group will have access to the full functionality provided; critical functionalities such as accessing maps of fire behavior simulations will be provided only through the corresponding credentials.

3. System’s administrators: This group contains the Web Server’s administrators as well as the Cloud’s platform administrator.

Because of the large number of the services offered, Figure 2 shows only the use cases for each user group. Table 3 lists the detailed services for each use case, as well as the user group that has access for the specific service.

Specifically, the first user group is able to access these kind of data and tools that are not considered private and confidential for fire prevention and management. Thus, the basic group of users is able to visualize fire risk maps, burn probability maps, real time weather data, forecast weather data as well as historical data. However, they are not able to utilize geospatial analysis tools (e.g. routes to the closest water sources) and conduct fire behavior simulations. Only authorized users can exploit these tools. Finally, administrators cannot only use geospatial analysis tools and conduct fire behavior simulations, but they also can manage the web server as well as the allocation of the Cloud resources.
<table>
<thead>
<tr>
<th>Service</th>
<th>Use cases</th>
<th>User’s group</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Visualization of fire risk maps</td>
<td>visualization of maps</td>
<td></td>
</tr>
<tr>
<td>• Visualization of burn probability maps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Visualization of real time weather data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Visualization of forecast weather data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access to fire management data, i.e.:</td>
<td>fire management data</td>
<td>Basic</td>
</tr>
<tr>
<td>• Base maps / satellite maps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• road network</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• water sources</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• evacuation sites</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• cover types</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• high risk areas</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Exploitation of Google Earth (KML) data</td>
<td>mapping tools</td>
<td></td>
</tr>
<tr>
<td>• Online map creation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Map printing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Access to historical fire risk data</td>
<td>historical data</td>
<td></td>
</tr>
<tr>
<td>• Access to historical weather data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Access to historical burn probability maps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Routing</td>
<td>geospatial analysis</td>
<td>Privileged</td>
</tr>
<tr>
<td>• Finding the closest routes to water sources</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Calculate drive times from a specific location</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Location tracking of fire vehicles on duty</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Visualization of new fire spots</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access to web cameras</td>
<td>Web cameras</td>
<td></td>
</tr>
<tr>
<td>Fire behavior simulation</td>
<td>Web server administration</td>
<td></td>
</tr>
<tr>
<td>Administration</td>
<td>Web server and Cloud administrator</td>
<td>Administrators</td>
</tr>
</tbody>
</table>

Table 3  
Alignment of the proposed services in corresponding user’s groups
6. ARCHITECTURAL COMPONENTS

Figure 3 shows the software components of the proposed platform. In the front-end layer, end-users will utilize a wide range functionalities directly in the web browsers of their PC’s (through the web application) and smartphones (through the corresponding mobile app for Windows Phone devices), with no other need to install any specialized tools in their devices.

The ESRI ArcGIS API\(^1\) is used for the construction and implementation of the platform’s graphical interface, based on the Silverlight framework. The API allows the development of web-based applications that integrate geo-processing and mapping services provided by the geographical server ArcGIS Server. Among the hosting of the Silverlight application, the Web Server (IIS) is responsible for the retrieval of the meteorological data collected from the remote automatic weather stations. Data are collected at the remote stations by special sensors with the help of special software and are sent to the server where they are collected, analyzed and visualized. The values are stored in specific databases so that users have access to weather data from any desired time period. Data from external sources (e.g.

\[\text{https://developers.arcgis.com/silverlight/}\ ^1\ \text{https://developers.arcgis.com/silverlight/} \]
new fire ignitions captured by satellites, images from web cameras and weather forecast data) are collected through the http and ftp protocols.

The ESRI ArcCatalog and ESRI ArcMap software is used for digitization, analysis and management of geographic data. Field data are collected from all the study areas to create spatial and non-spatial databases. Spatial data include road networks, vegetation cover types, fuel types, water sources, topography, dispatching resources, social structures and urban areas, etc. Existing databases, if available, are updated based on these data. In order to create a reliable fire risk and behavior database, we were obligated to undertake field inventories to acquire information about vegetation cover types of the study areas; and then, to create spatial datasets through the use of remote sensing (by acquiring recent satellite images), spatial statistics and GIS techniques. The result are spatial datasets of fuel models, canopy cover, canopy bulk density and stand height; i.e. the necessary inputs for predicting surface and crown fires with fire behavior models. The data are organized into geographic databases and are published in the form of geographic services from the server ArcGIS Server. The geographical services involve viewing static data (cartographic, road network, hydrant locations, evacuation areas, land use, high-risk sites, vegetation types), and services like finding shortest route between points, finding the nearest hydrant seats set point, calculating the extraction region from the designated point at a specified time and digitization fire incidents and hydrant locations.

Furthermore, Cloud-based and HPC-based components are utilized to ensure both power and speed of the calculations. The internal operating mechanisms of the Cloud platform ensure high connectivity and high availability of data and services through appropriate Service Level Agreements (SLAs). This ensure the continuous functionality of the provided services. The available nodes in the Cloud (beyond the visualization of results) undertake the calculation of fire ignition risk maps, the calculation of the burn probabilities maps and the conduction of the fire behavior simulations. To reduce the overall cost of Cloud resources usage, the number of available nodes is increased or decreased, depending on the actual needs. The number of available virtual machines is increased when new forecast weather data are available (i.e. every morning at a predefined time). After scaling up, every virtual machine downloads external data (i.e. real time weather for a specific hour of the next five days data), internal data (i.e. static data stored in the Cloud such as topography, vegetation, fuel types and socioeconomic inputs) and generates the corresponding maps. Execution is conducted in the Cloud by assigning a task to a specific virtual machine (Figure 3). Following job execution, the deployment is automatically scaled down. The results of the calculations are stored in special files BLOBS (Binary Large ObjectS) and are accessible to end users through the graphical interface of the application. Table 4 lists the proposed services and the software component, which participate in the development of the service.

<table>
<thead>
<tr>
<th>Service</th>
<th>Architectural Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Visualization of fire risk maps</td>
<td>Web app, Windows Phone app</td>
</tr>
<tr>
<td>• Visualization of burn probability maps</td>
<td></td>
</tr>
<tr>
<td>• Visualization of real time weather data</td>
<td></td>
</tr>
<tr>
<td>• Visualization of forecast weather data</td>
<td></td>
</tr>
<tr>
<td>Access to fire management data, i.e.:</td>
<td>ArcGIS Server</td>
</tr>
<tr>
<td>• Base maps / satellite maps</td>
<td></td>
</tr>
<tr>
<td>• road network</td>
<td></td>
</tr>
<tr>
<td>• water sources</td>
<td></td>
</tr>
<tr>
<td>• evacuation sites</td>
<td></td>
</tr>
<tr>
<td>• cover types</td>
<td></td>
</tr>
<tr>
<td>• high risk areas</td>
<td></td>
</tr>
<tr>
<td>• Exploitation of Google Earth (KML) data</td>
<td>Web app</td>
</tr>
<tr>
<td>• Online map creation</td>
<td></td>
</tr>
</tbody>
</table>
Table 4  Services and corresponding software components.

7.  CONCLUSIONS

This article presents the requirements analysis of the wildfire prevention and management platform AEGIS. One of the compelling advantages of the system lies in leveraging GIS capabilities without the need for extensive training on commercial or complicated GIS applications; especially that evidently, operational forces lack the know-how and expertise to develop and operate their own computing and IT systems. With the AEGIS innovative and advanced programming tools, firefighting personnel, emergency crews and other authorities will design an operational plan to encompass the forest fire, pinpointing the best ways to put it out with new levels of precision. Fire management professionals will locate vehicles of the Fire Service and other resources online and in real-time. Fire patrol aircrafts using GPS tracking and communications will send coordinates for each item to the system depicting them on an electronic map; and detection cameras could augment these data by transmitting images of high risk areas into the AEGIS system. The methodology described can be used as a framework in the development of forest fire management systems. Analysis of the requirements is essential, because stakeholders and the team of analysts have different conceptual approaches for the functionalities and capabilities of the system.

The current prototype is applied in 7 different study areas of Greece. Further research plans may include the utilization of the application in different geographical areas and in larger spatial contexts. By applying results and outcomes of this research on the study areas, knowledge is to be gained and tools to be developed that may allow the expansion of the system to the rest of Greece or other areas with minimal effort and resources.

8.  ACKNOWLEDGMENTS

The research project “AEGIS: Wildfire Prevention and Management Information System” (Code Number 1862) is implemented within the framework of the Action ARISTEIA of the Operational Program "Education and Lifelong Learning" (Action’s Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State. Final results and outcomes of the project are expected by September 2015.
References


### APPENDIX I

**QUESTIONNAIRE FILED BY THE END USERS**

<table>
<thead>
<tr>
<th>Functional Requirements</th>
<th>Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>How important do you consider the calculation of the risk?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>not important</td>
</tr>
<tr>
<td></td>
<td>quite important</td>
</tr>
<tr>
<td></td>
<td>necessary</td>
</tr>
<tr>
<td>What is the preferred time frame for the creation of the fire risk maps?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>24 hours</td>
</tr>
<tr>
<td></td>
<td>72 hours</td>
</tr>
<tr>
<td></td>
<td>120 hours</td>
</tr>
<tr>
<td>Within the time frame chosen in the previous question, how many risk maps would you like to be provided by the application?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 map / hour</td>
</tr>
<tr>
<td></td>
<td>1 map / 6 hours</td>
</tr>
<tr>
<td></td>
<td>1 map / 12 hours</td>
</tr>
<tr>
<td></td>
<td>1 map / 24 hours</td>
</tr>
<tr>
<td>How important do you consider the fire behavior simulation?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>not important</td>
</tr>
<tr>
<td></td>
<td>quite important</td>
</tr>
<tr>
<td></td>
<td>necessary</td>
</tr>
<tr>
<td>What is the preferred time duration of fire simulations?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Up to 3 hours</td>
</tr>
<tr>
<td></td>
<td>Up to 3 hours</td>
</tr>
<tr>
<td></td>
<td>Up to 24 hours</td>
</tr>
<tr>
<td></td>
<td>Up to 3 days</td>
</tr>
<tr>
<td>What is the preferred time frame of historical fires?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 decade</td>
</tr>
<tr>
<td></td>
<td>1 decade</td>
</tr>
<tr>
<td></td>
<td>1 decade</td>
</tr>
<tr>
<td>For which fires are you interested to have access to burned / reforested areas?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>For fires above 10 hectares</td>
</tr>
<tr>
<td></td>
<td>For fires above 10 hectares</td>
</tr>
<tr>
<td></td>
<td>For fires above 100 hectares</td>
</tr>
<tr>
<td></td>
<td>For fires above 1000 hectares</td>
</tr>
<tr>
<td>How important do you consider the functionality of accessing historical fires based on the time/date of the fire ignition?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>not important</td>
</tr>
<tr>
<td></td>
<td>quite important</td>
</tr>
<tr>
<td></td>
<td>necessary</td>
</tr>
<tr>
<td>How important do you consider the functionality of accessing historical fires based on the size of the burned area?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>not important</td>
</tr>
<tr>
<td></td>
<td>quite important</td>
</tr>
<tr>
<td></td>
<td>necessary</td>
</tr>
<tr>
<td>Question</td>
<td>Options</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
</tbody>
</table>
| How important do you consider the recording of weather conditions from   | • not important  
| remote automatic weather stations?                                       | • quite important  
| • necessary                                                               |
| Which of the following weather forecast maps do you prefer to be        | • Maps of wind speed / wind direction  
| incorporated in the system as?                                           | • Maps of Wind speed / wind direction  
| • Maps of Air temperature                                                | • Maps of Relative humidity           |
| • Maps of Cloud precipitation                                            | • Maps of Fuel moisture 10-h          |
| What are the preferred measurement units of wind speed?                  | • m/s                                      |
| • km/h                                                                    | • Beaufort                                 |
| How important do you consider the functionality of retrieving historical | • not important  
| weather data recorded by the remote automatic weather stations?          | • quite important  
| • necessary                                                               |
| Which of the following fire management data do you consider necessary    | • Road network                           |
| for visualization purposes?                                              | • Water sources                          |
| • Vegetation and cover types                                             | • Evacuation areas                        |
| • High risk areas (gas stations, houses in urban areas etc.)             | • Satellite images                        |
| • Satellite images                                                       | • Aerial photography                      |
| • Annotation                                                             | • Annotation                               |
Which of the following services of fire risk analysis should be incorporated in the final system?

- Online Editing
- Real – time images from web cameras
- KML layers support
- Geo-location of fleet tracking
- Closest routes to water sources
- Drive times from a specific location
- Routing

Which of the following services should be incorporated in the mobile app?

- Tracking the current location of the user
- Closest routes to water sources
- Access to real-time weather measurements from the closest weather station
- Tracking the current location of the user
- Closest routes to water sources
- Access to real-time weather measurements from the closest weather station

**Non-functional Requirements**

| Efficiency: How important do you consider efficiency for the provided services? | not important | quite important | necessary |
| Availability: Which is the most preferable period in which the provided services will be available? | May-September | October-April |
| Usability: How important do you consider providing a manual / operation guide for each of the functionalities of the system? | not important | quite important | necessary |
| Security: How important do you consider accessing services only through credentials (user name & password)? | not important | quite important | necessary |
TOWARDS THE DESIGN OF TRUSTWORTHY WEBSITES FOR CULTURAL ORGANISATIONS:
A VISITOR'S PERSPECTIVE

Evangelia Kavakli, Cultural Informatics Laboratory, Department of Cultural Technology and Communication University of the Aegean, Mytilene, Greece, kavakli@ct.aegean.gr

Philemon Bantimaroudis, Department of Cultural Technology and Communication University of the Aegean, Mytilene, Greece, pbantima@ct.aegean.gr

Christos Kalloniatis, Cultural Informatics Laboratory, Department of Cultural Technology and Communication University of the Aegean, Mytilene, Greece, chkallon@aegean.gr

Stefanos Gritzalis, Information and Communication Systems Security Laboratory, Department of Information and Communications Systems Engineering, University of the Aegean, Samos, Greece, sgritz@aegean.gr

Konstantina Avgeri, Department of Cultural Technology and Communication University of the Aegean, University Hill, GR 81100, Mytilene, Greece, ctm12011@ct.aegean.gr

Abstract

Internet and the World Wide Web are increasingly playing an important role in facilitating communication in the cultural domain. Cultural institutions increasingly use the web in order to make their contents available on a more global scale thus broadening their audiences. In the online environment, trust emerges as a key asset for effective organizational communication. In this study, we assess the role of different factors that influence users’ trust toward museum websites. We adopt an exploratory approach in the form of a survey, regarding visitor’s choices in two small museums located on a Greek island. The results indicate that alongside technical considerations, perceived organizational reputation, a significant theme in management and marketing studies, as well as users’ needs and gratifications, investigated in the field of media theory, constitute significant social predictors of online trust. The proposed model serves as an exploratory tool for small, cultural organizations that strive to improve visitors’ trust toward their websites. By understanding the drivers influencing online trust, they will be able to provide more effective online experiences for their visitors and this is particularly important in the current environment of restricted funding.

Keywords: Trustworthy information systems, online services, cultural organisations, uses and gratifications theory, organizational reputation

1. INTRODUCTION

Information and Communication Technologies (ICTs) and especially the Internet increasingly play an important role in the cultural domain, facilitating cultural institutions such as museums, to make their contents available on a global scale. A number of studies indicate that museum websites are valuable tools for reaching out to audiences (Thomas and Carey, 2005; IMLS, 2008; MTM London, 2010). At the same time, the Internet has
transformed the expectations of online users who now expect that museum websites will provide the information they need, when they need it (Marty, 2008).

As museum websites become ubiquitous, attention is focused toward factors that impact online success (Finnis, Chan & Clements, 2014). In our previous work (Pavlidis et al., 2013) we have claimed that trust is an important parameter affecting the overall experience of online visitors. To this end, we have defined trust in museum websites as the positive expectation that the system will provide an intended function that will satisfy the user’s communication goals, deriving from the Uses and Gratifications Theory (Katz, 1995). In addition, we have proposed a conceptual trust model taking into consideration both objective, technical parameters such as website security and privacy protection, as well as social parameters related to users’ subjective perceptions relating to their online expertise or to the organization delivering the website.

In this paper, we conducted a survey with the participation of visitors of two museums located on a Greek island in order to assess the proposed trust model. The results of the survey indicate that basic model variables, namely, perceived reputation, the use of technology and personal needs as a core construct, constitute significant predictors of trust. As a result, they become essential for museums managers as they effectively design, promote and evaluate online content and services.

The paper is structured as follows. Section 2, discusses related work in the area. Section 3 describes the conceptual trust model. Section 4, presents the survey regarding visitor’s choices in two small museums located on a Greek island. Finally, section 5 concludes with a discussion on the findings of this survey and its potential implications.

2. RELATED WORK

A number of studies examine the use of museum websites from a user-centred perspective (Marty, 2008). They address usability and accessibility issues, or try to identify specific usage patterns in terms of the activities performed, duration, day and time of use, and so on. This type of analysis is facilitated by the use of powerful and easy-to-use web statistics tools. However, such studies focus mainly on ‘how’ online visitors use websites rather than on ‘why’. Understanding why museum audiences use museum websites is essential for museums managers as they effectively design, promote and evaluate online content and services.

To this end, researchers’ attention has shifted toward an initial understanding of online visitors’ motivations and how these motivations have an impact on the way people engage with the museum’s website (Haley Goldman & Schaller, 2004; Peacock & Brownbill, 2007; Fantoni, Stein & Bowman, 2012; Phippen, Sheppard, & Furnell, 2004). In a similar manner visitors’ motivation has been correlated with actual behaviour in the museum’s physical space (Ellenbogen, Falk & Haley Goldman, 2007).

A common assumption in these studies is that online visitors use museum websites to satisfy information needs. Thus, the definition of motivational categories of online visitors is based on information seeking activities, like gathering information for an upcoming visit or finding specific information for professional or personal interest. However, as websites shift from being information media to becoming communication media, with social media dominating the time that users spend online, the motivations of online visitors are expected to change substantially.

Shifting to social media generates a number of security and privacy issues out of the users’ perception when using traditional websites. Protection of security and privacy is very
important for shaping the system’s trustworthiness on the mind of the end user. However, besides the technical solutions that are usually applied on a system in order to satisfy users’ security and privacy requirements other key factors play an important role for elevating the system’s trustworthiness. The work presented in this paper aims to explore how individual user needs identified in the form of uses and gratifications are associated with users’ trust toward museum websites.

With regards to online trust, a number of studies have been conducted in order to identify the elements that communicate trust in e-commerce and e-government websites, mainly from a Human Computer Interaction (HCI) or systems security perspective (Corritorea, Krachera, & Wiedenbeckb, 2003; Bart, Shankar, Sultan & Urban, 2005; Yan & Holtmanns, 2007). In the context of cultural institutions, there is a growing body of work related to trustworthy digital repositories and the determinants of trustworthy information sharing (Prieto, 2009; Yakel, Faniel, Kriesberg & Yoon, 2013). This research is relevant as online shopping and information provision are applicable to museum websites; however it excludes fundamental aspects that motivate online visitors such as social interaction and learning.

Empirical studies across different web site categories have shown that online trust mediates the relationship between website characteristics and user intentions (Bart et al., 2005). In other words, a website is “trusted” because it meets the needs of the user communities for which they are designed (Prieto, 2009).

In the context of the current study, we explore how individual user needs identified in the form of uses and gratifications are associated with users’ trust toward museum websites. The literature on uses and gratification highly represented in media studies, has recently been applied in order to understand consumers’ attitude and intention of using the Internet in different scenarios mainly social networking and e-commerce (Papacharissi & Rubin, 2000; Kaye & Johnson, 2004; Ridings & Gefen, 2004; Raacke & Bonds-Raacke, 2008). However, few studies explore the relation between user needs and the formation of online trust or about the specific user needs and gratifications in the context of museums (Stylianou-Lambert, 2010; Gladysheva, Verboom & Arora, 2014).

3. **A Conceptual Trust Model for Museum Websites**

This paper focuses on online trust between people and museum websites. In this context, online trust can be defined as the positive expectation that the system will provide an intended function that will satisfy the users’ (communication) needs. Thus, a trust model examines the potential influence of technical website properties and user perceptions on the satisfaction of user needs, thereby influencing users’ formation of trust in museum online services.

The proposed model (shown in Figure 1), builds upon the ‘hard trust’ and ‘soft trust’ approach proposed in (Yan & Holtmanns, 2007). “Hard” trust builds upon technical system specifications based on objective regulations and standards. “Soft” trust considers trust based on users’ subjective perceptions relating to their online expertise or to the organization supporting the system. We claim that in the context of museum websites both aspects of trust are applicable as they coexist with one another in order to provide a trustworthy system.
A number of hard/technical trust parameters relate to web technology mainly security and privacy protection. Most system users (especially users that depend upon online systems to take advantage of specific services) feel more convenient when they are provided with assurances that their personal data they provide for the realization of a specific transaction are protected. Even the existence of a privacy policy usually appeases online users even though no evidence of specific types of usage is provided. Security deals with the integrity and confidentiality of data. Implementing security mechanisms for the user and data protection is of vital importance when a museum tries to build trust. In addition, website availability is another important security factor towards trust. Eligible users should always have access to e-services based on their access rights. Unjustified denial of service to eligible users lowers their trust. Reliability is also a key issue, as repeated visitation is related to satisfying users’ requests. Furthermore, usability is related to trust as users favor easy-to-use, easily accessible websites (Corritorea, Krachera & Wiedenbeckb, 2003; Prieto, 2009). Privacy also plays an important role. Issues like anonymous browsing, authentication with the use of pseudonyms and unlinkable transactions play an important role towards the implementation of trustworthy websites.

From a social perspective, trust relates to the user’s expectation of the cultural organization delivering the specific website, (i.e., the reputation of the museum). Organizational reputation reflects on the trustworthiness of a website affecting its functionality (Bedi & Banati, 2006). In this sense, museums’ reputation as trustworthy sources of information raise similar expectations with respect to the information provided in museum websites. In a similar manner, if museum visitors trust the museum to promote ethical behavior, this trust is also reflected on its website for example, in the context of expected personal data protection and privacy policies.

In the online world, reputation systems, also known as recommender systems, provide mechanisms for judging trustworthiness when users lack a personal history of past experience.
(Corritorea, Kracher & Wiedenbeckb, 2003). Familiarity on the other hand, is the stage where people rely on their previous interactions. Users’ familiarity with the website as well as previous experience with the Internet (Bart et al., 2005), are likely to have significant effects on website trust. Familiarity builds consistent expectations of a website and increase users’ confidence. Greater confidence is associated with reduced uncertainty and therefore may lead to greater trust toward a website.

4. **FORMULATING TRUST IN MUSEUM WEBSITES: A VISITOR SURVEY**

Due to the limited amount of literature on uses and gratifications theory and its relationship to online trust in the context of museum websites, an exploratory study was conducted. The study focused on the visitors’ perspective addressing two of the three perspectives discussed in the previous section, namely user needs and social trust issues.

The central objective of the study is to assess visitor’s trust toward the museums’ website/digital services, based on the museums’ perceived reputation as well as visitors’ needs. Therefore, the central research question of this study is as follows:

**RQ: To what extent perceived organizational reputation, visitor’s needs and familiarity with web technology influence visitors’ trust toward museums’ website services?**

To this end, a survey was conducted in the premises of the two most noteworthy museums of Lesvos – the Archaeological Museum and the Natural History Museum of Lesvos. The researchers posed exploratory questions in regards to visitor’s choices in small museums located on a Greek island. These two organizations located on the third largest island of Greece were deemed representative of small-to-medium size cultural organizations of the European periphery. Both museums are located in Lesvos and thereby they are a part of an island society, with a limited access to resources available in metropolitan areas and they are a part of a secluded economy. Both organizations employ less than 50 employees, their yearly visitation does not exceed 30,000 visitors and they are subject to seasonality factors, as their visitation peaks during the summer months. The survey was conducted in a non-random sample of 103 visitors during September 2013 and was distributed both in English and in Greek. A Likert scale from 1 to 5 where 1 signified minimal agreement and 5 reflected maximum agreement was utilized as a standard measurement device for surveys like this.

The questions that visitors answered provided evidence for the following variables:

1. Gender – a nominal scale, demographic variable, included in the survey to assess gender differences in relation to the dependent variable.

2. Age – an ordinal scale, demographic variable included in the survey to assess age differences in relation to the dependent variable.

3. Education – an ordinal scale, demographic variable, included in the survey to assess educational differences in relation to the dependent variable.

4. Previous Visitation – A control variable through which the researchers investigate whether a previous visit at the two museums contributes toward trusting their Internet services.

5. Long-Term Use of Website/On-line Services – A control variable through which the researchers attempt to assess whether familiarity with on-line services is related to trusting the museums’ Internet services.
(6) Organizational Reputation – A primary independent variable through which researchers assess a relationship between a museum’s perceived reputation/fame and visitors’ trust of its Internet services.

(7) Personal Needs – A primary independent variable through which researchers assess visitors’ personal needs as a factor that contributes toward their trusting of the museum’s Internet services. This is a key variable for assessing visitors’ needs in relation to trusting museum web services.

(8) Information/Knowledge – This variable represents one type of personal needs derived from the Uses and Gratifications literature as one of the primary factors that lead audiences to seek information. In this context, we assess whether it can be linked to visitors’ trust of museums’ Internet services.

(9) Entertainment/Escape – This variable represents one type of personal needs derived from the Uses and Gratifications literature as one of the primary factors that lead audiences to seek information. In this context, we assess whether it can be linked to visitors’ trust of museums’ Internet services.

(10) Social Relationships – This variable represents one type of personal needs derived from the Uses and Gratifications literature as one of the primary factors that lead audiences to seek information. In this context, we assess whether it can be linked to visitors’ trust of museums’ Internet services.

(11) Personal Upgrading – This variable represents one type of personal needs derived from the Uses and Gratifications literature as one of the primary factors that lead audiences to seek information. In this context, we assess whether it can be linked to visitors’ trust of museums’ Internet services.

(12) Previous Information about the Museum – A control variable through which the researchers attempt to assess whether previous knowledge or information about the two museums contributes toward trusting their Internet services.

Because of missing values traced in different cases, 96 of the questionnaires were analysed. A total of 12 questions were included in the questionnaire, while assessing three constructs related to the core question of this study: (1) Trust toward the museum’s website services, (2) the museum’s perceived reputation and (3) visitors’ needs derived from the Uses and Gratifications tradition. To improve the overall reliability of the questionnaire, it was pretested before its distribution. Based on respondents’ comments, certain items were reformulated and some were eliminated, in order to maximize the overall reliability of the designed instrument. Furthermore, to assess the overall reliability of the questionnaire, we measured Cronbach’s Alpha, which registered at acceptable levels (0.805). To analyse the data, both descriptive as well as inferential statistics were utilized. Frequencies and cross-tabulations and calculated as we attempted to answer our central questions. However, only statistically significant findings are reported.

5. FINDINGS

Our descriptive analysis yielded some interesting results. While dealing with the dependent variable – trust toward museum websites – visitors were asked: “When you visit the website of a cultural organization, do you trust it for the safe management of your personal data (or its privacy policy)? 61.5% (59 respondents) answered “yes” while 38.5% (37 respondents)
answered “no”. Although the majority of museum visitors indicate that they trust museum websites, there is a substantial percentage of respondents indicating their distrust toward museum websites. Approximately one third of respondents lean toward distrusting museum websites, a finding that merits additional attention.

Figure 2. Visitors’ ranking “reputation,” “web technology,” and “personal needs” in relation to trusting museums’ websites

Museum respondents were asked to rank according to their perceived importance the factors that influence trust toward museum websites (see Figure 2). Web technologies was ranked as the most important variable (45.3%), followed by the museum’s perceived reputation (33.7%) and the visitors’ uses and gratifications (21.1%).

<table>
<thead>
<tr>
<th>Evaluation of needs satisfied by museums</th>
<th>Not at all</th>
<th>Slightly</th>
<th>Moderately</th>
<th>Much</th>
<th>Greatly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categories of Needs and Gratifications</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge</td>
<td>16.7%</td>
<td>19.8%</td>
<td>32.3%</td>
<td>16.7%</td>
<td>14.6%</td>
</tr>
<tr>
<td>Entertainment</td>
<td>32.35%</td>
<td>27.1%</td>
<td>26%</td>
<td>14.6%</td>
<td>0%</td>
</tr>
<tr>
<td>Social Interaction</td>
<td>44.8%</td>
<td>27.1%</td>
<td>18.8%</td>
<td>9.4%</td>
<td>0%</td>
</tr>
<tr>
<td>Personal Upgrading</td>
<td>28.1%</td>
<td>22.9%</td>
<td>25%</td>
<td>18.8%</td>
<td>5.2%</td>
</tr>
</tbody>
</table>

N=96

Table 1. Museum Uses and Gratifications
In terms of specific uses and gratifications that museum visitors seek -- such as knowledge, entertainment, social interaction and personal upgrading, -- responses vary as expected (summarized in Table 1). Thereby knowledge and personal upgrading register as more important than entertainment and social interaction. Specifically, 31.3% of respondents indicated that museums satisfy or greatly satisfy their need for knowledge while 24% responded that museums satisfy or greatly satisfy their need for personal upgrading. In regards to entertainment and social interaction the respective percentages are 14.6% and 9.4%. This is an expected finding as museums are considered primarily information providers rather than entertainment centres.

As we cross-tabulated demographic variables, – namely gender, age and the level of education – with the visitors’ responses in relation to their trust toward websites, gender and age do not seem to influence other variables as no statistically significant differences were observed. On the other hand, the level of education seems to be related to visitors trusting their personal data in museum websites. Specifically, the more educated a person is the higher the trust he/she displays toward museum websites. Almost all respondents holding university or postgraduate degrees (91.5%) indicated that they trust museum websites. The respective percentage for respondents with no university degree is 8.5%. The Chi-Square result was statistically significant at the .05 level. Furthermore, university education is related to entertainment in the context of the museum visit. People with university degrees seem to perceive the museum experience as “very entertaining” (42.2%) compared to respondents without a university degree (14.3% for high school graduates). This is a finding that merits additional exploration as the level of education of museum visitors differentiates them from people with low education, in terms of the type of experience they acquire from their visit. Highly educated museum visitors seem to attribute an entertainment value to their museum visit, in addition to acquiring knowledge and information.

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE</th>
<th>Wald</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>-1.072</td>
<td>.594</td>
<td>3.257</td>
<td>.071</td>
</tr>
<tr>
<td>Age</td>
<td>-.140</td>
<td>.174</td>
<td>.649</td>
<td>.420</td>
</tr>
<tr>
<td>Education</td>
<td>1.149</td>
<td>.437</td>
<td>6.919</td>
<td>.009</td>
</tr>
<tr>
<td>Visited the organization in the past</td>
<td>-.855</td>
<td>.662</td>
<td>1.672</td>
<td>.196</td>
</tr>
<tr>
<td>Long-term use of websites</td>
<td>1.214</td>
<td>.581</td>
<td>4.360</td>
<td>.037</td>
</tr>
<tr>
<td>Organizational Reputation</td>
<td>1.593</td>
<td>.598</td>
<td>7.098</td>
<td>.008</td>
</tr>
<tr>
<td>Satisfaction of personal needs</td>
<td>1.703</td>
<td>.633</td>
<td>7.232</td>
<td>.007</td>
</tr>
<tr>
<td>Previous knowledge about the organization</td>
<td>.414</td>
<td>.303</td>
<td>1.859</td>
<td>.173</td>
</tr>
<tr>
<td>Constant</td>
<td>-6.308</td>
<td>2.063</td>
<td>9.347</td>
<td>.002</td>
</tr>
<tr>
<td>Cox &amp; Snell R Square = .390</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nagelkerke R Square = .529</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Logistic Regression: Predicting Trust toward Museums’ Websites
Finally, analysis regarding the relationship between the variables examined in the survey, indicates that the two core variables, namely reputation and personal needs constitute significant factors which seem to be related to trust (see Table 2).

6. CONCLUSION

Providing trustworthy information, both onsite and online, constitutes a key activity affecting the credibility of modern museums (Bomboy & Sherman, 2014). A widely cited survey commissioned by the American Association of Museums found that 87 percent of those surveyed saw museums as “as one of the most trustworthy sources of objective information” (Lake Snell Perry & Associates, 2001). In 2006, an Institute for Museum and Library Services survey of American adults concluded that “museums evoke consistent, extraordinary public trust among diverse adult users” (IMLS, 2008). Similar findings are reported in the surveys conducted by the Canadian Museums Association (Gerald, Muise & Northrup, 2009) and the British Museums Association (Britain Thinks, 2013). However, trust is not automatically reflected on the online content delivered at the museum website. By understanding the drivers influencing online trust, museums will be able to provide more effective online experiences for their visitors and this is particularly important in the current environment of restricted funding.

A number of studies dealing with trust in the context of online systems, have emphasized the technological aspects that potentially affect people’s trust – the reliability and efficiency of digital services, the quality of multimedia interface design and the precautions taken to ensure users’ security and privacy rights.

In this study, we have provided valuable feedback and justification towards the fact that according to the users’ perspective trust is not only influenced by technical factors. In particular, we considered two significant social variables that contribute toward visitors’ trust, namely organizational reputation, a significant theme in management and marketing studies as well as users’ needs, explored in the field of media theory. This is not by any means an exhaustive list of social predictors of trust. However, the study contributes toward an initial discussion that encourages interdisciplinary considerations in the context of online trust.

This survey is subject to certain limitations. We explore small samples of visitors in one island society. As island societies are unique environments displaying distinct market characteristics, we are hesitant to generalize those results to other markets. Future research should seek evidence from larger populations seeking to incorporate additional factors that might influence formulation of trust. In particular, the relationships between technical and social trust variables and how they collectively affect users’ trust towards online systems of cultural organizations requires further investigation and this is the subject matter of our current work.

References


LEARNING BY NEGOTIATION – IMPLEMENTING A JOURNAL MANAGEMENT SYSTEM

Ozgun Imre, Linköpings Universitet, Sweden, email: ozgun.imre@liu.se

Abstract

This study investigates how an international academic journal invests in an information system. Following the criticism that learning in an organisation is not conflict free the study sketches the decision making around the information system as a series of negotiations. The case shows that during these negotiations, the information system played the role of a boundary object that resulted in the editorial team understanding the system in different ways than previously. These negotiations paved the way for learning in the organisation, with the information system highlighting differences among the participants involved with the journal.

Keywords: Learning, Negotiation, Journal management system, Social world

1 INTRODUCTION

Recent years has seen a steady increase in academic outlets in variety of disciplines (Larsen & von Ins 2010), resulting in newly established academic journals that compete with existing journals. This competition is not an easy one: the new journals have to attract good papers/authors, be able to process these papers and add value to them, and then spread the word around of their existence and attract readers—three actions/processes that have taken the established journals years to hone.

Information technology (IT), needless to say, has played a role in this increase of outlets. It has enabled access to information and people, as well as opened up distribution channels that did not exist before. Thus it is not surprising that IT offers solutions to help these new journals in competing with the existing ones.

However, to be able to offer any help, the IT solutions has to be adopted by the organization. This issue is further aggravated, as there are no barriers for the established journals to use same solutions themselves. As Carr (2003) pointed out, IT would not automatically provide advantage to the organization, especially as the best practices are easily copied. For the implementation to be successful and bear this potential advantage, the organization has to go through a learning process.

This learning process is what motivates this paper. In a small international journal setting, the process of learning in a journal management system implementation is investigated. Following the arguments that organizational learning is not conflict free (see: Contu & Willmott, 2003), this paper sketches this learning process as a series of negotiations. By highlighting the challenges encountered by the editorial team and how they managed these challenges through these negotiations, the paper aims to provide an account of how an academic journal (hereafter Business Journal) makes investments in information systems (IS).

Following this brief introduction, the paper continues with section two, a short literature review on organizational learning, drawing mainly from IS perspective. Section three provides a theoretical framework for the study, followed by the research design in the next section. The case is presented and analysed in the two following sections, with section seven concluding the paper.
2 LEARNING AND IS

The issues related to learning and organizational learning are not new to the IS researchers. There are many potential benefits associated with IS implementations, often associated with benefits through standardization, automation and centralization. However, as noted by Orlikowski and Robey (1991) introduction of IS more often than not means an organizational change. It is these changes that Brynjolfsson and Hitt (2000) argue would result in gains from the IT investments, and to be able to cope with these changes, organizational learning is necessary (Robey & Sahay 1996; Gregor et al. 2006).

One obvious choice to learn is by formal learning. However, as Lave and Wenger (1991) argue, formal learning is just one way, and not necessarily how learning is done in practice. They argue that learning is situated in everyday practice and that learning takes place by observation and imitation, and by socializing within their communities. IS researchers have similarly recognized the value of learning by participation (Robey et al. 2000). As the review by Robey et al. (2000) shows, there are myriad of factors that are important in successful implementation, but as Wastell (1999) notes learning by participation is a essential for the success of IS development and implementation.

This learning process means involvement of different stakeholders. Within the developer-user context, the need for effective communication and understanding what the other party wants from the IS has been recognized for a long time (Boland Jr. 1978; Wastell 1999). The political use of emails as a “rebellion” against top managers resulted in the project of introducing emails in a university setting deemed as unsuccessful, showing that though learning took place, stakeholders did not view the system in similar fashion, and that organizational politics play a role in learning (Romm et al. 1996). From a soft system methodology strain, a similar issue is taken by Westelius (2006) in which a large Swedish organisation tries to implement web applications. Though the parties involved agree that the web applications are important and change occurs, how they view the change and what they expect from it differs: for example the local clubs resisting the headquarters, as they see the initiative as a tool for centralisation. Coupled with this issue of different stakeholders, and their different stakes, the automation provided by the IS, and the probability of this leading to de-skilled workforce and loss of knowledge is also a concern touched in the literature (see: Orlikowski and Barley (2001)).

The inclusion of the users to the development/implementation process might be a result of different ideals, as Bjerknes and Bratteteig (1995) argue: to acquire relevant knowledge; to reduce resistance and to increase work place democracy. However, similar to the criticism raised by Contu and Willmott (2003), Mengiste and Aanestad (2013) argue that though the issues related to learning and participation can be found in the literature, power relations and how they affect the learning is not taken into consideration often. As Romm et al. (1996) argue, even the mention of the power relations itself is usually taken as a detrimental factor. Following this brief literature review, following Mengiste and Aanestad (2013), this paper aims to bring the negotiations among the stakeholders and how the power relations were settled in this learning process the journal has gone through.

3 THEORETICAL UNDERPINNINGS

Similar to Mengiste and Aanestad (2013), social worlds provide a part of the theoretical lens for this research. Social world is a concept used by Strauss (1978) and later by Clarke (1991), and has seen use in various disciplines, with several examples in science and technology studies (for a review see: Clarke and Star, 2008). In Clarke’s (1991) argument, social worlds are “the fundamental building blocks of action” and “at least one primary activity (along with related activities), ... sites where activities occur ... [and] technology (inherited or innovative means of carrying out the social world’s activities)” (Strauss, 1978 cited in Clarke, 1991, p. 131). For social worlds there is no formal membership and they might be geographically
spread. What ties them, however, is effective communication (Clarke 1991). Social worlds are, in line with the action orientation found in works of Strauss, spaces of “perceptions and action” and “groups with shared commitments to certain activities, sharing resources of many kinds to achieve their goals, and building shared ideologies about how to go about building their business.” (Soeffner 1991, p.363)

Social worlds is closely related to the concept of arena: “a field of action and interaction among a potentially wide variety of collective entities” (Clarke 1991, p.128) and are characterized with clashes between the social worlds, a “means of battle” (Soeffner 1991). Arenas are where “conflict, competition, cooperation, exchange, and negotiation” take place, where a negotiated order is achieved. In these arenas, “various issues are debated, negotiated, fought out, forced and manipulated by representatives” of the participating social worlds (Strauss 1991, p.239). Negotiated order is a perspective to look at the interaction between the structure and the human agency. Following from Strauss’ work on mental hospitals, the social order is attained by negotiations among different actors where the members of the team had to hammer out their own “line of action in relation to the team as a whole and to the operational philosophy that came to predominate the ward” (Strauss et al. 1964, p.142). This order is a result of constant negotiations, and thus is saddled with occasional breakdowns, and to attain the order, the stakeholders have to engage with each other. Even rules, Strauss et al. (1964) argue are a result of on-going negotiations, even though they are usually conceptualized as exempt from the negotiation process. In such negotiations, the individuals represent their social worlds in the arena, and negotiate.

It is this focus on negotiation, and the order created as a result, that makes the arena/social world theory a useful tool to look at the implementation of the IS. Following Clarke’s (1991) adoption of boundary object to the social world/arena theory, one can conceptualize the IS as one. Boundary object is a concept to explain things that exist at the junctures where different social worlds meet in an arena of mutual concern (Star & Griesemer 1989). In its role as the boundary object IS acts as an “object that is translated to address the multiple needs or demands placed upon it by the different social worlds that “meet” around” (Clarke, 1991, 134), and acts, as Jonsson et al. (2009) argue more than just a mediator.

As can be seen, most of the concepts mentioned are rather similar to concepts used by other theories mobilised within IS research, among others Giddens’ structuration theory. Similar to structuration theory, social world/arena and negotiated order argue that structure and agents co-create each other, and should not be taken as separate. The agency of the individual is constrained by the structure, but the agents are still autonomous enough to affect the structure. Such similarities between these two theoretical lenses were noted by Carey (2002, p.202) who argued that the sociology promoted by social interactionism – Strauss being a member of this group – was “a sociology of structuration before Anthony Giddens invented the word”. However, as implied before, negotiated order is a dynamic way of looking into the changes and structure, and puts forward negotiations, and the social worlds involved, as the focus where the structuration is taking place. Thus negotiated order, and social worlds/arena tells where (negotiations) and what (social worlds) to look for (Barley 2008). As such, negotiated order and social worlds/arena theory should be taken as complementary to structuration theory, arguing that it is the situated behaviours that one should focus to observe the changes in structure and agency.

4 Setting and Approach

In this study, a single case study design was selected. As the aim of the study is to highlight the negotiations that took place in the learning process of an organization, the study has longitudinal elements that makes single case study design appropriate (Yin 2009). Another reason that motivated such a design stemmed from the fact that one of the editors of Business Journal works at the same institution as the author, providing ease of access to the case setting. This ease of access, as well as the aim to show how the negotiations took place,
further cemented the use of single case design, which as Siggelkow (2007) argues has the strong potential of displaying dynamic processes.

The study was conducted with an interpretivist approach (Walsham 1995; Walsham 2006), informed by the principles laid out by Klein and Myers (1999). As such the study makes no claims of being objective in the positivist sense. As the construction of meaning between the author and the editors resulted in several iterations of interpretation, the understanding of the case has changed for both parties. Similarly, the interpretation of the case has changed as new data was incorporated to the study. Both of these issues mirror two of the principles laid out by Klein and Myers (1999), and as such deviates from the objective research espoused in the positivist sense.

The data for the study was collected via qualitative semi-structured interviews, as well as observations. The easy access to one of the editors has made it possible to see how the editors engaged with each other as well as other stakeholders as challenges arose during work. More importantly, such access made it possible to see how the editor participated and negotiated with other parties in informal settings, for example in a coffee room talk. The notes taken during the observations were later on re-written and used for the analysis. For the interviews, three out of the five editors were interviewed, with the interviews averaging one hour. The interviews were digitally recorded and were summarised. Follow up interviews were conducted to further clarify some issues that came up during the interviews and observations, resulting in an iterative process of creating a full picture by using pars of the accounts provided. The follow up interviews also provided an opportunity for the editors to check how their accounts were interpreted by the author, giving them the chance of clarifying their accounts as well as to see the Business Journal in a new setting.

5 BUSINESS JOURNAL

Business Journal was founded 5 years ago by a group of colleagues that at the time were working at the same university. This group of would-be editors has decided to establish a journal that has studies of emerging markets as its main focus, with emphasis on empirically grounded papers. Situated in Business Studies and Economics, the Journal has been posed as a venue for studies from varying sectors and backgrounds, and was able to attract papers ranging from macro level studies – oil prices and stock exchange – to micro level studies – concerning value making in a journal.

The role of the IT/IS for the running of the Journal was not something that was consciously touched upon in a strategic manner when the Journal was founded. The role of the IT/IS – and how the editorial team treated it – has evolved through time. For the sake of this study, two episodes relating to the IT/IS decisions taken by the editorial team are presented. While these episodes might seem as two distinct points in time, the decisions taken at the first phase depend on previous decisions, just as the later episode depends upon the first. In the following parts, the decision about using a cloud system, as well as the idea of moving to an open source journal management system, are presented.

5.1 Growing with cloud

As mentioned, Business Journal was founded by a group of colleagues, who at the time were working at the same university. One of the main reasons for establishing the journal was a reaction against the established system. As the would-be editors were starting their academic careers, they felt that they were not getting the support that is needed from their environment. The system argued for publishing more, but at the same time it was set up in a way that favoured senior faculty that have already published papers. From another angle, a general dissent was found among the editorial team and their colleagues that the "gatekeepers" of the publishing industry were, for one reason or another, against them. One way to solve this issue was to set up their own journal.
In the early days, most of the tasks the editorial team did were supported by informal ways: having face-to-face meetings when there are problems, discussing it at one of the rooms, having a chat over coffee. The IT solution that was used predominantly at this time was emails. However, as the journal began to grow, the number of submissions to track increased. The editors were not able to see who was involved with a submitted paper, and what the progress of the paper was within the review system. The issues related to this knowledge management “mess” were further aggravated as the editors kept track of what they had done – and perhaps more importantly why they made a particular decision – mostly in an uncodified, implicit manner.

The unconscious adoption of email for Business Journal turned out not to be enough for the editorial team. Around this time, the cloud services, like Dropbox and Google Drive were gaining more visibility, and one of the editors put the issue on the table. If they would be using such a service, they would be able to see the process of the papers in “real time”, and see who is involved with the paper.

While the agreement to have a cloud service was easily achieved, this did not mean that the emails were to be scraped out. It was the main communications tool among the editors, as well as how they communicated with the other parties involved with the Journal. While the cloud service would be able to handle some communication among the editors, email was a familiar tool for all involved, and necessitated no maintenance on the user side.

Though the agreement was easily achieved, this did not mean that there was only one voice during the discussions. The editors were all familiar with using the office programme provided to them in their universities: should they pick a system where they put the documents in this form in like Dropbox, or should they adopt one that has these documents already within as with Google Drive, which included Google Documents?

This small divergence of ideas was expressed as having a full package, where everything can be settled by using the tools the system provided, versus using the tools the editors already used but store the resulting documents in the cloud.

Editor 1 commented in the interview that the choice of going for a cloud service perhaps was not the best decision, but at the time it was the only decision they could have. The editorial team was expecting that they would get some backing from the university, as they felt establishing Business Journal fell within the university’s call for increased internationalization and visibility. However, the response they got from the university was that it was not possible for the university to provide support.

This clash between the university and the editors, situated in different social worlds was one of the more formal points of negotiation. More recognizable for in the academia perhaps is the clash of the expectations that a faculty member has. A quote from Editor 1 highlights these tensions:

> I guess we fall to the same mistakes that we warn our students about. We saw it [cloud service] as a solution, but we didn’t think too much about it somehow. We knew it might not be enough but at the time it was what we did. Tight schedules, exams to correct, you know. You can not not do them, so you give up from your own time to keep the Journal alive. (Editor 1)

This quote highlights the pressures that arose from different social worlds that the editors belong to. On the one hand they are employed by their respective university/department, meaning that they need to fulfil some officially defined roles. On the other hand, the social world of being an editor of a newly established journal demands that the editors engage with the daily operations of the journal, thus clashing with the university related responsibilities. These responsibilities were further aggravated, as the editors recall, as all of them had to fill some roles – giving lectures, grading exams, supervising students – with very short notice.
In these early phases of Business Journal, IT/IS strategy was not fleshed out. However, as mentioned before, it became clear that the role that IT/IS played came to the forefront of the editors minds. The move from the emails – which was an established form of communicating – to use Google Drive necessitated handling different opinions: what should the spreadsheets contain as information; when is the appropriate time for uploading documents, i.e. copyright notice of the submission, the final draft?

The five colleagues take turns to be the “chief” editor for the volume/issue they are working with. This resulted in the editors playing different roles during the work process of the Journal. While the chief editor is responsible for fulfilling the editorial role, the other editors’ networks were used to find reviewers. In this case, who should contact the reviewer? Should the reviewer contact her friend or the chief editor? Who will keep track of these email conversations? Though all of them felt the frustration of trying to track down who has done what, the problems associated with this lack of knowledge sharing and knowledge management persisted.

Another stream was the issue of security of the cloud. While the editors would benefit from moving some part of the work involved to the cloud service, that would mean that an external party would be hosting the information gathered in the Journal. Aside from the issues related to downtime or loss of the stored documents, this meant that the author’s contributions have to be secured in such a way that there won’t be a breach of confidentiality. This resulted in selecting a well-known provider for the service, rather than the one offered by one of the universities that one of the editors was affiliated with.

These concerns led the editors to consciously engage how they operate, and what the concerns for various parties in the Journal’s work system could be. They had to negotiate among themselves to see how they could benefit from the system. The cloud system also made them realise that though the solution provided helped to solve some of the issues they faced, other issues persisted. In this manner, going from emails to the cloud enabled the editors to see the Journal from other social worlds – as a reviewer, as an editor, as a faculty member – and resulted in consolidation of what needs to be done for the next step. The use of spreadsheets and what information needed to be put there showed a reflection of the experiences of the editors as “editors” as well as “managers”. As the editing of the submissions progressed the more managerial tasks of keeping track of the process began to take precedence, resulting in the editors trying to balance their quest for good academic papers that fit Business Journal, with adopting work processes to a system so that they could follow the submission.

5.2   Aiming for maturity

As previously mentioned, the learning process is an on-going one, and as the Journal evolved, the need for a more capable IS became apparent. Though the cloud system complimented the email communication, the editors still kept most of their knowledge implicit. The new cloud system enabled them to see if the copyright notice was there, but that presupposed that somebody would put it there. Thus while using Google Drive has complemented the emails, it only offered ad-hoc support. This resulted in another round of negotiations about what to do, and the editors decided a more formal IS was needed.

The issue was put on the table after the editors realized that the cloud system was not providing much of a support to the team. Editor 1 related to this problem as a continuity issue. She argued that if one of the editors left the journal, the newcomers would have a hard time to decipher what is in the cloud service. She argued that especially with the voluntary nature of the work, and with all the other pressures they face in their work lives, they were not able to have a concrete way of establishing a system that would enable them to “pass on the torch” to newcomers. In her own words, “We needed to have a system that I can rely on and understand as a newcomer that hasn’t worked with this journal before” (Editor 1).
However, as mentioned by all the editors, the voluntary nature of the work, and the pressures exerted by other obligations, makes it hard to set up a system that will satisfy all the needs. Another colleague of the author has remarked that the situation is what she faces with her own research, where the studied projects are designed by today’s standards, which might become obsolete as soon as the users of the project changes. This concern about the future resulted in a more intense negotiation phase – this time more consciously done – about how to solve the situation.

One part of the negotiation was the formal request of the editors to have backing of their universities. They felt that if they had such backing it might provide them with more legitimacy. Being hosted by the university, even if the Journal is not recognised as affiliated with it, would make the readership wider. Similar ideas were expressed in earlier stages too, but they were not as consciously sketched out. However, when one of the editors met with the university representatives, the answer they got was not what they expected. The editors thought that by establishing the Journal they were implicitly helping the visibility of the universities they are employed at. Similar to this, they thought being an editor also fulfilled the call for increased internationalisation. The answer that was given to them to their request, however, was that since this was not a university project the university cannot back Business Journal.

This official response led to the editors asking for help from their respective IT departments. They had discussions about the systems that the editors had heard about, and if the IT personnel had an idea about what might be suitable. These brief negotiations helped the team to realise that they might be able to find a suitable solution, where the IT personnel’s knowledge would be used to match the systems to the requirements of the editors. Unfortunately, these helpful instances occurred only when the parties could arrange a time to meet. As the Journal was not a university project, the help the IT departments would give had to be on voluntary basis.

Learning from their experience with the cloud, and realising that just a change of IT/IS might not be enough, the editors also decided to sit down and discuss what the job entails, and what parts the IT/IS can help solve. This resulted, once again, in once again a series of brief negotiations among the editors. The process of shortlisting the journal management systems, the issues about simplicity, and how in the future the system would look like was discussed. The editors that had experience from using such systems for other work, e.g. as a conference chair, were more willing to adopt a system they thought would benefit them and have the functionalities needed. This was countered by the idea that this might mean increased financial costs: the already existing set-up was near free, so any additional system of this magnitude would mean significant increase. As Editor 2 puts it: “In the end we had several options – some matched us better than the others, but if you don’t have the backing of the university, and if you don’t expect the authors to pay for the cost of using such a system, you actually don’t have that many options.”

This negotiation phase resulted in the editors scrutinising how they actually work. Though the existing system was not that intuitive to help them, the team has realised that even the best system would not perform perfectly unless the participants came to an understanding about how to use the system. By arguing over what is needed in their IS the editors entered a phase of learning about publishing and knowledge management. As editor 2 mentioned, “It takes time to get a grip about what you need. You know it [publishing process], but you don’t know it until you are in it. We had to learn the tricks to ease the process.” By forcing the editors to express their wants and needs, and to see the constraints in place – like an unwilling university administration and short term demands posed by teacher roles – the IS played the role of a boundary object that facilitated learning.

Similarly, the IS played another role in pushing the editors to think about the competition. Some time ago, one of the editors has realised that a new journal with a very similar name, and similar scope and aim had been established. This increased competition was introduced to
the social world of the editors and made them aware that now they had to think about how the IS would be seen from the outsiders. By employing some of the journal management systems, they would be able to have a system that would be used not for the internal demands of the editors and reviewers, but would also be able to present a unified, corporate face to the authors and readers. This necessity to show a corporate face was similar to the earlier episode where the editorial role was contrasted with the managerial role – perhaps an ownership role, as the editors’ account showed that they are emotionally engaged with the Journal.

6 A BRIEF ANALYSIS

Following the case description above, this section revisits the case and highlights the issues using the theoretical lens of social world/arena theory. In both of the episodes, the IT/IS, at first a cloud service and later a journal management system, has played the role of a boundary object. In both episodes, these objects resulted in the editorial team being engaged with negotiations, among the team itself, as well as with other external parties.

One such external party was the university that the editors wanted to have as backer of the Journal. The idea that the university might be willing to help them with setting up a journal management system, and providing them with a server, however was shot down quickly. The editors believed that the Journal fits to the University’s strategy, whereas the University argued that the Journal was not a university project. This clash of interest to some extent stemmed from the question of ownership, as the editors did not want to let the Journal be managed by the University, whereas University did not want to commit to a project that it can not influence. This situation can be taken as a power struggle: the editors asking the University a second time, after the Journal solidified itself further can be seen as a tactic to offset the power struggle they faced before, and the University holding its ground. This clash of social worlds resulted in the editorial team taking the matters in their own hands.

This rejection itself has further triggered another set of negotiations, in which the editorial team asked help from the IT departments, to get an idea of what to expect from a journal management system, and to feel out if they would be willing to help if the need arises. This in turn resulted in a mutual learning process of both the IT department and the editors of what needs to be in a system, as well as what can go wrong.

As the help from the IT department would be limited due to the rejection of the University to provide support, this part of the negotiations were rather informal. A more formal negotiation occurred at the same time among the editors. As they were also providing their time voluntarily – and financing the existing systems – there was the issues of how to finance a new system. The shortlisting of the systems that was helped by the IT department – as well as the informal networks the editors used – was negotiated over from the financing issue, with some editors showing reluctance to a new system. A direct result of this discussion was the decision to look at open source solutions for the journal management system.

Through these negotiations, the editorial team not only learning about the IT/IS component of their work system, but also learned about the whole process of publishing. As the case description highlighted, the IT/IS has played the role of the boundary object, through which the editors had to negotiate their different demands due to belonging different social worlds. The experience they had as an author and reviewer for other venues made it possible for them to see the issue at hand from two different but connected sub-social worlds present in the social world around publishing.

Just as this negotiation phase was going on, similar to the ownership dilemma between university and editorial team, the clash of the editors social world and the managers social world necessitated them to map out what their work process is like and create a requirement list, rather than opting for ad-hoc solutions. Though editors, acting from the social world of editor were more interested in attracting good quality papers and the publishing process around it, when they were put into the manager position, they had to see the situation from
another point of view. The need for rationalisation and standardisation, so that a common ground for knowledge sharing can be built upon was realised, not for the ease of operations, but for the future survival of the journal.

As these negotiations shown, the understanding of the editorial team of both the IT/IS and the Journal and its environment itself have changed over time. This learning process, as some of the earlier literature about learning in organisations has not been a smooth process, but has involved multiple parties, coming from different social worlds. The ability – and the necessity – of the editors to shift between different social worlds have highlighted how this process is not a homogeneous, linear one, but rather a fragmented one where the decisions are negotiated. Similar to what is discussed in the social world/arena theory and negotiated order, such decisions are temporally limited, when consequences change, a new order has to be negotiated. One example of such change occurred when one of the editors has realised that competition was increasing, and they needed to have a corporate front to be counted as a legitimate journal. Clashing with the structure that the editorial can not change, the pressures were incorporated to their discussions, and the negotiations had to include a new set of criteria that the IS needs to fulfil.

The negotiations between different social worlds are briefly depicted in Figure 1 below. The figure for simplicity purposes depicts different sub-worlds as distinct from each other. However, as noted before an individual belongs to different social worlds, e.g. the editor can also be a member of university administration, as well as a reviewer. Similarly, for easier reading, boundary objects are connected to the social worlds with lines, whereas they cross boundaries. Several social worlds were only implicated in the negotiations, as shown with dotted lines in connection to the boundary objects, depicted as clouds. Some issues raised during the negotiations and some learning results are depicted in rectangles.

Figure 1. Social worlds and negotiated order of Business Journal

This paper has followed the stream of research that criticized the conventional understanding of learning as a smooth process, i.e. Contu and Willmott (2003), and followed the recommendation of Mengiste and Aanestad (2013) to use social worlds/arena theory as a theoretical lens. The case presented however, differed from their case, as Business Journal has included fewer participants in their work system, and the negotiations took place in a
microcosm, as opposed to the hierarchical national level agencies found in their research. In this case, there were power structures. However, aside from the University exerting their official power, most of the power struggles remained subtle. This to some extent has been a result of the editors being involved with similar sub-worlds – as an author, reviewer, and editor – thus enabling them to shift between their social worlds swiftly. However, this does not mean the process was conflict-free: when issues related to adopting a system that necessitates further financing, the managerial and editorial views have clashed. Even though all these people have worked together, such differences necessitated them to look at their own work practices, not just the IT/IS components that they intended to discuss.

The tensions shown in a small case such as presented here shows that the issues related to learning in an organization is still not a conflict free area. Though the editors all had very similar goals and to some extent background, their perception of the challenges, and thus how to solve them resulted in a series of negotiations. By playing the role of boundary object the IT/IS has made those differences visible to the editorial team, forcing them to contemplate the issues from other angles to survive, thus facilitation iterative learning cycles with every new negotiation.

7 CONCLUSIONS

This study has aimed to and provide an account on how an academic journal invests in an IS, by sketching how the learning in the organization took place as the editors contemplated their IT/IS solutions. In line with the criticism raised by Contu and Willmott (2003), and following Mengiste and Aanestad (2013), the learning process was conceptualized as a series of negotiations, where the end result is a semi-stable order. In such a scenario, going from IT solutions that provided ad-hoc solutions, to a more formal IS similar to a case management system, IT/IS has played the role of a boundary object upon which the actors involved with the Business journal has negotiated their interests.

The case has shown that even though all the editors come from similar background and similar aims, they had different roles during the investment process, shifting from the editor to a manager role, as well as seeing the journal from the perspective of a reviewer or an author. External actors has also shaped the decision process, necessitating a series of negotiations among the editors, University and IT department, through which the editorial team increased their understanding of the IT/IS as well as the Journal and publishing process itself.

Though in Mengiste and Aanestad (2013), the object of the study and the involved parties were on a macro-level, the case has shown that learning in a relatively homogenous organization is not a conflict free process. Conversely, though subtler than the power struggles than noted by them, the case here has shown that power struggles play a role in learning and should be taken into account.

Some implications can be drawn from the case. Adopting social worlds/arena and negotiated order can provide a situated analysis of IS development and implementation in organisations, showing the dynamism how the structure and agency co-create each other. Using negotiations and social worlds, the role of IT/IS as not merely mediating communication but also making the differences that were not visible or spoken about before better understood.

Continuing this strain, the case has highlighted how the IT/IS has played the role of a boundary object. However, of notice is that for the object to play this role, it has to be brought to the table and negotiated over. The case has shown that negotiations are the instantiation of this boundary object role, where the IS’s potential to reveal and bring together the different perspectives occur. This can be taken as an indicator that further research can be focused on the negotiations to study change and IS as boundary objects.
Considering how the issues raised by the editors closely resemble factors for adopting other IS systems, e.g. accommodating increased system capacity, creating a formal face, elimination of delay in ERP literature (Markus & Tanis 2000), use of social worlds/arena and negotiated order can provide a useful lens to look at adoption and implementation processes in general. As mentioned, in its role as boundary object, the IS made the differences among stakeholders visible. This can be taken as a practical implication, highlighting how even in a homogenous group – the editorial team – differences due to different social worlds are present. Identifying the possible stakeholders, and their social worlds would help in anticipating possible scenarios in negotiations, thus easing an organisational change process. By ironing out the differences to the extent that a workable solution can be agreed upon, the wider resistance to organisational change can be mediated, highlighting the need for being open to negotiations and willing to listen to others. From another perspective, negotiations can also highlight the stakeholders that are suppressed and further problematize why such a situation occurs and how this unbalanced representation can be solved.

ACKNOWLEDGEMENT

This research has been financially supported by the Swedish Research School of Management and IT. I am grateful for the comments of the anonymous reviewers, and participants of the EIS seminar at Linköping University.

REFERENCES


ORGANISATIONAL ISSUES INVOLVED IN PROJECT PORTFOLIO MANAGEMENT IN TERMS OF ORGANISATIONAL STATE-TRANSITION APPROACH

Hamed Sarbazhosseini, University of Canberra, Hamed.Sarbazhosseini@Canberra.edu.au
Amy Hart, Australian Catholic University, Amy.Hart@Outlook.com

Abstract

Managing project portfolios has become significantly important for organisations. In addition, interest in project portfolio management (PPM) has increased amongst both academics and practitioners. Project portfolio management is defined as a decision-making process that projects and/or programs are evaluated, selected and prioritised, and resources are allocated to achieve strategic goals. This research aims to understand the issues involved in PPM in terms of Organisational State-Transition Approach (OSTA). OSTA has been defined as a well-tried and reliable method to understand PPM’s complexity. This paper discusses the findings from part of a PhD study which has used OSTA in PPM. However, the finding presented in this paper focuses on organisational issues that are explored through case study research from seven Australian cases which were reviewed.

Keywords: Project Portfolio Management, Organisational State-Transition Approach, Organisational issues.

1 INTRODUCTION

For organisations to deliver better outcomes, managing concurrent projects is essential. Project Portfolio Management is discussed as a regime to manage multiple projects to gain benefits from projects and programs.

The academic review of PPM indicates there is a lack of clarity and understanding in the field. This is surprising while the number of PPM publications continue to increase (Killen et al. 2007; Arto and Dietrich 2004; Patanakul and Milosevic 2009). These publications were not able to clearly identify the value of PPM for organisations. Instead, PPM publications tend to emphasise the development of new methods (Iamratanakul and Milosevic 2007; Milosevic 2003), mathematical formulas (Iamratanakul et al. 2008; Dickinson et al. 2001; Kremmel et al. 2011), processes (Jeffery and Leliveld 2004; Cao et al. 2006) and frameworks (Archer and Ghasemzadeh 1999; Reyck et al. 2005) to implement PPM (Jonas, 2010). In addition, the number of conferences focused on PPM has significantly increased.

The widely cited Standish Group (CHAOS, 2013) report that in 2013, 43% of projects were challenged (late, over budget, and/or with less than the required features and functions) and 18% failed (cancelled prior to completion or delivered and never used). While these numbers represent an improvement in the success rates when compared with the results from previous studies, it still shows that over 60% of projects are not delivering at all or not delivering what it was promised. Continuing project failure rates suggest that theories of project management may not be complete, and PPM perhaps is part of the solution.

While there is a lack of theories to explore the use of PPM, McDonald and Sarbazhosseini (2013) discussed Organisational State-Transition Approach (OSTA) as a possible framework for a better structure and understand PPM. OSTA’s on-going and dynamic characteristic makes it useful to understand PPM’s theory.

This paper is part of PhD research that studied how organisations considered PPM from three main states of OSTA’s perspective namely: state indicators, desired states and transition actions. This paper highlights the explored issues in seven organisations studied during the PhD research.
The OSTA states were studied from three main perspectives in the PhD study (see Table 1 below); however, this research has only focused on identifying the issues that organisations faced in project portfolio management and this discusses the findings from state indicators in organisations as highlighted below.

<table>
<thead>
<tr>
<th>Research Questions</th>
<th>OSTA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What are the current situations of PPM literature; including issues and scope?</td>
<td>State Indicators</td>
</tr>
<tr>
<td>2. What is the desired state in PPM?</td>
<td>Desired States/Goals</td>
</tr>
<tr>
<td>3. What are the actions, methods, and tools in order to move from state indicator to a desired state?</td>
<td>Transition Actions</td>
</tr>
</tbody>
</table>

Table 1. The highlighted research question for this paper

2 PROJECT PORTFOLIO MANAGEMENT

Briefly, PPM is defined as “a collection of projects and/or programs and other work, that are grouped together to facilitate effective management of that work to meet strategic business needs” (PMI, 2006). The literature indicates that there is an increasing focus on establishing and improving PPM in organisations. This interest is related to an intensified emphasis on improving PPM methods and processes and the trend towards a more strategic perspective for the management of projects (Killen and Hunt 2012).

PPM is a “dynamic decision-making process in which new projects and programs are evaluated, selected and prioritised and balanced in the context of the existing projects and programs within the portfolio” (Cooper et al. 1999).

According to Killen and Hunt (2010), PPM includes policies, practices, activities, procedures, methods, and tools that assist ongoing resource allocation and reallocation among a portfolio of projects to maximise the success of the organisation.

Portfolio theory was implemented into ICT management sphere in 1952 with the focus on financial investments (Markowitz, 1952). Portfolio theory was first adapted to IT projects paving the way for Modern Project Portfolio Management (PPM) literature (McFarlan, 1981). In the widely read book, Third Generation of R&D (Cooper et al. 1999), predicted in 1991 that portfolio analysis and planning will “grow in the 1990s to become the powerful tool that business portfolio planning became in the 1970s and 1980s”.

The purpose of this original formulation of portfolio theory was to determine the particular mix of investments to maximise returns at a given level of risk for the investor (Reyck et al. 2005). Most research on portfolio management of the project was published from the mid-1990s onwards and focused on IT and ICT projects.

Organisations are attracted to PPM in their ICT and IT projects because of the claims made for it. Kersten and Verhoeof (2003) and Verhoeof (2002) argue that firms reduce IT (Information Technology) spending by 10 to 40 percent using PPM. Laslo (2010) claims that PPM allows an organisation to maintain agility while avoiding wasteful investments and Thorp (1999) argues PPM techniques are fundamental to getting value from IT projects. According to Rongzeng et al. (2005), banks considering ways to cut costs and eliminate waste in IT expenditure are advised to start taking a portfolio management approach.
2.1 **Main Resources of PPM Literature**

Project portfolio management’s literature has been categorised into three major literatures to better understand its theory. The academic literature discusses definitions, goals and actions in PPM. The industry literature explains standards and processes that managers should follow to implement PPM. Finally, the software literature that explains what software often can offer for their customers. Analysing these literatures help to clarify PPM research.

The academic literature of PPM indicates that PPM has three primary goals including, 1) aligning with strategic goals, 2) maximising value of the portfolio, and 3) achieving a balanced portfolio.

The industry literature can be found among PPM industry standards. There is a substantial amount of ‘grey’ literature regarding PPM. Industry standards perhaps provide the most reliable information about the practice. In PPM, the two major standards are the Project Management Institute (PMI, 2008), and the UK’s Office of Government Commerce (OGC, 2012). These standards define processes and tools in detail and provide guidelines and support to organisations in the application of portfolio management practices. In addition, the OGC has introduced a P3M3 (Portfolio, Program, and Project Management Maturity Models) in which organisations can assess their portfolio, program and project management and can take actions accordingly.

According to PPM’s software literature (Sarbazhosseini and Young 2012), software developers offer better customer satisfaction, taking timely actions, and increasing competitive advantages. Reviewing software literature shows that there are some differences in PPM goals from academic perspective and industry.

3 **Organisational State-Transition Approach**

As a domain of study, PPM is characterised by undisciplined and diverse opinion as to its industry approach to management standards, business literature and software support. In this paper, OSTA is used in PPM to add a clear and coherent structure to the field.

It is the ongoing and continuous nature of PPM that makes the State-Transition Approach (STA) useful in looking at the organisational state, change and responsiveness. Because of its dynamic advantage, the State-Transition Model (STM) is being used in Clinical (Siebert et al. 2012) and Ecological (Stringham, 2003) applications in order to analyse the state of the ecosystem and to examine patient’s status in different medical time frames. It also has application in software engineering where it is used in Unified Modelling Language (UML) and state charts (UML, 1997).

OSTA takes a view of the whole organisation as a system, one that is ongoing and integrated.

*The issues for OSTA as a way of describing systems and their change are (a) to determine and represent the current state of complex systems, (b) to express desired states in terms of systems indicators and (c) to frame action that changes the system in a way that it produces the desired change to state indicators without damaging other states of the system. (McDonald and Sarbazhosseini 2013)*

General form of OSTA can be seen as below Figure 1.

![General form of OSTA](image)
As an example of an OSTA, consider the Capability-Maturity Model (CMMI, 2010). This model conceptualises and defines a series of five capability levels (states) and the kinds of actions (transitions) that can lead to changes in that state. According to McDonald and Sarbazhosseini (2013), example of the early states is:

Initial: The ... process is characterised as ad hoc, and occasionally even chaotic. Few processes are defined, and success depends on individual effort.

Repeatable: Basic project management processes are established to track cost, schedule, and functionality. The necessary process discipline is in place to repeat earlier successes on projects with similar applications. (Paulk et al. 1993, pp. 8-9)

and the kinds of transitions that might be taken to move from 1) to 2) include instituting formal process proposal assessment such as explicit resource allocation, and responsibility specification.

3.1 State Indicators in PPM Literature

According to PPM literature, there are number of reasons for applying PPM concepts, as there are:

- Too many active projects
- Projects not linked to strategic goals
- Wrong projects
- An unbalanced portfolio (Kendall and Rollins 2003)

In addition to above challenges, there are other issues discussed as below:

- Too many projects in the pipeline
- High degree of conflict over existing resources
- Slowing project progress
- Reducing successful competition rate (Morris and Pinto 2007)

However, there is lack of evidence and research on issues involved in PPM theory. The present research study explores the major issues from organisational perspective in PPM.

4 Research Methodology

Selecting an appropriate research methodology is essential for any research. For purpose of this research, the case study methodology has been chosen. The scope of the case study is, according to Yin (2009, p. 18),

1. A case study is an empirical inquiry that
   1.1 Investigates a contemporary phenomenon in depth and within its real-life context, especially when,
   1.2 The boundaries between phenomenon and context are not clearly evident.

2. The case study inquiry
   2.1 Copes with the technically distinctive situation in which there will be many more variables of interest than data points, and as one result.
   2.2 Relies on multiple sources of evidence, with data needing to converge in a triangulating fashion, and as another result.
   2.3 Benefits from the prior development of theoretical propositions to guide data collection and analysis.
Validity and reliability of data were examined in different processes and protocols to make sure that the data analyses are valid and reliable. During this research, seven cases were studied in five months. The interviews were conducted with portfolio management offices. The documents were reviewed to better understand the context of the organisations and their scope from applying PPM.

The interview questions were designed based on OSTA’s three main states. Data collected from documents and interviews were analysed from those three main states discussed in Section 3.

The findings in the next section were analysed based on different data coding techniques used in the PhD research study namely: structural coding and thematic coding.

Structural coding is a question-based code that “acts as a labelling and indexing device, allowing researchers to quickly access data likely to be relevant to a particular analysis from a larger data set” (Namey et al. 2008, p. 141). Saldana (2012, p. 67) discusses structural coding as more suitable for interview transcripts than other data.

During the structural coding of data, it was discovered that there were possible new categories other than the categories mentioned in OSTA. For example, these themes were identified during structural coding of the data: Challenges and Issues, PPM Success Measurement, and PPM Benefits.

Following the structure of data coding, the thematic analysis was used in order to identify major themes highlighting issues and challenges involved in organisations. According to Saldana (2012, p. 139), a theme is “an outcome of coding, categorisation, and analytic reflection, not something that is, in itself coded.” Therefore, the primary findings from structural coding were used to identify major issues in PPM. In addition, the analysis of the thematic coding explored the most common and significant themes from organisation’s perspective with considering OSTA’s states.

5 ORGANISATIONAL ISSUES IN PPM IN TERMS OF OSTA

This section presents the findings from use of OSTA in PPM in a state indicator (illustrated in Table 1). Throughout the data analysis, it was found that organisations are facing different challenges and issues. The following significant issues and challenges are discussed commonly amongst all cases.

5.1 Lower Visibility and Transparency

Organisations commonly had lower visibility in their organisation. Therefore, it is significant for organisations to have better visibility in order to reduce risks. Organisation can clearly see the interdependencies among projects and, ideally, be able to identify and mitigate risks in the early stages so that there are fewer unplanned disappointments or setbacks. It was discovered that they apply PPM because they do not want to have surprises in meeting objectives.

With a transparent flow of information, portfolio managers can see how managers are operating projects and what progress is being made. Due to complex interdependencies between projects in a portfolio, it is significantly important for organisations to aim for a higher transparency across projects. One of the cases discussed that PPM can help their organisation to better structure their process in order to have efficient information sharing.

5.2 Challenges with Stakeholder Management

A concerning issue that was identified during this research and from each of the cases was the difficulty in managing stakeholders. It was discussed in numerous cases that because the organisations operate in a top-down system, the participation from all stakeholders was difficult. Managing project portfolios is a process that requires a lot of people to work together and collaborate in the best way possible. From the top down, policy and procedures need to be followed in order for portfolios to be managed successfully.

For example, the senior manager’s engagement was raised as being a significant issue. It was mentioned that senior managers have to take responsibility for delivering the portfolio of projects.
Senior managers, by leading and managing portfolios, can give confidence to all stakeholders that the actions they are taking for achievement of portfolio goals are necessary.

5.3 Challenges with Budget Cuts

A significant matter of concern for Government cases was the act of cutting budgets. It was no surprise to hear that budget cuts have an impact on managing projects in a portfolio. The outcome of budget cuts means that decision makers are required to assign priority to certain projects meaning that projects may need to be postponed or ceased. While projects are defined for one or few strategic goal/s, killing one project might cause a domino effect and impact negatively on other projects.

The cases described that the reality that when budget cuts occur projects can change, or new projects can be added to a portfolio. They mentioned that the difficulty with this is that policies and procedures do not always permit organisations to define a new project.

5.4 Challenges with Change Management

Along with all the elements of managing a portfolio, change management or a change in the scope of projects was considered a challenge for all organisations. Change of scope was one of the issues that organisations do not manage well. Similarly decision makers, as a result, need to make complicated changes either by changing, adding or, in some cases, ending a project to fit in line with the change in scope. It was mentioned by a case that “quite often, people and the business changes were their major problem rather than the delivery of projects or programs”.

5.5 Cultural Issues

A common challenge identified throughout the cases was the need to create and maintain a positive culture in the project management process. An issue arises when there is a need for change leading from either a new processes or a change in the scope of projects. With any of these changes, there is a need to obtain full support from senior management. If this support is not apparent, cultural issues arise. The flow on from poor management in any of the identified themes such as poor stakeholder engagement, low consistency (not following procedures), budget cuts and poor change management, was considered to lead to significant cultural issues. Cultural issues have a direct impact on portfolio achievements.

6 STATE INDICATORS IDENTIFIED IN EMPIRICAL STUDY

The state indicators represent current state of the organisations in this study. Issues in PPM are explored and an explanation of the situations is discussed in above section. This section discusses the above findings.

It was surprising to find that indicators differed to those mentioned by Kendall and Rollins (2003) in academic literature of PPM who identified the main reasons for implementing PPM being “too many active projects, projects not linked to strategic goals, wrong projects and an unbalanced portfolio.”

It was discovered in the PPM literature that organisations should apply PPM when there are multiple projects, unlinked strategic goals, unbalanced portfolios and wrong projects. The studied cases, however, showed that organisations apply PPM in order to gain a framework throughout the organisation. With having the framework, they would be able to have better visibility, accountability and consistency throughout the portfolio.

The study shows that organisations apply PPM when they have lower visibility and transparency. These new issues have never been previously discussed in PPM theory; however, they have a significant role in managing portfolios. The empirical study discovered that in order to manage their portfolios efficiently organisations they needed to consider these new criteria. The study demonstrates that these criteria play an important role in PPM.
7 CONCLUSION

This paper has presented findings from part of a PhD study which applied OSTA in PPM to better frame it and to clearly understand its complexity. OSTA’s ongoing characteristics made it useful in understanding issues and challenges involved in PPM theory. Seven cases where reviewed for this purpose. It is found that PPM is complex and there are so many interdependencies in this theory. Understanding issues and challenges are essential for any organisations to better manage their portfolio which leads to success.

Throughout this research, it was evident that organisations have lack of visibility and transparency of information. Therefore, it was commonly discussed that organisations require a better visibility and transparency of information to better manage portfolios of projects and/or programs. In addition, this research showed that organisations faced some challenges including stakeholder management, budget cuts, change management, and cultural issues. Not appropriate planning for managing stakeholders led to cultural issues. Furthermore, budget cuts resulted in managing changes which also created concerns for stakeholders and finally led to cultural issues in organisation.

Organisations are invited to apply OSTA in their PPM processes to better manage their decision-making processes in ICT and IT projects to achieve a better outcomes and to better manage their portfolios.

There are several potential opportunities for future work that could advance the knowledge of PPM theory and OSTA. Issues identified in this research can be further study to better frame PPM and deal with challenges. The OSTA can be applied in to different environments to better understand OSTA practices.

Acknowledgments

Hamed Sarbazhosseini & Amy Hart © 2015. This paper represents the findings of a Hamed’s PhD research project.

References


Kendall, G. and Rollins, S., 2003. “Advanced project portfolio management and the PMO:
multiplying ROI at warp speed”, J. Ross Publishing.
technology”, McGraw-Hill.
Security and Privacy Protection for Information Systems

Chairs: Dr. Tsohou Aggeliki
Table of Contents

Track – Security and Privacy Protection for Information Systems

2. [PID-41] Ryma Abassi and Sihem Guemara El Fatmi, “COUNTERING THE COLLUSION ATTACK IN A TRUST-BASED MANET”
4. [PID-68] Ehinome Ikhalia and Alan Serrano, “A FRAMEWORK FOR DESIGNING AN EFFECTIVE SECURITY AWARENESS SYSTEM FOR ONLINE SOCIAL NETWORK USERS”
5. [PID-74] Norilyz Figueroa and Ramakrishna Ayyagari, “TRAINING USERS ON INFORMATION SECURITY: EVIDENCE FROM JAVA APPLETS”
ON THE LOW DIFFUSION OF PRIVACY ENHANCING TECHNOLOGIES IN SOCIAL NETWORKING: RESULTS OF AN EMPIRICAL INVESTIGATION

Vemou Konstantina. Department of Information and Communication Systems Engineering, University of the Aegean, Greece. kvemou@aegean.gr

Mousa Gavriela. Department of Information and Communication Systems Engineering, University of the Aegean, Greece. icsdm12030@icsd.aegean.gr

Karyda Maria. Department of Information and Communication Systems Engineering, University of the Aegean, Greece. mka@aegean.gr

Abstract

This paper discusses the low adoption of PETs among SNS users, based on the results of an empirical investigation among users of social networking services. 170 members of 5 popular social networks provided information on how they protect their privacy, as well as on the most important factors guiding their decision to use privacy preserving tools or not. Research findings suggest that awareness of PETs is still low among social network users and that quality, effectiveness, cost and ease of use are critical factors influencing PETs adoption. A small number of users was also found not to employ any PETs, despite the fact that they reported being familiar with some of them. This paper enhances our understanding of PETs diffusion from the perspective of users and argues that usability aspects need to guide their design and implementation.

Keywords: Social Network Services, Privacy-Enhancing Technologies, adoption, usability.

1 INTRODUCTION

Social Networking Services (SNS) are constantly gaining popularity, with four of them (Facebook, Twitter, LinkedIn and Pinterest) reaching the top 15 most visited websites, in February 2015 (Ebiz/MbaA 2015). Facebook is estimated to have almost 1.4 billion active registered users, followed by Google+ with 343 million and Twitter with 284 million active users (Statista 2015). While offering a rich networking experience, by allowing users to create a network of friends/followers and communicate by sharing different types of information, such as posts, photos and location data, SNS entail privacy risks, due to the large amount of personal information that is published. These include unauthorized secondary uses of personal information, surveillance, identification theft, bullying and stalking (ENISA 2007). Users are increasingly worried about potential threats (Acquisti and Gross 2006, Boyd and Hargittai 2010), while 59% of teenager users perceive SNS as unsafe (StatisticBrain 2014). SNS users have been reported to follow several strategies to protect themselves, such as self-censorship (Marwick and Boyd 2014). Such an approach, however, cannot eliminate privacy threats, as users are still revealing personal data in an indirect way, as well as secondary data (ENISA 2007). Privacy-enhancing Technologies (PETs) such as access controls and privacy wizards (METAGROUP 2005) provide the means for protecting personal information while online. Relevant research, however, indicates that users of PETs are still far from reaching a critical mass (London Economics 2010, ENISA 2012). This is a complex issue to explain, as a multitude of factors can influence individuals to use privacy tools, including awareness, requirements for special IT skills, their complexity, the diversity of existing tools, costs, inadequate feedback of protection results, partial
addressing of users’ privacy requirements, responsibility misconceptions, and culture (Vemou and Karyda 2013).

Extant literature on PETs deployment is dominated by studies on their technical characteristics, functionality and constraints (e.g. Yeung et al. 2009) discuss the advantages and disadvantages of decentralized SNS platforms), whereas individual and social aspects are only scarcely tackled (e.g. Balsa et al. 2014, Fahl et al 2012, Wang et al 2014). On the other hand, several studies as well as technical reports claim that specific tools offer user-friendly functionality (e.g. MyPermissions\(^1\): “…even my mom could manage her permissions now”, Secure.me\(^2\): “…providing simple and easy actions that you can take to secure yourself”, PrivacyFix\(^3\): “…simple way to manage your online privacy settings through your mobile device. Access... one main dashboard that quickly & easily shows you what and with whom you’re sharing stuff…”).

This paper explores low PETs adoption in the context of SNS from a user perspective and discusses the findings of an empirical evaluation of PETs deployment. 170 SNS users of different social media provided us with information on their habits concerning their interaction with social networks, which (if any) privacy tools they have used, out of a list of 8 well-known applications (MyPermissions, PrivacyFix, Friendwheel, Bitdefender Safego, ZoneAlarm Privacy Scan, Secure.me, Priveazy lockdown, Safe Secure and Private Facebook messages) and rated the importance of a set of factors that have been identified as important for PETs adoption in relevant literature.

Analysis of survey findings suggests that knowledge of PETs is still limited among SNS users; interestingly, however, 65 out of the 170 respondents replied that that they were familiar with at least one of the tools but had never used them. We also found that users are mostly interested in the effectiveness, quality and cost of a privacy tool. They consider as less important, comparatively, the time needed to get familiar with the tool, user support and requirements for technical expertise or prior knowledge.

This work contributes to the understanding of limited PETs deployment, as it discusses the views of users with regard to these tools. Findings are not to be generalized; they can, however, provide us with insights as to what stimulates or discourages users from adopting privacy enhancing tools and applications, which can inform the design and development of PETs. Research findings also point to several issues that need further investigation, such as the user awareness and training as well as the economics or the cost to use technology.

2 ADOPTION OF PRIVACY ENHANCING TOOLS BY SNS USERS

Following research on rising privacy concerns of SNS users (Acquisti and Gross 2006, Boyd and Hargittai 2010), several Privacy-enhancing Technologies (PETs) have been developed to mitigate privacy risks. Such technologies include a wide range of applications, such as access controls, audience segregation, privacy-signalling tools, third-party tracking tools, social identity management systems, visualization tools and decentralization of Social Network Services (Vemou and Karyda 2013). While some PETs are embedded in SNS platforms (e.g. basic access controls), others adopt a different approach, altering the SNS platform architecture, as in the case of decentralized SNS. However, the majority of PETs concerns tools that can be used as add-ons to the SNS functionality (e.g. browser plug-ins, such as Scrambls\(^4\), MyPermissions\(^5\) and Friend Inspector (Cetto et al 2014)) and users need to initiate their use, in terms of download and installation (Jaatun et al 2011).

Despite the abundance of privacy enhancing tools and applications and the claims of software providers to satisfy users’ privacy needs\(^6\), users of SNS platforms do not show preference in their use.

\(^1\)http://mypermissions.org/story/
\(^2\)https://www.secure.me/en/about/
\(^3\)https://play.google.com/store/apps/details?id=com.avg.privacyfix
\(^4\)https://scrambls.com
\(^5\)http://mypermissions.org/
\(^6\)http://mypermissions.org/story/
For instance, adolescent SNS users prefer to apply their own strategies on posted data, such as censorship and social steganography (Marwick and Boyd 2014) and show little interest in applying add-on PETs (Balsa et al. 2014). The same applies to embedded PETs in SNS platforms, such as privacy settings, that usually go unnoticed by users (Boyd and Hargittai 2010) or are not used effectively, leading to unintended sharing of personal data (Madejski et al 2011).

Low adoption of PETs among SNS users is credited to different factors, such as lack of users’ awareness (Acquisti and Gross 2006, Pötzsch 2009), requirements for special IT skills (Compañó and Lusoli 2010) and diversity of existing tools (Vemou and Karyda 2013). There are also usability issues, such as the time needed to learn a new tool (Vemou and Karyda 2013). Developing user-friendly PETs has been identified as a key requirement for their design (Le Métayer 2010, Jaatun et al 2011) and previous research provides evaluation on several aspects such as non-intrusiveness (Balebako et al. 2011), intuitive interfaces (Kolter et al 2010) and low performance degradation (Saint-Jean and Feigenbaum 2010). Moreover, literature identifies technical issues affecting usability (Madejski et al 2011, Vemou and Karyda 2013) as affecting the adoption of PETs.

Wästlund et al. (2010) developed a questionnaire-based tool for users to assess the usability of PETs, either at a general level, or with regard to specific privacy-critic areas, such as data management and data release. Leon et al. (2012) evaluated a set of available block-tracking tools, based on user experience, leading to conclusions on users’ understanding of the interface and configuration capabilities, while Balsa et al. (2014) asked SNS users to evaluate usability of cryptographic access control tools.

Lately, research on privacy tools adoption revealed that user background factors, motivations and social network site experiences influence the use of PETs offered within SNS platforms (Litt 2013).

Conclusively, extant literature identifies a multitude of factors influencing the use of privacy enhancing technologies, emphasizing on awareness and usability issues. This paper adopts a user perspective to explore how SNS users protect their privacy and identify their attitudes towards the use of PETs.

3 PRIVACY ENHANCING TOOLS AND APPLICATIONS FOR SNS

As mentioned above, several types of privacy-enhancing tools, such as access controls, third-party tracking tools, and social identity management systems have been developed to assist SNS users in protecting their privacy. Several types, such as access controls, are now embedded in SNS platforms, however, they only address partial privacy needs and their effectiveness is questionable (Madejski et al 2011). Thus users still need to employ several PETs, add-ons to the SNS platform functionality, to ensure their privacy is protected.

Through research we have identified a set of add-on applications that protect user’s privacy. The list was not meant to be exhaustive, but to cover different social media platforms, such as Facebook, Twitter and Google+. We found no official information about the use of these applications, thus our intention was to compile a list of different types of tools and applications available to SNS users for protecting their privacy. Also, as the focus of this research is on SNS users, we excluded general privacy preserving applications such as blockers of third-party tracking entities.

MyPermissions is a browser extension assisting users in managing the access of third-party applications to their profile and is applicable to several SNS platforms, including Facebook, Google+ and Instagram. It reports the applications having access to personally identifiable information, per type of access (e.g. read access, post in the name of user access, location information access) and allows users to directly remove them from their profile, thus offering a single management interface for third-party applications. Another feature, offered as a premium service, is real-time notifications on new applications added and gained access to user’s personal data.

7https://disconnect.me/help
PrivacyFix\(^8\), developed by AVG, is a browser extension scanning user’s privacy settings and proposing changes to apply stricter controls, based on usual privacy risks. The tool then acts as an interface for the user to amend privacy settings. It also presents a list of entities tracking the user’s navigation to SNS, along with the choice to block them and an awareness raising feature, presenting the value of the user’s data in money.

Friendwheel\(^9\) is a network visualization application for Facebook and Twitter users. Browsing a user’s friend list and the respective lists of his or her friends, it creates a radial colourful graph-wheel presenting all user’s friends and how they are connected to each other. It also offers the ability to switch data parameters and create a visual wheel for users’ networks.

Bitdefender Safego\(^10\) is a free application for Facebook and Twitter scanning each post on the user’s profile and informing for malicious links (virus infected links or spam messages). At the same time, it warns the user’s friends about findings. Another feature, still in a pre-mature level, is scanning the user’s posted information to reveal posts of sensitive data, compromising privacy (e.g. public address information) and directing the user to the respective webpage, in order to amend settings.

ZoneAlarm Privacy Scan\(^11\) (renamed to SocialGuard Privacy Scan) is a free Facebook application accessing recent activities in the user’s profile (posts, tags, likes) and the respective audience settings, to identify posts that could be privacy compromising, e.g. posts that are visible to friends of friends. It then presents the results to the user, along with a grade for user’s privacy during the past month. It also provides recommendations to change privacy settings, in the form of a Frequently Asked Questions tutorial.

Secure.me\(^12\), developed by Avast, is a free Facebook application scanning users’ profiles and settings to find potential privacy risks by publicly available personal information. It then provides recommendations on audience access control settings. The tool also provides warnings about questionable posts on the user’s profile, posted by him or third-party applications. It also provides parents with the capability to observe their children’ profiles and be notified for privacy risks.

Priveazy lockdown is a deprecated free web browser extension providing user notifications on privacy and security configurations related to several privacy risks. It could be used with Facebook, LinkedIn, Twitter and several other platforms. Priveazy lockdown provided privacy wizards with detailed information, helping the user change his privacy settings and even notified when changes in the platforms’ setting boards had been applied, to revisit them.

Safe secure and private Facebook messages\(^13\) (renamed to Secret Wall) is a free browser extension encrypting the user’s Facebook private messages and timeline posts. The user creates an audience group and invites friends, who receive the necessary decryption key to see posted content. It is available for Firefox, Chrome and Internet Explorer 8+ browsers.

4 \textbf{THEORETICAL BACKGROUND: ADOPTING PRIVACY PROTECTION TECHNOLOGY}

According to the Diffusion of Innovations theory (Rogers 2010), four main elements influence the spread of a new idea or technology: the innovation/technology per se, the communication channels, time, and the characteristics of the social system. Rogers (2010) defines an innovation as "an idea, practice, or object that is perceived as new by an individual or other unit of adoption”. Despite the fact that PETs have been developed since more than two decades now, when cryptographic applications became widely available (Danezis and Gurses 2010), for most users they are still a novelty.

\(^8\)https://privacyfix.com/
\(^9\)https://friend-wheel.com/
\(^10\)http://www.bitdefender.com/solutions/bitdefender-safego.html
\(^11\)https://www.facebook.com/games/spprivacy/
\(^12\)https://www.secure.me/en/
\(^13\)www.secretwall.me
Drawing on innovation theory, use of PETs can be considered as an individual decision-making process, comprising the following steps: i) individuals become aware of the innovation and have some idea of how it functions (knowledge stage), ii) they form a favourable or unfavourable attitude toward the innovation (persuasion stage), iii) they choose to adopt or reject the innovation (decision stage), iv) they employ the innovation (implementation stage) and v) finally users evaluate the results of the innovation and finalize their decision to continue using it (confirmation stage). Consequently, to understand the diffusion of PETs we need to take into consideration not only their technical characteristics, but also social and communication issues (e.g. the role of technology opinion leaders and the nature of communication between SNS users), as well as aspects of individual decision-making.

To investigate how users’ shape their decisions with regard to employing privacy enhancing technologies we derived a set of factors from relevant literature and we composed a questionnaire, based on these factors.

- **Quality of the tool/application**: Quality refers to the tool properties and characteristics that contribute to meeting users’ privacy needs.
- **Cost**: The cost required to acquire a tool, in terms of direct buy or subscription costs, has been found to influence user behavior towards or away from using a tool (Acquisti 2010, Vemou and Karyda 2013). Providing a free tool (or free features of it) encourage users to try and, in a later stage, to adopt it. However, as a percentage of users are willing to pay a certain cost for privacy tools (Acquisti 2010), trial versions, similar to those of various antivirus software packages (e.g. ESET 30-days trial), may convince them to overcome cost barriers.
- **Time to learn the tool**: The amount of time that is required for a typical (non-expert) user to become familiar with a tool affects its adoption (Vemou and Karyda 2013). It refers to whether the interface is intuitive (Kolter et al 2010), whether the user is required to perform simple, easily-remembered steps to accomplish a task, but also to wizards or other helping features guiding the user through his first experience of using the tool. Considering the diversity of tools offering similar functionality, e.g. access controls, time to learn differentiates tools.
- **Ease of use**: Ease of use (also described as perceived complexity) has been identified as a major factor influencing the adoption of an innovation (Davis 1986, Rogers 2010). Users prefer applications which are easy to use, both during installation/configuration and every day operation (Leon et al 2012). In terms of installation, steps need to be few and easily understood (London Economics 2010), or even to be provided with a wizard, a demo or a presentation that guide the user to understand and apply without fear of severe mistakes. In terms of operation, users require few and easy steps in order to remember application. This is very important considering the diversity of existing PETs, fulfilling partial privacy requirements, thus requiring the user to apply several tools for protection (Vemou and Karyda 2013). Also, impact on performance of everyday SNS operations is important, as users wish to protect privacy without experiencing low latency of the service (Balsa et al. 2014).
- **Requirement for technical expertise**: Several privacy-enhancing tools involve tasks which require advanced technical skills, beyond that of an average user (Vemou and Karyda 2013). For instance, users of Safe secure and private Facebook messages need to manage encryption keys, whereas to use MyPermissions one needs to be familiar with managing browser extensions. As requirements for technical expertise can discourage users from employing PETs, Balsa et. al. (2014) stress the need to further research how to securely automate some tasks on cryptographic tools and has been one of the main reasons mandating promotion of built-in privacy in SNS platforms (Vemou and Karyda 2014).
- **Technical support**: Assistance could be provided through the tools’ website, by phone, email or even live chat. In addition, demos may guide users through installation and configuration, thus achieving satisfaction.
- **Effectiveness Feedback**: Users’ perceptions on the effectiveness of PETs influence their adoption. This can be attributed to the way PETs communicate, or rather fail to do so, their results, and to the way they give feedback for actions they have performed to protect the user (Balsa at al. 2014). Also, as the way privacy related dangers are presented relates to perceived effectiveness of a tool (Vemou and Karyda 2013), offering the user transparency of successful
operations on personal data protection and adequate risk presentation is one of the factors influencing towards the adoption of a tool. In addition, brief awareness features, such as warnings for third-party applications access rights, prior to installation (e.g. my permissions cleaner) or scores of achieved privacy level (e.g. Bitdefender Safego) are features that directly and non-intrusively catch the user’s attention and can positively influence PETs adoption.

- **Reputation/popularity of the tool**: It is common for users to avoid being the first to try a tool or a technology, fearing side-effects of technology immaturity (Wu and Wang 2005). However, positive reputation of a tool and popularity motivates users to try it (Rogers 2010). Thus, users base their choice of tools on assessments from other users and even seek for comments in social media or forums. Positive assessments and trust by other users, as well as adoption by other users in the same SNS platform positively influence users to adopt a tool, and so does popularity of the tool provider, being well known from other applications.

5 **Empirical Research and Data Collection**

5.1 **Research design and data collection**

To explore users’ preferences with regard to PETs adoption, we conducted a survey involving SNS members, using an online questionnaire created with Google Drive Forms. Online questionnaires allow anonymous answers and provide sufficient time for respondents to answer, while they support the collection of large amounts of information in a short time (Henerson et al. 1987). After a pilot test involving ten SNS users, who helped identifying and correcting ambiguous wording, the questionnaire was disseminated via public profiles in Facebook and Google+, created by the authors for the purposes of the research.

Overall, we received a total of 170 completed questionnaires. 61% of the respondents were female, 82% were between 22 and 45 years old, 12% between 16 and 21, and 6% were over 46 years old. The great majority of users (89%) have finished college. 48% of the respondents work in the private or public sector and 30% were students.

5.2 **The questionnaire**

The questionnaire (provided in appendix A) included 17 multiple-choice questions arranged in 4 sections and began with an introductory paragraph explaining the purpose of the survey, which was followed by 4 questions on personal status (age, sex, education and occupation). Respondents were then asked to provide information on the social networking platforms they use, out of a list of the five most popular ones (Facebook, Pinterest, Twitter, Google+ and LinkedIn) according to eBiz/MBA (Ebiz/MbaB 2014) and how frequently they use them. The first four platforms are general purpose SNS offering the user a wide range of networking services, from sharing comments and photos to adding third-party applications (except from Pinterest that does support third-party applications). LinkedIn belongs to niche SNS platforms (ENISA 2010), offering special purpose networking services, related to one’s professional sphere. While we expect users to be more cautious on posts in LinkedIn (Martensen et al 2011), we included this platform in the questionnaire because of the high impact a social network can have on professional life (Black and Johnson 2012).

Respondents where then asked on the types of personal data they provide, whether they provide real or false data and if they communicate with persons they do not know. They were also asked whether they use applications provided by third parties, if they control their privacy settings and whether they maintain private or public profiles.

In the following section respondents were asked to indicate which privacy preserving applications they were familiar with and which of these they had actually used (out of a list of 8 tools). Finally, respondents evaluated a set of factors (elicited by relevant literature as described in Section 4) with regard to their importance for using a privacy tool, on a 5-point Likert scale.
6 RESULTS ANALYSIS: SNS USERS EXPERIENCE WITH PETs

6.1 Users’ privacy behavior

All users reported that they are members of at least one of the social networks in the questionnaire, namely Facebook, Google+, Pinterest, Twitter and LinkedIn. Very few respondents used the option “other” provided and specified additional SNS: Tumblr, Delicious, Foursquare, Instagram, StumbleUpon, We heart it, Favim and ask.fm. Interestingly, 79% of the respondents replied that they are members in more than one social networking service and 57% are very frequent users.

Most respondents (85%) post their real name and personal data, and half of them (51%) replied that their social networks include individuals not previously known to them. Most users publish their personal photos (81%) and date of birth (67%), their education background (67%) and many indicate their occupation (48%). However, most users refrain from posting information about their sexual, religious and political preferences (only 18%, 11% and 6% respectively, provide this information). 40% provide their email address, 6% their home address and only 2% publish the number of their mobile phone. Several users publish their marital status (19%), the name of their partner (11%) and names of members of their family (26%). 23% of users publish photos very frequently. It is also worth noting that publishing false information is not an exception: 21% replied that they use fake names, 15% that they do not publish the real date of their birth, 10% provide fake email addresses, 10% provide false information about their family and personal status and 15% deliberately provide false information with regard to their occupation.

37% of the respondents use third-party applications, such as games and contests, and 36% replied that they are aware of the fact that such applications access their personal data. 19% of them consider this an unimportant issue, whereas 31% check privacy settings when they use these applications. Few users (13%) have never used third-party applications. Most users (64%) chose to maintain a private profile, 57% control who has access to their personal information and 2% replied that they do not know if their profile is public or private. About half of the respondents (53%) use the privacy settings provided by the SNS platform very often or often, while the rest use them seldom or occasionally (44%) and only 2% have never used them. However, fewer users (33%) consider privacy settings easy to use, whereas more than half (55%) find them difficult or relatively difficult to use.

6.2 Deployment of privacy tools

Overall, the majority of respondents were not familiar with the PETs indicated in the survey (see Table 1). 18 users have used Safe secure and private Facebook messages and 9 used Secure.me. MyPermissions, PrivacyFix and ZoneAlarm Privacy Scan have been used by 7 users (each), while 5 users have used Bitdefender Safego. Last but not least, Friendwheel and Priveazy lockdown have been used by 1 user each.

<table>
<thead>
<tr>
<th>Tools</th>
<th>Have used it (%)</th>
<th>Aware but have not used it (%)</th>
<th>Not aware (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MyPermissions</td>
<td>4</td>
<td>13</td>
<td>83</td>
</tr>
<tr>
<td>PrivacyFix</td>
<td>4</td>
<td>12</td>
<td>84</td>
</tr>
<tr>
<td>Friendwheel</td>
<td>1</td>
<td>8</td>
<td>91</td>
</tr>
<tr>
<td>Bitdefender Safego</td>
<td>3</td>
<td>9</td>
<td>88</td>
</tr>
<tr>
<td>ZoneAlarm Privacy Scan</td>
<td>4</td>
<td>18</td>
<td>78</td>
</tr>
<tr>
<td>Secure. me</td>
<td>5</td>
<td>21</td>
<td>74</td>
</tr>
<tr>
<td>Priveazy lockdown</td>
<td>1</td>
<td>11</td>
<td>88</td>
</tr>
<tr>
<td>Safe secure and private Facebook messages</td>
<td>11</td>
<td>19</td>
<td>70</td>
</tr>
</tbody>
</table>

Table 1. Use of privacy tools.
Interestingly 65 users (out of 170) replied that they have never used any of these tools, despite the fact that these users had previously replied that they are aware of at least one of them. For instance, 21% replied that they are familiar with Secure.me but had not used it. This finding suggests that limited awareness should not be solely accredited for low PETs adoption.

When asked to evaluate the importance of the factors that would influence their decision in deploying a privacy preserving application, respondents identified the following factors as important: **quality** (84% reported it is important or very important, 14% not so important or little important and 2% not important at all), **effectiveness feedback** (79% reported it is important or very important, 17% not so important or little important and 4% not important at all), **cost** (77% reported it is important or very important, 19% not so important or little important and 4% not important at all), **ease of use** (65% reported it is important or very important, 31% not so important or little important and 4% not important at all) and **time needed to learn how to use the tool** (59% reported it is important or very important, 36% not so important or little important and 5% not important at all). Users were divided with regard to the importance of **requirements for technical expertise**: 51% consider it important or very important, 43% not so important or little important and 6% not important at all. Half the respondents consider that it is important or very important that other SNS users use the same privacy tool (**popularity**), 45% consider it of not so important or little important, of and only 5% consider it of no importance at all. Finally, **technical support** is identified as important or very important by 60% of the users, while 34% consider that it is of moderate or low importance.

7 DISCUSSION: THE USER PERSPECTIVE

The findings of the survey presented in this paper provide an insight to the privacy behavior of SNS users and the deployment of Privacy-enhancing Technologies from the user perspective. These findings are not to be generalized, as the sample was relatively small and self-selected; the set of privacy tools included in the questionnaire was also small and not exhaustive, in order not to drive away respondents. The results of the survey are in line with relative research and move the discussion on the diffusion of PETs one step further by a) providing the user’s perspective; b) indicating that lack of awareness is not the only obstacle to the diffusion of privacy enhancing technologies and c) providing feedback on which factors users consider more influential in adopting a privacy preserving application.

Overall, survey results provide us with interesting findings with regard to the privacy behavior of the average SNS user: most are quite active on publishing their personal information and share this with strangers, however, they often provide false data and avoid providing information of sensitive information, such as religion and sexual preferences. Very few did not know the difference between private and public profiles and many (more than half) regularly or some times change their privacy settings. It seems, thus, that SNS users are familiar with privacy issues and take some actions to protect their personal information. On the other hand, we found that most SNS users experience difficulties in managing the privacy settings offered by the SNS platforms or consider them as relatively difficult, aligning with findings of Madejski et al. (2011). Thus, users do not see SNS platforms as allies in their effort to protect their privacy.

Another interesting finding is that SNS users seem to ignore, or underestimate, privacy implications by third party applications. These applications are quite popular among SNS users, but it seems that users do not realize the extent of secondary uses of information. This is an area indicating that privacy awareness needs to be enhanced.

Most respondents were not familiar with any of the privacy enhancing tools included in the questionnaire. We believe that this result would not have been dramatically different even if we had selected a different set of tools. This finding has been cited elsewhere as well (Flash Eurobarometer 2008). The fact, however, that some users (not very few) chose not to use privacy enhancing tools, despite the fact that they are familiar with at least one of them, needs further investigation.

In terms of the Diffusion of Innovation theory (Rogers 2010), this means that users are still not aware of the innovation (PETs). Moreover, findings suggest that informed users may form an unfavourable
attitude toward privacy tools. Due to the nature of this research, we were not able to further investigate this issue. Future research needs to focus on how we can influence this stage towards a favourable decision.

In general, people are more likely to adopt an innovation if they believe that it will enhance their utility. This means they need to have an idea for the benefits they will get from using it. People also consider related costs and the degree to which the innovation would disrupt or change their ordinary actions. They evaluate how compatible the innovation is with their existing habits and values, if it is hard to use, and evidently, its effectiveness (Davis 1986). Some also take into account how socially acceptable the innovation is and what will other people think of them using it (reputation/popularity).

Survey results also suggest that SNS users consider quality, effectiveness feedback and low cost as the most important factors, in order to adopt a privacy tool. Obviously, we need to investigate further, the factors that can influence the decision of SNS users to deploy privacy tools, based not only on self-reported attitudes of users, but also on hands-on experimenting of such tools. Furthermore, as some factors may have different involvement in some types of PETs (e.g. in cryptographic PETs use of the tool by the user’s friends is a prerequisite) research should be conducted separately for different categories of PETs.

We can also derive a set of directions for designers and developers of privacy enhancing applications:

a) privacy tools need to enhance their functionality to let the user know which actions were performed by the tool to protect user privacy (e.g. which entities were restricted access to personal data from his profile), and which were the possible privacy threats; b) privacy tools need to fulfil multiple privacy requirements (for instance, while a user may be in need of assistance to set audiences of his posts, he is at the same time in need to protect his personal information from the SNS platform itself). Thus a tool proposing groups of friends for access controls or only encrypting messages could not be derived of high quality by the user, because it would address only a part of his privacy requirements. Organization of several privacy enhancing technologies under a single, simple tool would be more attractive to users; c) Last but not least, as high costs can be an inhibitor of privacy tools use, providers should study users’ perceptions over PETs costs prior to setting a tool’s price or consider offering some of its functionality for free, to convince users into trying the tool.

8 CONCLUSIONS AND FURTHER RESEARCH

This paper provides insights on the privacy behavior of SNS users and corroborates similar research on showing that adoption of PETs among them is still limited. This is mainly attributed to low awareness of privacy enhancing tools and applications among SNS users. It also argues that awareness is not the only factor strongly influencing PETs adoption and user weighting of several adoption factors, such as effectiveness feedback, costs and requirements for technical expertise should be further explored.

We were able to identify a number of factors that, according to the users, would significantly influence their decision in adopting a privacy tool. This also needs to be validated through future qualitative research, minding the different role each factor could have in different types of PETs (e.g. use of the tool by other users would be a prerequisite for its use). The outcome of such a research could provide more accurate directions for PETs developers, in order to focus on SNS users’ needs. For instance, as effectiveness was stated to influence users’ adoption attitude, providing transparency of threats the tool protected the user from, should be a leading consideration during its design and development.

Online questionnaires give respondents the opportunity to provide their views anonymously and without the pressure of time that would be present in an interview; however there is always the possibility of inaccurate responses, lack of spontaneity and the restriction to specific answers. Survey findings presented in this paper are not to be used for generalizing and do not serve as the only basis for the design and implementation of privacy preserving applications for social network users.

They can be used to understand the low diffusion of PETs in terms of limited user awareness, and indicate that user-centred factors need to be considered for designing and developing privacy tools.
References


ENISA 2010. ‘Online as soon as it happens’. Report 2010

ENISA 2012. ‘Privacy considerations of online behavioural tracking’. Report 2012


Appendix A – Survey questionnaire (translated in english)

User profile (personal status) 1. Sex

- Male
- Female

2. Age

- 16-21
- 22-28
- 29-35
3. Education
- Primary
- Secondary
- Higher (e.g. college)

4. Occupation
- Private sector
- Public sector
- Freelancer
- Unemployed
- Student
- Retired
- Other:

Use of social networking services
1. Are you a member of social networking services? (More than one answer is allowed)
   - Facebook
   - Google+
   - Linkedin
   - Pinterest
   - Twitter
   - None
   - Other:

2. How often do you use them?

<table>
<thead>
<tr>
<th>Frequency</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Select a value from a range of 1, Very difficult, to 5, Very easy.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. As a name of your profile you use:
- Real
- Alias
4. In social networks your contacts consist of:
   - Only those who I know
   - Mostly known and less strangers
   - Acquaintances and strangers
   - Mostly strangers and less known

Privacy attitudes

1. What information do you provide as a member of the social network you belong to?

<table>
<thead>
<tr>
<th>Information</th>
<th>I do not provide this information</th>
<th>I provide this information, but deliberately not complete and accurate</th>
<th>I provide this information and it is complete and accurate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Profile photo</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name day</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sexual preferences</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Languages</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Religious preferences</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Political preferences</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-mail</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobile phone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telephone (home phone)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alias</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Home address</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Website</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
I do not provide this information | I provide this information, but deliberately not complete and accurate | I provide this information and it is complete and accurate
---|---|---
Marital status | | |
Name partner | | |
Family relatives | | |
Profession | | |
Education | | |

2. Which of the following activities do you perform?

<table>
<thead>
<tr>
<th>Activity</th>
<th>Very often</th>
<th>Seldom</th>
<th>Little - not at all</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post photos</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Submit a video</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post a comment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participation in games</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participation in competitions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Check in</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Do you know which of your personal data you allow access to by third-party applications (apps), such as contests, games, etc, in order to use them?

- Yes
- Whenever I download an application, I check the privacy settings
- I do not pay attention
- I do not know what applications are
- I have never used such an application

4. You've set your profile as:

- Private so that only my contacts can see
- Partly private so that contacts of my contacts can see
- Public, which is open to everyone
- I do not know
5. If you have private or partially private profile:

- Limit which contacts can have access to information I provide
- Everyone can see the same
- I do not know

6. How often do you check the settings of your personal data?

<table>
<thead>
<tr>
<th>Very often</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Never</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select a value from a range of 1, Very difficult, to 5, Very easy.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

7. In social networks, the management of your personal data appears:

Select a value from the range 1 very difficult to 5 very easy

<table>
<thead>
<tr>
<th>Very difficult</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Very easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select a value from a range of 1, Very difficult, to 5, Very easy.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Tools to protect privacy**

1. Do you know and / or use any of the following tools?

<table>
<thead>
<tr>
<th>I've used</th>
<th>I know but I have not used</th>
<th>I do not know</th>
</tr>
</thead>
<tbody>
<tr>
<td>myPermissions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PrivacyFix</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friendwheel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bitdefender Safego</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZoneAlarm Privacy Scan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secure. me</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Priveazy lockdown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Safe secure and private Facebook messages</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2. How important do you consider the following factors for the use of protection tools? (rating from very important to not at all important)

<table>
<thead>
<tr>
<th>Factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Cost</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Ease of use</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Time to learn the tool</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Requirement for technical expertise</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Technical support</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Effectiveness</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Reputation/popularity of the tool</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
</tbody>
</table>
COUNTERING THE COLLUSION ATTACK IN A TRUST-BASED MANET

Ryma Abassi, Higher School of Communication, Supcom, University of Carthage, Tunis, Tunisia, ryma.abassi@supcom.rnu.tn

Sihem Guemara El Fatmi, Higher School of Communication, Supcom, University of Carthage, Tunis, Tunisia, sihem.guemara@supcom.rnu.tn

Abstract-
Using trust management can be highly beneficial for MANET (Mobile Ad hoc NETwork) security. In fact, it allows the establishment of trustful relations between collaborating nodes based on reputations. A reputation is the perception that a node can have about another one based on its past actions. It is generally computed from direct observations or neighbors’ observations exchanged using recommendations. Unfortunately, for malicious reasons, such method may be faked by cheaters: several nodes collude in order to rate each other with the maximum value and at the same time decrease other nodes’ reputations by giving negative recommendations about the latter. The main proposition in this paper is then, securing trust-based MANET against collusion attack. This is achieved by detecting colluding nodes through the calculation of a recommendation deviation and the punishment of these nodes by discarding them from further communication. These two processes are based on a formal modeling of collusion attack and its environment. Finally, several simulations were achieved proving the robustness and the efficiency of our proposal.

Keywords: Security, MANET, Trust Management, Collusion, Inference System, Validation.

1 INTRODUCTION

MANETs (Mobile Ad hoc NETworks) (Arijita et al., 2012) are wireless mobile nodes dynamically self organizing in arbitrary and temporary network topologies. Their nodes can be internetworked without a pre-existing communication infrastructure. Therefore, such networks are designed to operate in widely varying environments, from military networks to low-power sensor networks and other embedded systems. Dynamic topologies, bandwidth constraints, energy-constrained operations, wireless vulnerabilities, and limited security are among the main MANET characteristics.

In order to overcome the security problem, trust is generally used (Ruohomaa and Kutvonen 2005; Arijita et al., 2012). Trust enables collaborating nodes to counter their uncertainty and suspicion by establishing trustworthy relationships. Due to the importance of the challenge, trust is associated to a unified approach allowing its specification and formalization called Trust Management (TM). Recently, we defined trust relations between a Trustor (trust provider) and a trustee (trust beneficiary) as a binary decision relationship allowing their collaboration in a given situation with a given security level. This level is proportional to trustee’s reputation i.e. a perception a party creates through past actions about its intentions and norms (Ruohomaa and Kutvonen 2005). It is obtained through direct observations made by the node itself and/or by indirect observations. These latter correspond to the received appreciations from neighbours who have had interactions in the past and have evaluated and rated each others.
Paradoxically, success of trust based schemes depends on cooperation among the nodes (Arijita et al., 2012). In fact, the TM process may constitute a security weakness due to its vulnerability to the collusion attack where several malicious nodes may collaborate in order to decrease a benevolent node’s reputation. This is what we call collusion attack. It is defined as “A collaborative activity that gives to members of a colluding group benefits they would not be able to gain as individuals” (Qureshi et al., 2011). Several works dealt with collusion attacks [Kannhavong et al., 2006; Babu et al., 2008; Marmol and Perez, 2009; Grafii et al., 2007] but few were interested by collusion attack in the context of a TM process.

The main contribution of this paper is then, securing a trust-based MANET against collusion attack. This proposition extends an earlier work proposing a reputation-based environment using direct and indirect observations. Direct observations are obtained through the watchdog mechanism whereas indirect ones are exchanged using recommendations. Two main activities compose the proposition: detecting colluders and punishing them. Detection is achieved through the calculation of a recommendation variance from other one since obviously colluder estimation will be divergent from legitimate ones. Punishment discards detected colluders and prevents them from participating in future communications. This was made through a formalization of the main needed concepts as well as simulations proving the feasibility of the whole proposal.

The remaining part of this paper is structured as follows: Section 2 recalls some existing works dealing with collusion attack. In Section 3, we present a trust environment and a collusion attack modelling in such environment. Section 4 depicts our main proposition: how a collusion attack is detected and how a colluder is punished? Some simulations and results showing the feasibility and the efficiency of our proposal are given in Section 5. Finally, Section 6 concludes this paper.

2 RELATED WORK

Several works dealt with Trust in MANET (Michiardi and Molva, 2002; Buchegger and Boudec, 2002, Movahedi et al., 2012). “CORE”, a collaborative reputation mechanism (Michiardi and Molva, 2002), was designed in order to use the “Watchdog” mechanism as a monitoring mechanism. This latter, allows maintaining a reputation table at each node to keep track of reputation values of other nodes. Only positive rating factors can be distributed among nodes since a selfish node may send false negative rating factors to other nodes and may cause disruption of the reputation system. In (Buchegger and Boudec, 2002), CONFIDANT Cooperation Of Nodes: Fairness In Dynamic Ad-hoc Netwroks was proposed based on four main components namely a monitor, a reputation system, a path manager, and a trust manager. These components are required to be implemented in every node. Recently, Movahedi et al (Movahedi et al., 2012) proposed ATMS (an Automatic Trust Knowledge Monitoring Scheme), a knowledge monitoring scheme for trust management based on autonomic principles. ATMS is based on five main components: the monitor collecting local and global information, knowledge containing tables in which collected information are stored, analyzing verifying whether policies are violated, planning executing the knowledge monitoring optimization algorithm and execution enforcing decisions taken by planning component.

Paradoxically, trust schemas may be vulnerable to some attacks. Because the high interest perceived by the researchers, several works dealing with collusion attack in MANET were made. This kind of attack targets the trust process securing routing protocols. Most of them were interested by OLSR based environment where nodes collude during communications. In (Kannhavong et al., 2006), Kannhavong et al. introduced a collusion attack model against OLSR as well as a technique to detect the attack by utilizing information of 2-hops neighbours. In fact, according to this model, a node must be able to learn topology up to 3-hops in order to check whether link info advertised by its 1-hop neighbours is reliable or not. However, in this proposition, a node is not able to differentiate between mobility induced topology change and the collusion attack. In (Babu et al., 2008), authors present CAP-OLSR a Collusion Attack Prevention protocol for OLSR. This protocol prevents the collusion attack based on trust information using an information theoretic framework. Nodes establish trust relations among them using direct observations and uncertainty calculation as well as using recommendations gathered from other nodes. Entropy functions have been used to represent the
uncertainty involved in a node’s behaviour. Recommendations are exchanged through two new introduced messages, Trust Request (TREQ) and Trust Reply (TREP).

The main proposition of this paper is the proposition of collusion detection and prevention process suited for trust based MANET.

3 COLLUSION ATTACK BASIC CONCEPTS

In this section, we present the main concepts inherent to TM in MANET. Then, we introduce our collusion attack modeling in such environment.

3.1 Trust modeling

In a recent work (Abassi and Guemara El Fatmi, 2012), we proposed a trust-based modelling for MANET using the main trust characteristics such as asymmetry, reflexivity, permanence, scalability, etc. This modelling uses reputation in order to establish trust relations between a Trustor and a trustee and is based on four steps:

1) Collecting information about a given node by asking other nodes their recommendations.
2) Aggregating received information properly in order to calculate reputation.
3) Selecting the most trustworthy node.
4) Punishing or rewarding according to the obtained satisfaction in order to adjust the reputation of the node.

In fact, a Trustor can request the establishment of a trust relation as follows:

\[ \text{Trust (T, t, a, o, \{c\}) \rightarrow level} \] (1)

Where ‘T’ stands for the Trustor, ‘t’ the trustee, ‘a’ the action concerned by the trust, ‘o’ the object of the action, ‘c’ the relation constraints and ‘level’ for the trust level, ‘c’ is an optional field. Let’s note that T and t represent any generic MANET node.

The trust level is proportional to the trustee’s reputation. Moreover, we assume that reputation value can range from \{-3, -2, -1, 0, 1, 2, 3\}. The use of these values was made in order to have more expressiveness concerning trust relation evaluation and to cover three potential degrees by analogy to the low, medium and high degrees. Hence, a trustworthy node is associated to a \( rep \in \{1, 2, 3\} \) where an untruthful node is associated to a \( level \in \{-3, -2, -1\} \). The \( rep == 0 \) is used to characterize an unknown node and thus to integrate it in the communication process and avoid the starvation problem i.e. avoiding the exclusion of an unknown node without a reputation.

In this paper, we are interested by reputations that are calculated based on direct and indirect observations collected through recommendations. These latter are weighted according to the reputation of recommending nodes. A node solicits a recommendation concerning a trustee by broadcasting a recommendation request Recd-REQ.

\[ \text{Recd-REQ (Trustor, trustee)} \] (2)

Each node receiving this request and having already communicated with this trustee responds by a Recd-RESP containing a trust-level depicting corresponding to the trustee’s reputation such as perceived by the node identified by its ID for a given context cxt described through an action on an object.

\[ \text{Recd-RESP (ID, Trustor, trustee, trust-level)} \] (3)

Let’s note that a weight \( w_i \) is associated to each node \( i \) as follows:

- If \( rep_i \in \{-3, -2, -1, 0\} \) then \( w_i = 0 \);
- If \( rep_i = 1 \) then \( w_i = 1 \);
If $rep_i = 2$ then $w_i = 2$;
If $rep_i = 3$, then $w_i = 3$.

Once collected, these recommendation levels are sent to the reputation calculation engine which is in charge of assessing the corresponding reputation as follows:

$$rep_{a,b} = \frac{3 \times repl_{a,b} + \sum_{i=0}^{n-1} w_i \times repl_{i,b}}{|n| + 1}$$  \hspace{1cm} (4)

Where $w_i$ is the weight associated to the node $i$, $n$ is the number of evaluating nodes and $repl_{a,b}$ is the reputation given by $a$ to $b$. The value $repl_{a,b}$ is used when a given node $a$ has already its own appreciation about node $b$ and is updating it. It is calculated as follows where $m$ is the number of communications between $a$ and $b$, $fw_{a,b}^k$ represents a satisfaction value when node $b$ forwards packets from node $a$ for the $k^{th}$ times:

$$repl_{a,b} = \frac{\sum_{k=1}^{m} repl_{a,b} + fw_{a,b}^k}{\sum_{k=1}^{m} repl_{a,b}}$$  \hspace{1cm} (5)

$$fw_{a,b}^k = \begin{cases} 1 & \text{if forwarded} \\ -1 & \text{else} \end{cases}$$  \hspace{1cm} (6)

The $repl_{a,b}$ value is pondered by weight $3$ given that trust is a reflexive relation implying that each node trusts itself.

Finally, we extended the neighbor table in order to store the reputation of each neighbor such as depicted by Table 1.

<table>
<thead>
<tr>
<th>NodeID</th>
<th>Reputaion</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>-1</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1. Neighbour Table Extension

In the rest of this section, our collusion modeling is detailed.

### 3.2 Collusion attack modeling

There are certain attack models we are not addressing in this paper. Therefore, we made some assumptions here:

1. Reputation and local trust values are securely protected.
2. At least, two nodes have to cooperate to achieve such attack.
3. Colluders can always collaborate with each other and attack the network.
4. A colluder may not be neighbor with the Trustor. This is because the recommendation request is broadcasted and any node in the network may respond.
5. Each colluder must be neighbor with at least another colluder.
6. A colluder may not be neighbor with the trustee. In fact, a colluder is cheating so it can do it even when it is not neighbor with the trustee and didn’t communicate with him before.
7. Two colluding behaviors are considered: the continuous and the camouflaged.

More precisely, in this paper, we focus on collusion attack in which two or more attackers collaborate to disturb the TM process. Attackers rate each other with the maximum value and on the other hand, decrease the reputation of benevolent nodes by giving negative recommendations about them. Hence, in the rest of this work, a collusion attack is seen as a group of nodes selectively giving wrong recommendations for a given node (victim) in order to decrease its reputation and consequently to discard it from communication.
A great challenge concerning collusion is its detection (Kamvar et al., 2003). In fact, attackers try to be hiding in legitimate traffic. In this section, the main contribution of this paper is presented: securing a trust management process against collusion attack. This is made following two steps operating in parallel: detecting the colluders and preventing this kind of attack.

4.1 Collusion attack prevention

Let’s recall that recommendations are exchanged when an unknown node arrives in the network. The process depicted by Figure 1 is proposed in order to avoid collusion attack. In fact, when a node needs to evaluate the trustworthiness of an unknown node, it uses its neighbors table in order to select trustworthy nodes to which it will send the recommendation requests \textit{recd-req}. Two cases are conceivable: (1) no trustful node responds because none has already communicated with this node (2) there are some trustful responding nodes.

![Figure 1. Prevention Process](image)

In the first case, this node i.e. \textit{trustee} is affected by a neutral reputation that will be updated after communications through direct observations. In the second case, recommendations are used in order to calculate a first estimation of this reputation. If the node is considered as trustful, then communications will be triggered in parallel with the monitoring process else the punishment process is triggered.

4.2 Collusion attack detection

Obviously, colluders’ recommendations are divergent from legitimate ones. Hence, in order to detect colluders, a reputation variance is calculated. This latter serves as a quality measure of the reputation value computed based on recommendations. Assessing this deviation may contribute to colluders’ detection. Let’s have $x_i$ is the obtained reputation by node $x$ on node $y$’s using its behavior and received recommendation at the instant $i$. 
Let’s have $X: \Omega \rightarrow \mathbb{R}$, we measure how far a set of reputations is spread out using the following variance

$$V(X) = E((X - E(X))^2)$$  \hspace{1cm} (7)

where $E(X) = \sum_{x \in \Omega} xP(X = x)$

Variance measures the deviation between a reputation and the average of received reputations. Moreover, in order to assess the reliability of the estimate, this variance is used to create a confidence interval $I$ as follows:

$$I = \pm \frac{\sigma(X)}{\sqrt{n}}$$  \hspace{1cm} (8)

Where $n$ corresponds to the number of recommending nodes and $\sigma(n)$ is the variance square root. This latter is triggered by every node having to calculate the deviation of another node identified by ID when receiving recommendation response for a given trustee $t$.

To detect colluders, a threshold value $\delta$ is used indicating the maximum allowed variance. Using the presented metrics, the whole detection process is described by Table 2.

### Table 2. Collusion attack detection inference system

Let’s consider a finite domain $P$ containing all received recommendation responses. The rules presented in this system apply to a quadruplet ($\emptyset, D, 0$ and $\emptyset$) whose first component is the detected colluders set, whose second component $D$ is a subset of $P$, whose third component $n[R]$ (initially affected to zero) is the number of variances detected for node identified by its ID and whose last component is the calculated variance. We write $C \vdash^* C'$ where $C'$ is obtained from $C$ by application of the inference system. The inference rule `recurcall` is applied when a node receives a recommendation response in order to check whether this response is greater than the chosen threshold. In such case, $n[R]$ is incremented. The inference system stops when $n[R]$ reaches the value 3.

### 4.3 Collusion attack punishment

Prevention is also achieved through colluders’ punishment as presented in Table 3. The rules presented in this system called inference rules apply to a 5-uplet $(\text{rep}_ID, B, \text{rep}_t, G$ and $n[ID])$ whose first component $\text{rep}_ID$ is the node reputation; whose second component $B$ is the Blacklist; whose third component $\text{rep}_t$ is the trustee reputation; whose fourth component $G$ is the Greylist and the last component $n[ID]$ is the number of deviations.
4.4 Dealing with camouflage

As introduced previously, a special type of collusion exists where attackers do not send fake recommendations continually but in \( p \%) \) of times: this is what is called camouflage attack. In order to detect such attack, we propose a detection scheme based on two components: a classifier and a correlator. Once a recd-resp is received, the classifier affects each response to one of the following three classes designed by class1, class2 and class3. Class1 regroups all recommendations having a recommendation level lesser than 0. Class2 regroups all recommendations having a recommendation level equals to 0 and Class3 regroups all recommendations having a recommendation level \( \in [0; 3] \) . The next step is then the generation of the behavioral pattern followed by the malicious node. This is made by the correlator periodically, by analyzing the class column in the reputation table for a given trustee and generating statistics in order to characterize the behavioral pattern of the node.

This behavior can be described by the following equations. Let’s have \( \text{Recd} \) ranging over finite number of recommendations \( r_1, ..., r_n \) associated respectively to the following probabilities \( p_1, ..., p_n \) where \( n \) is the number of recommendations given by a node. The standard deviation \( \sigma \) is expressed as follows:

\[
\sigma = \sqrt{\sum_{i=1}^{n} p_i (r_i - \bar{r}_i)^2} = \sqrt{\left(\sum_{i=1}^{n} p_i r_i^2\right) - \bar{r}^2}
\]

Having that the law of \( \text{Recd} \) is uniform over a finite set of values then \( \sigma \) can be reduced as follows:

\[
\sigma = \sqrt{p(\sum_{i=1}^{n} r_i^2) - (\sum_{i=1}^{n} p_i r_i)^2}
\]

According to the normal distribution, this deviation is acceptable if it is not exceeded with a probability \( \text{prob} \) of more than \( 3/1000 \):

\[
\text{prob} = (\sigma - \bar{r}) \geq \frac{3}{1000}
\]

Let’s note that in order to minimize false positive error, the observation time \( t_{\text{obs}} \) must be well chosen. In fact, a large value of \( t_{\text{obs}} \) may increase the false positive rate. For example when only two recommendations are considered (\( t_{\text{obs}} \) is small), the standard deviation will be large especially when the considered trustee is malicious and consequently, the Recommendor will be considered as a camouflaged node.

5 SIMULATIONS AND RESULTS

5.1 Simulation environment

We run several simulations using NS-2 (Issariyakul and Hossain, 2008) and AODV source code. The parameters used in these experiments are shown in Table 4.

<table>
<thead>
<tr>
<th>Simulator</th>
<th>NS-2 (vers.2.35)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation Time</td>
<td>100 seconds</td>
</tr>
<tr>
<td>Transmission range</td>
<td>250m</td>
</tr>
<tr>
<td>Link bandwidth</td>
<td>2Mbps</td>
</tr>
</tbody>
</table>
Table 4. Simulation Parameters

<table>
<thead>
<tr>
<th>Traffic type</th>
<th>CBR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data payload</td>
<td>512bytes</td>
</tr>
<tr>
<td>Delay between packets</td>
<td>0.25s</td>
</tr>
<tr>
<td>Number of nodes</td>
<td>50 nodes</td>
</tr>
<tr>
<td>Area</td>
<td>1000m×1000m</td>
</tr>
<tr>
<td>Node Mobility</td>
<td>between 0 and 20 m/s.</td>
</tr>
<tr>
<td>Mobility model</td>
<td>Random Direction Waypoint model</td>
</tr>
<tr>
<td>Misbehaving nodes</td>
<td>Chosen randomly</td>
</tr>
</tbody>
</table>

5.2 Experimental results and discussion

To test the feasibility of the attack, we consider malicious nodes colluding in order to unfairly praise themselves while trying to decrease the reputation of actually benevolent ones. The results obtained after performing such experiments can be observed in Figure 2 where reputation values evolution has been observed according to five scenarios, the first one without any colluder and the last four scenarios with an increasing percentage of colluders (this percentage is obtained according to recommending nodes). Let’s note that in this simulation, the victim is a well behaving node. Hence, without colluders, reputation is stationary. In fact, it starts with value “-1” and is increased until reaching a stable value which is in this case “3”. This is due to the fact, that reputation is modified progressively based on watchdog and received recommendations. Moreover, when there is an attack, the reputation value decreases dramatically. More precisely, as the ratio of colluders with respect to total number of nodes increases, reputation values decrease. This decrease is not abrupt neither sudden because the watchdog value is more important when calculating a node’s reputation.

Reputation values in Figure 3 are sensibly lesser when watchdog isn’t used even without any colluder. This is due to the fact that each node performs its own observations of its neighbors and that this appreciation is pondered by its weight when it is recommending another node. The weight itself is obtained thanks to the appreciation of node receiving the recommendation and can be zero if the node hasn’t observed this node yet.

The same simulations were achieved in presence of colluders and results are shown in Figure 4. We note that the reputation values calculated based on recommendations only are lesser to the one calculated based on watchdog and recommendations. In fact, when the node has its own observations
even in presence of colluders, the calculated reputation will be more realistic. However, when only recommendations are used, the reputation will be more faked.

![Figure 4: Variation in reputation value in presence of colluders](image1)

![Figure 5: Colluder detection rate varying with the number of colluders](image2)

A second experiment was carried out to determine the colluders’ detection rate of our proposal. As shown in Figure 5, the detection rate of colluders is around 100%. For example, when the percentage of colluders is 10% we detect 100% of them where when their percentage is 90%, we detect around 80% of them. Considering these results, our detection algorithm can be considered as efficient.

In order to verify the efficiency of our detection, we evaluated the number of false positive detected varying with the number of colluders. Figure 6 depicts the obtained results. One can see that the false positive rate increases with the number of colluders but remains very negligible compared with colluders’ number.

The last experiment was conducted in order to evaluate the performances of our proposal in terms of induced overhead as well as energy consumption. Figure 7 depicts the measured overhead for AODV and AODV enriched with our trust scheme and thus using 50 nodes. We can clearly see that the added overhead due to our trust scheme is acceptable.

![Figure 6: False positive rate varying with real number of colluders](image3)

![Figure 7: Packet overhead compared with total transmitted packets](image4)

Finally, Figure 8 depicts the energy consumption of our scheme compared with AODV. The results show that whether for AODV or our trust scheme, the energy consumption is sensibly the same: added messages and interactions don’t drain nodes batteries more than a simple routing protocol without any security mechanism.
Obtained results demonstrate that our proposition can improve significantly reputation based scheme credibility. In fact, it allows colluders detection and isolates them with an acceptable false positive rate. Moreover, the performances of our proposal in terms of routing overhead and energy consumption are sensibly similar to the original routing protocol (without any modification). Although, a reactive protocol was used in these simulations, our proposal is routing protocol independent and the same results can be obtained through the use of a proactive one, too.

6 CONCLUSION

In a collaborative environment such as MANET, evaluating the reliability of nodes based on how trustily a node acted in the past is vital. Trust Management can be used to promote such healthy collaboration. Unfortunately, this process is faced with various threats especially in MANET.

This paper dealt with one of these threats, the collusion attack. Using this attack, malicious nodes try to drive down the reputation of other reliable nodes by sending fake recommendations during the reputation calculation.

Our proposition is then, the detection of colluders as well as the punishment of such behavior. Detection is made by assessing recommendations variance compared with the recommendations average. The punishment process discards colluders from the routing process. Simulations results demonstrate the feasibility and the efficiency of our proposal. First, the colluder detection rate varying with the number of colluders was evaluated and obtained results shows that this rate is included between 100% when the number of colluders is lesser than 20% of the total number of nodes in the network and 80% when the number of colluder is approximately around 90% of the network nodes.

Second, false positive rate, overhead and energy consumption were evaluated. Obtained results are encouraging since the false positive rate is lesser than 10%, the overhead and energy consumption are close to the ones generated by AODV without any addition.

In future works, we intend handling more complex attacks such as Sybil camouflage, whitewashing attacks, etc.

References


MANAGEMENT ISSUES FOR BRING YOUR OWN DEVICE

Martin Brodin, University of Skövde, martin.brodin@his.se
Jeremy Rose, University of Skövde, jeremy.rose@his.se
Rose-Mharie Åhlfeldt, University of Skövde, rose-mharie.ahlfeldt@his.se

Abstract

Bring Your Own Device (BYOD) is an emerging research area focusing on the organisational adoption of (primarily mobile) devices used for both private and work purposes. There are many information security related problems concerning the use of BYOD and it should therefore be considered an issue of strategic importance for senior managers. This paper presents a systematic literature analysis using a BYOD strategic management framework to assess developing research trends. The analysis reveals early work in the analysis and design aspects of BYOD strategies, but a lack of research in operationalizing (planning, implementation and evaluating) strategy – the action phase. The resulting research agenda identifies twelve management issues for further research and four overall research directions that may stimulate future research.

Keywords: BYOD Bring Your Own Device, information security management, strategic management.

1 INTRODUCTION

During the last decade it has become commonplace for employees to have dual-use computing devices (devices used both at home and at work) - often for a mixture of private and professional purposes. One popular way of labelling this trend is Bring Your Own Device (BYOD). A recent survey indicates that 95% of companies allow employees some use of their own devices, that 36% offer full support for all employee-owned devices, and that 48% support selected devices (Barbier et al., 2012). Harris et al. (2012) report that one third of organisations allow privately owned devices (a result confirmed by Disterer & Kleiner (2013)) - and another third deploy company-owned dual-use devices. Some large companies sanction extensive BYOD programs; for instance Intel’s program involves 10,000 personal devices (Miller & Varga, 2011). The use of privately owned devices may also be widespread in companies that do not sanction them. Harris et al. (2012) report that 36% of employees ignore company policy and choose to use the device they feel appropriate. BYOD is predicted to be ubiquitous in the near future (van der Meulen & Rivera, 2013).

Though dual-use of devices is widespread, the term BYOD covers several different interpretations in the literature. BYOD implies that the employee owns the device and transports it to the workplace, a phenomenon associated with consumerization (Niehaves et al, 2012). However it may be more common for companies to supply consumer devices (for example a mobile phone) and allow home use (Oliver, 2012). Dual-use also implies that the device is used for a variety of work and personal tasks, implying shared or duplicated data, software and network connections. Where the device is used at home it may be connected to the computing environment of the workplace (Stevenson, 2012), and to external third party services. The nature of the device may be less significant than the extension of access to webmail, cloud services and content management systems (Morrow, 2012). BYOD in this study refers to computing devices which are mobile (used in the office and outside it, including the home) and/or dual-use (used both for professional and private purposes), whether provided by the employer or the employee.

The rapid spread of BYOD probably has many causes, including the popularity of mobile devices, efficiency gains for users in synchronising home and work resources, and productivity gains for employers in the expansion of the work sphere and better integration of information resources.
Employers may hope to transfer some of the device costs to their employees, or use the devices as attractive perks. However, both IT managers and information security experts express concern (ReadWrite, 2013; Intel, 2012). Whereas most information management approaches strive for standardization, consolidation and reduction of complexity (Disterer & Kleiner, 2013), widespread adoption of BYOD implies reduced standardization and increased complexity. There are major problems concerning integration with existing infrastructures, device support, and increased exposure to a variety of information security hazards, such that BYOD should be considered an issue of strategic importance for information security managers - and probably also for the senior managers of information-dependent organisations. Research indicates the importance of choosing an appropriate model for governance and support (Barbier et al., 2012). Strategic management of BYOD covers both the determination and execution of policy.

An early, but rapidly accelerating literature studies these phenomena, so that the management of BYOD may be considered an emerging research area. The objective of this article is to investigate how this literature deals with these issues using literature study techniques. We will address the following research questions:

- RQ1: Which managerial issues are highlighted in the emerging literature?
- RQ2: What are the research gaps in the early BYOD literature, from a strategic management perspective?

The paper is structured as followed. In section 2 the research method and analysis model are explained. Section 3 presents the analysis of the literature according to the model. Finally, section 4 gives the results and conclusions of the analysis, and offers directions for future research.

2 Research method

The search for relevant literature in this review was derived from Webster and Watson’s (2002) structured approach for determining the source material. These were the principal steps:

1. An extensive literature search using the WorldCat search engine with the search terms: Bring Your Own Device, BYOD, BYOT, BYOS, Bring Your Own, office-home smartphone, smartphone+information management, smartphone+policy, personally owned, consumerization, shadow IT and mobile computing, in combinations with information management, policy, security management, private, privacy, user-driven and dual-use. The search was filtered for peer-reviewed articles in English. This step resulted in 2865 article abstracts.

2. Manual screening for relevance (where relevance requires that the article both falls within the mobile/dual-use definition and focuses on policy, management or strategic issues, rather than technical issues). The articles were screened first by reading the abstracts. This screening removed many articles where BYOD had a different meaning (for instance a term in chemistry), articles which were tangential to the theme of the paper (for instance concerned with pedagogies and BYOD) and articles dealing with primarily technical issues. The remaining articles were downloading in full text and screened again, resulting in 69 unique articles.

3. Backward chaining by reviewing the citations in the articles identified as relevant in step 2. This step revealed many white papers and non peer-reviewed articles but only one new article.

4. Complementary forward chaining search in Web of Knowledge, Academic Search Elite, ScienceDirect, ACM, Emerald, Springer, IEEE and Wiley. This revealed 15 new relevant articles, leaving a total of 85 articles as the literature selection.

The search was considered complete since the complementary searches revealed few new articles of relevance.
2.1 Analysis framework

Webster and Watson (2002) also require that a literature review be concept-centric, where the concepts determine the ‘organizing framework’ of the review. Concepts may derived from the analysis, but a common practice is to adopt a suitable conceptual framework from the literature. The chosen BYOD management framework (Brodin, 2015) is adapted from Jonson and Scholes (1997) seminal work on strategic management, and the international standards ISO/IEC 27001 (2013) and ISO/IEC 27002 (2013) Information Security Management Systems (ISMS). The three main categories in the model are analysis, design and action.

Analysis concerns the assessment of opportunities and threats involved in the adoption of BYOD, where expectations refer to the opportunities in the form of BYOD benefits that are targeted, environment points at BYOD threats originating from outside the organisation (for example information security threats) determined through risk assessment, and resources and capability indicate the organisation’s current ability to realise benefits and mitigate environmental threats.

Design concerns the development of strategic information and security governance strategies or policies for BYOD, where options represent distinct strategic directions, development refers to the adaptation and enumeration of options, and selection refers to choosing the appropriate strategy or policy.

Action concerns the operationalization of the chosen BYOD strategy, where planning precedes the policy implementation, and evaluation is carried out to determine the success of the BYOD strategy and its implementation.

Most articles in the literature selection covered several of these areas, but in table 1 they are classified according to their primary, or dominant purpose.
Management issues for BYOD

In this section the principal management issues investigated in current BYOD research are analysed.

3.1 Analysis

Analysis concerns the assessment of opportunities and threats involved in the adoption of BYOD, including expectations, environment, and resources and capability.

3.1.1 Expectations

Researchers point to many expectations for BYOD related to benefits for both employee/users and management. The main expectations are for increased personal productivity, flexibility of time and place and increased user satisfaction.

IT Managers rank increased personal productivity as the most important expectation for BYOD (Intel, 2012). The BYOD-program at Intel reports that personal device users saved on average 47 minutes per day, amounting to more than 2 million hours per year (Miller & Varga, 2011). iPass (2011) claim that a dual use mobile user works longer than other employers: 240 hours more per year. In cash terms, productivity benefits may amount to between $300 and $1300 per year per employee, depending on job role (Barbier et al., 2012). One reason for increased productivity may be that employees who are able to satisfy their psychosocial needs at work invest more of themselves (Kahn, 1990; Pfeffer 1995). However many of the existing studies of BYOD benefits are sponsored by large industry players (Intel, Cisco) with vested interests in promoting BYOD, and these results should be treated with caution.

BYOD increases flexibility of time and place, allowing employees to work outside the office and normal working hours. Some managers expect this to be the most significant BYOD benefit (Singh, 2012). One way this can be measured is by monitoring business related emails and access to corporate resources from non-corporate devices after office hours. Harris et al. (2012) refers to a study where 14% of employees connected to corporate resources after work hours and 22% used a private mobile phone to check corporate emails before they went to bed. Logically BYOD also helps the employee to...
manage their personal affairs from the office, but this is not investigated. Increased freedom to manage personal work in terms of time and place may have positive health effects (iPass, 2011). However constant work availability facilitated by BYOD is implicated in increased personal stress (Green 2002), and the extension of work into home life, may affect family relationships, for example the amount of time spent with children (UNICEF 2014).

A secondary expectation for BYOD is increased user satisfaction (Miller & Varga, 2011). This is associated with the convenience of reducing the number of devices; for example one mobile phone for both private and business use (Disterer & Kleiner, 2013). Harris et al. (2012) report that many users enjoy having advanced technology devices at work and home, but relatively few believe that it contributes significantly to work satisfaction.

3.1.2 Environment

In the BYOD literature the organisational environment is largely perceived as an information security threat, in which dual use devices are at greater risk. Threats are assessed through risk assessment, and increased risk stems from user behaviours and expectations for their devices, particularly when they also use them for personal purposes and consider that they own them. Thus the principal managerial issues for BYOD in relation to environmental threats are data control and device protection.

A major BYOD concern is control of corporate data, especially where data is stored outside company premises, when the device that it is stored on is lost or stolen, or if the employee leaves the company. Particularly difficult is the question of who is accessing corporate data, since BYOD devices (which may connect to confidential corporate data systems) are seldom physically secure, and may be attached to multiple networks. A company owned device can be retrieved when an employee leaves, or remotely wiped if it is stolen. The data, if stored, may be encrypted, and the company’s information security policies enforced by the IT department. Even with these precautions, sensitive corporate data is routinely recoverable from second-hand hard disks (Jones et al. 2012). Dual use device owners tend to resist the installation of encryption and remote wipe software (or other kinds of software associated with managerial control) since they consider that it encroaches on their privacy (Pettey & Van Der Meulen, 2012). Only a third of private device owners use encryption for company data (Camp, 2012). Private device owners freely install software of their own choice and join networks other than the company’s protected network. If it is too complicated to access the secure network, users may go for the less secure guest network instead (Kehoe, 2013). They may store data on multiple hard disks, including their private cloud (Dropbox, OneDrive, iCloud, Google Drive). A particular problem arises when the key or password protecting the data is personal, whereas the data is corporately owned (Walters, 2013). How can this data be monitored and audited? An employee leaving a company takes their privately owned device with them – how does the company ensure that sensitive corporate data is removed?

A related managerial issue is protection of BYOD devices, since devices storing sensitive corporate data are routinely lost, stolen or hacked (Wilson, 2012). If the IT department does not control the device they cannot force operating system updates or ensure that the antivirus program is up to date (Morrow, 2012). Most private users have poor protection habits: they do not update software regularly (Skype et al. 2012), or use the auto-locking facilities provide for them. Researchers expect those behaviours to remain when their device is used for work-related purposes (Disterer & Kleiner, 2013). Camp (2012) estimates that “less than half of all devices in the BYOD category are at greatest risk from the most basic of security measures”. Users should back-up their own devices (Wong 2012) since the organisation cannot be responsible. IT managers are thus required to protect corporate data they may not even control (Walters, 2013). Faced with non-standard devices and non-compliant users (Tokuyoshi, 2013) they may give up. Difficulties in supporting security, encryption and remote wipe are the most common explanations for not restricting BYOD use.

3.1.3 Resources and capability

Resources and capabilities represent the organisation’s current ability to realise benefits and mitigate environmental threats from BYOD. Two significant managerial issues here are awareness and support.
Awareness describes an organisation’s capacity to monitor and react to the BYOD threats in its environment. Allam et al. (2014) propose a model for smartphone information security awareness based on accident prevention techniques. The model is designed to help monitor the information security position and tailor security policies and procedures to threats. However Ashenden and Lawrence (2013) believe that awareness programmes are limited and their effect on behavioural change doubtful. Instead, they propose a social marketing framework that will be more effective. They identify the user behaviour they want to change, analyse why users exhibit those behaviours, identify benefits for users from potential change which increase security, design an intervention, and evaluate the impact.

A significant resourcing and capability issue for BYOD is support. BYOD devices run many operating systems on many platforms, with diverse software. IT managers anticipate many compatibility problems with existing IT infrastructures (Intel, 2012). However, users expect the same level of support they had with their standardised company-owned devices (Brooks, 2013). IT professionals experience the frustration of increasing support costs and administration time, which reduces productivity in other areas (Walters, 2013). Intel (2012) claims that BYOD comes with no impact on support and with relatively low cost (Miller & Varga, 2011). Organisations that transferred purchase costs for devices to their users saved some money. However Harris et al (2012) report that these savings were eaten up by the increased cost of managing the IT environment.

3.2 Design

Design concerns the development of strategic information and security governance strategies or policies for BYOD, where options represent distinct strategic directions, development refers to the adaptation and enumeration of options, and selection refers to choosing the appropriate strategy or policy.

3.2.1 Options

Strategic options represent different choices that managers have in relation to the adoption of BYOD, where the two extreme positions are (i) to forbid any kind of dual use device, and (ii) to allow each and every form of BYOD without restrictions. Mourmant et al. (2013) do not examine BYOD as an independent option, but as part of their model for intrapreneurial freedom; BYOD is part of freedom of materials and resources. Harris et al. (2012) present a model for IT consumerization with 6 strategic options that range from strict (tight control, few standard devices) to complete freedom. The only option that allows privately owned devices is laissez-faire, where management allow external devices and applications without any restrictions. However no research advocates this strategy, although some researchers and standards discuss trade-offs and the acceptance of risk. Holleran (2014) proposes a compromise option, where BYOD is prohibited, but in return employees are allowed to use their mobile devices for personal purposes. Another way of developing strategic options is through analysis of the managerial control space. Yang et al., (2013) proposes a risk management quintet, which looks at the mechanisms for technology adoption, control, liabilities, user perception, and user behaviour.

One prominent article genre in this category was the opinion piece from an acknowledged industry expert (e.g. Millard (2013); Steiner (2014); Thielens (2013); Walker-Brown (2013)). Though apparently peer reviewed, these articles are based on personal experience and do not display any conventional research method. They are not considered further here.

3.2.2 Development

Regardless of choice of strategic option, there is universal agreement that the first development step is information security policy update (Oliver, 2012; Harris et al., 2012; Wong, 2012; Gatewood, 2012; Caldwell, 2012; Simkin, 2013; Montana, 2005; Vickerman, 2013; Yang et al., 2013). Though these researchers identify the policy as central to the success of BYOD, research in the information security management field indicates that policies are often broken. Younger people seldom obey information security policies (Simkin, 2013), though more than half of IT professionals believe they do. Users have poor understanding of policies (Oliver, 2012; Wong, 2012), if they are even aware of them.
Consequently, it is not enough to update a policy; it must also be communicated (Wong, 2012; Gatewood, 2012; Oliver, 2012).

Wong (2012) points out the need for users to understand the regulatory framework: for example which information is owned by the organisation and which is personal information that they may freely use. This problem is compounded by role confusion: when, and in what situations, is a user acting as a private person, and when they are acting as an organisational representative. Is it acceptable to post sensitive corporate information on a social network where you are profiled as a private person, or if you are no longer working for the company? Other central aspects in the development area are risk assessment, clarification of ownership of information, right to audit, privacy rights, security of business information, and registration of assets (Vickerman, 2013).

3.2.3 Selection

BYOD strategy decisions should be made by the appropriate people in the organisation after weighing benefits against information security risks: the business/security balance. Ring (2013) identifies organisations that gave BYOD both green and red lights after evaluating the risks. He concludes that the choice is ultimately “a business decision, not a security decision”. Borrett (2013) agrees, arguing that senior management target increased flexibility and/or cost-savings. Mooney et al. (2014) suggest that the entire c-suite (chief executives) should be involved in the process. Guinan et al. (2014) disagree, arguing that, depending on the organisation, the process may be top-down, middle-out or bottom-up, and that knowing where and with whom to begin may be the key to success. Silic and Back (2013) identify two must-win areas when selecting a strategic option: mobile strategy and security framework. Furthermore, they argue that stakeholder support is critical, both for making the change and for rooting new information security procedures in the culture.

3.3 Action

Action concerns the operationalization of the chosen BYOD strategy, where planning precedes the policy implementation, and evaluation is carried out to determine the success of the BYOD strategy and its implementation.

3.3.1 Planning, Implementation and evaluation

BYOD is a relatively new phenomenon, and few researchers directly address the action phase. Those that do, agree on the need for training. Walters (2013) focuses on the human and informational, rather than technologies, since a lot of the traditional layered approach to enterprise security do not apply anymore. What definitely apply is the human layer with information security education and awareness. Furthermore, Walters (2013) state that functional and organisational roles for data access must be determined before a BYOD implementation can start.

Gatewood (2012) emphasises information security training for all employees and points out that a forgotten and unlocked phone can lead to a disaster. The technical mechanisms are not worth as much if employees do not comply with the BYOD strategy and policies. Studies indicate that proper security training must be in place to get employees to adopt the new strategy (Hu, 2013; Markelj & Bernik, 2012).

When the policies and procedures are implemented it is important to evaluate opportunities and threats with respect to organisational context to determine if an update is needed (Niehaves et al, 2012).

4 DISCUSSION AND DIRECTIONS FOR FUTURE RESEARCH

Two research questions were posed for this review. In response to the first question: What managerial issues are highlighted in the emerging literature on BYOD, twelve issues were identified (represented in italics in the next sections).

Managerial expectations for BYOD include increased personal productivity, time/space flexibility and increased user satisfaction. These benefits coincide with expectations for mobile devices in general,
and researchers need to understand what the specific impact of dual use, personal ownership and personal choice of device have on these outcomes. In addition the methodological approach of this research requires more consideration, and there is a need to separate independent research investigations from those of major industry players. Researchers should also establish costs (and particularly hidden costs) of BYOD programs which may result from infrastructure integration, support and extra information security demands, the costs of information security breeches, and employees organising their personal affairs in work time, amongst other things. There is also need for research into employees’ dual use (home and work) patterns.

Environmental information security threats highlighted the need for improved approaches for information control and device protection. Many of these threats are known in the mobile security field, and researchers need to understand how (partial) loss of organisational control of information and devices, less standardization and transfer of responsibilities for protection/backup to users affect these threats. Important questions for researchers may be: which known threats are amplified by BYOD (and by how much) and have known responses that can be scaled up; which threats are amplified to the point where they can no longer be managed with known responses; and which threats are new and require improved management approaches. Many of these questions require empirical investigation and quantification. An unexplored question is whether there are information security threats that are reduced or removed by BYOD. A further issue that is not yet investigated is the effect of BYOD on employee privacy. Users have information rights (many of them are backed by law), as well as organisations.

Issues relating to organisational resources and capabilities include awareness and support. Organisational information security awareness may come to depend more on user-led reporting, manufacturer alerts and monitoring the information security communities. Patterns for support may change when there are many different devices and little standardization, with more reliance on users’ own capabilities and lower levels of information (which might also focus on information security guidance and instructions). Crowd-sourced solutions to these problems, with users doing much of the work themselves and IT professionals co-ordinating are not yet researched. A further issue needing investigation is information classification; this may facilitate many differentiated strategic options.

With respect to the design of BYOD strategies, researchers need to improve already established models of strategic options in order to complement the partial offerings available. Such models should offer differentiated BYOD strategies to managers, explaining the potential benefits, costs, risks and information security responses of different courses of action. Such strategic option models should be based on quantitative and qualitative evidence, with a theoretical departure point. Since development of new strategic positions involves an information security policy update, researchers may investigate how current information security standards (such as ISO/IEC 27000-series and methodological support for information security (MSB, 2015)) manage BYOD. However, at the same time they should investigate how the take-up of the information security regulatory framework as a whole can be improved, especially in the BYOD environment where users may perceive the regulatory framework as voluntary. Selection of options is based on the business/security balance. This is a particularly complex area for organisations and need to be researched, as it involves cross-disciplinary comparative assessments of benefits and risks, where neither organisation-wide benefits nor a complete empirically based picture of information security threats are yet available. Moreover the development of strategic options implies comparative assessments for several scenarios or contingencies. Most of the BYOD literature focuses on personal productivity, and its influence on team communication, group work, customer management, and at the organisational level information flow, workflow and management communication are not yet studied. Managers should understand what they could expect to achieve for their organisations with BYOD programs. Organisations with structured information security programs already in place are better placed to handle emerging BYOD difficulties. However many organisations lack information security classification and security risk management that might provide a firmer foundation for strategic decision-making. Managers should also be helped to understand the scale of risk to which their organisations are exposed by authorised (or unofficial employee-led) BYOD programs.
The action or implementation of strategies is not much investigated in current BYOD literature (see below), but researchers can translate BYOD compliant information security standards and methods into training materials and contribute more effective learning strategies.

Research directions for BYOD management issues are summarized in table 2.

<table>
<thead>
<tr>
<th>Framework category</th>
<th>BYOD management issues</th>
<th>BYOD research agenda</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Analysis</strong></td>
<td></td>
<td>benefits and costs should be established empirically by independent researchers using methodologically sound techniques.</td>
</tr>
<tr>
<td>expectations</td>
<td>1. increased personal productivity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. time/space flexibility</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. increased user satisfaction</td>
<td></td>
</tr>
<tr>
<td>environment</td>
<td>4. information control</td>
<td>cataloguing of known mobile information security threats and responses for BYOD area, and identification of new threats and responses; protection of employee privacy.</td>
</tr>
<tr>
<td></td>
<td>5. device protection</td>
<td></td>
</tr>
<tr>
<td>resources and capability</td>
<td>6. awareness</td>
<td>investigation of distributed and user-led information security awareness and support; information classification.</td>
</tr>
<tr>
<td></td>
<td>7. support</td>
<td></td>
</tr>
<tr>
<td><strong>Design</strong></td>
<td></td>
<td>improvement of normative models of strategic options based on empirical evidence and theory</td>
</tr>
<tr>
<td>options</td>
<td>8. strategic options</td>
<td></td>
</tr>
<tr>
<td>development</td>
<td>9. security policy update</td>
<td>development or improvement of policy and regulatory frameworks from existing information security standards and methods, and investigation of improved user compliance</td>
</tr>
<tr>
<td></td>
<td>10. regulatory framework</td>
<td></td>
</tr>
<tr>
<td>selection</td>
<td>11. business/security balance</td>
<td>cross-disciplinary comparative assessments of organisational benefit and information security risk</td>
</tr>
<tr>
<td><strong>Action</strong></td>
<td></td>
<td>(under researched area requiring further investigation)</td>
</tr>
<tr>
<td>planning</td>
<td>12. training</td>
<td>materials, methods and tools for communicating and disseminating regulations within organizations, (under researched area requiring further investigation)</td>
</tr>
<tr>
<td>implementation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>evaluation</td>
<td></td>
<td>(under researched area requiring further investigation)</td>
</tr>
</tbody>
</table>

Table 2. Research directions for BYOD management issues

In response to the second question (what are the research gaps in the early BYOD literature, from a strategic management perspective), the current distribution of research over the BYOD management framework (Brodin, 2015) is skewed. Table 2 shows that the largest part of the research concerns strategic analysis (expectations, environment, resources and capabilities), where the majority deals with information security threats. A smaller proportion concerns strategy design, with many recommendations based on experiential evidence and a widespread concern with information security policies. Much less research covers action – the operationalization of strategy phase. One reason for this absence may be that BYOD is an emerging phenomenon, so there are relatively few well-designed implementations to investigate. Another possibility is that BYOD presents relatively few new strategic challenges, and can be managed with incremental changes to information management and mobile security strategies within existing frameworks. Regardless, this still has to be investigated. Therefore it seems necessary to take the following steps to provide sound research that is helpful to practitioners.
1. Ground BYOD research in existing mobile security research in order to specify what can be inherited from existing research and what the new parameters are, such as ownership, decreased standardization etc.

2. Develop theory-based strategic options frameworks with suitable research methods (for instance design science).

3. Focus on strategic action (planning, implementation, evaluation) research by encouraging the empirical investigation of BYOD implementations using case studies, action research, and other qualitative methods, supplemented by for instance quantitative evaluation methods.

4. Encourage cross-disciplinary research to broaden the base of the research beyond the information security communities (see Györy et al., (2012)).

5 CONCLUSIONS

In this article we investigated the emergence of the widespread empirical phenomenon of Bring Your Own Device in research literatures. BYOD is linked to consumerization, as computing devices for personal use become widespread in affluent societies. Much of the research discussion is located in the mobile security research area, since data and device security is a major concern. There are technical strands of research (for example in chip design); however we chose to focus on the managerial implications of BYOD for companies. Though BYOD is difficult to separate from other aspects of dual use computing, two aspects of BYOD may become crucial for the development of computing in organisations. The first is the shared understanding that the user owns their device (regardless of who actually pays for it); the second is the consequent understanding that they have free choice – of device, the software that they install on it, and what they use it for. These factors effectively move the locus of control of the device (and the information accessed by it) away from the organization and towards the individual employee - a change widely assumed to be unstoppable and non-reversible. Such changes often require a strategic response from organisations. We analysed 85 articles focusing on these phenomena using a framework developed for the purpose from the strategic management and security standards literature. We identified 12 BYOD core management issues addressed by the literature and provided a focused research agenda for each of these existing issues. We also analysed prominent gaps in the literature and identified four overall research directions which can help address those gaps. The twelve management issues, together with these four overall research directions provide a basis for a stimulating and useful programme of research.

References


Kehoe B. 2013. 'BYOD - Proceed with caution'. Hospitals and Health Networks. 87(6): 17.
Oliver, R. 2012. 'Why the BYOD boom is changing how we think about business it'. Engineering and technology. 7(10): 28.
Pettey, C. and Van Der Meulen R. 2012. 'Gartner identifies three security hurdles to overcome when shifting from enterprise-owned devices to BYOD'. Gartner Inc. URL:  http://www.gartner.com/newsroom/id/2263115 (visited July 2013).


Vickerman, J. A., 2013. 'Managing the Risks of BYOD With the line between work and home increasingly blurred, companies must establish a policy that embraces progress'. Risk Management, 60(1): 38-41.


A FRAMEWORK FOR DESIGNING AN EFFECTIVE SECURITY AWARENESS SYSTEM FOR ONLINE SOCIAL NETWORK USERS

Ehinome Ikhalia, email: ehinome.ikhalia@brunel.ac.uk, Brunel University, UK
Alan Serrano, email: Alan.Edwin.Serrano-Rico@brunel.ac.uk, Brunel University, UK

Abstract

Online social networks have become the most susceptible platform for spreading malware. Malicious users are often successful in exploiting social networks because of the clustering structure of ‘friends’ connection and trust among legitimate users to spread their malware. Additionally, due to lack of effective security awareness, many users are now unknowingly becoming active partakers in the distribution of various malicious softwares growing exponentially. Despite efforts to create effective security awareness programs by platform providers and practitioners; some research works have criticised existing awareness techniques due to their sole emphasis on the content of their information security programs. In this work, we examine existing security awareness techniques used for the avoidance of malicious IT threats and identify the issues hindering effective security awareness for social network users. This paper argues that due to the unique risks faced by social network users, the diverse end-user functionalities and the deceptive characteristics of social engineering malware attacks executed through them, there is a need to develop mechanisms tailored at helping users avoid malware attacks carried out through online social networks.

Keywords: Information security, social networks, cyber-security awareness, malware attacks.

1 ONLINE SOCIAL NETWORKS AND CYBER THREATS

Online social networks (OSNs) are growing exponentially, becoming the most popular and cheapest communication medium for Internet users (Cheung et al, 2011). Concurrently, malicious users are exploiting the lack of security awareness of users by luring them into downloading malware on their computing devices (Fire et al, 2013; Luo et al, 2009). In the last decade, many research works have been concerned with the development of anti-malware softwares to solve the growing challenges of malware distribution through online social networks. However, many of these solutions have not been effectively evaluated to measure their intended objectives. Due to the growing success of social engineering malware attacks carried out daily on social networks, security analysts are now suggesting that end-user awareness could be the main solution in dealing with acts of malicious deception (Beye et al, 2010; Faisal & Ibrahim, 2013; Hasan & Prajapati, 2009; Sood, 2011).

The implications of malware attacks increasing through social networks have bizarre consequences for end-users and business organisations (Almeida, 2012; Makridakis et al., 2010; Perez, 2013). Some of the implications, includes, financial loss, reputational damage, and loss of important data (Thomas & Nicol, 2010). The most dangerous social network malware attack was first identified in 2008. The malware popularly referred to as - Koobface, targeted users of Facebook, Twitter and MySpace. The mode of operation of the malware involved luring end-users to click on fake links to online videos or pictures. Thereafter, the users were encouraged to install a prospective ‘Adobe Flash Update’, but are in fact malware (Thomas & Nicol, 2010). Cyber security reports suggests that since the evolution and massive success of the Koobface malware attack, malicious users are increasingly adopting its characteristics to lure more unsuspecting legitimate users to download malware into their computing devices for various malicious intentions (Boshmaf et al, 2013).

Unfortunately, platform developers and other stakeholders have focused on technical and procedural security measures while implementing their information security solutions (Abawajy, 2014; Furnell et al, 2002). Conversely, from information systems (IS) perspective, this work argues that solely relying
on technical security solutions is insufficient in protecting end-users due to the escalating amount of malware attacks executed via the complex functionalities that OSNs offers.

Some research works have suggested the prospective of using non-technical mechanisms in order to enable OSN users avoid malicious attacks (Al-Qasem et al, 2013; Faisal & Ibrahim, 2013). While technical mechanisms advocate the use of technical tools or integrating a variety of technologies to address malware threats; non-technical methods supports the adoption of a well-developed user awareness system for end-users (Arachchilage & Love, 2014).

According to (Aloul, 2012), an effective information systems (IS) security requires that users are aware of the threats facing the systems they use both individually and within an organisational setting. In addition, we posit that an effective security awareness program needs to be tailored to reflect the context or situation of the system through which the malicious threats occur. This paper focuses on identifying the factors that needs to be considered in developing an effective security awareness system for OSN users. To the best of our knowledge, this work is the first attempt to develop a security awareness system tailored for users of OSNs.

The paper is structured as follows;

Section 2 describes the malicious threats of online social networks. Section 3 describes suggested guidelines for developing an effective information security awareness system. Section 4 explains the research methodology adopted for this work. Section 5 presents the analysis of the systematic review on existing security awareness systems. Section 5.1 presents the limitations of existing security and their unsuitability for social network users. In section 6 describes the proposed framework for developing a security awareness system for users of online social networks. Finally, possible areas for future research are highlighted.

2 ONLINE SOCIAL NETWORK THREATS

Online social networks (OSNs) are emerging as one of the most significant sector in Web 2.0 community today. Individuals and business organisations are now integrating this platform as a central factor of their social and business ecosystem (Ikhalia & Imafidon, 2013). The huge benefits that social networks provide for individuals and organisations are increasingly being faced with malicious threats; which threatens its existence (Furnell, 2010). Users of social networks have the natural tendency to trust their co-users and the applications that run on the platform. Due to these relative “trust-based” relationships, malicious users are massively exploiting the platform to deceive legitimate users into downloading malware to their systems. Mansfield-Devine (2008) and Williams (2008) argue that in a trusting environment, user awareness plays an important role in their security.

To a large extent, the future existence of OSNs will depend on the ability of platform providers to ensure the provision of more pleasurable functionalities without declining the security of their end-users. To effectively examine the need for a security awareness framework for online social network users, we describe three key threats faced by end-users and their potential implications in the following sub-sections.

2.1 Threat 1: Breach from ‘friend’ connection

One of the fundamental features of OSNs is the enabling of connections and communication among users who share similar interests, philosophies, family ties et al. Prior to a ‘friendship’ connection between two users, social network providers e.g. Facebook, usually protect the privacy of users by default. However, the platform Facebook now allows the ‘follower’ type of connection, which implies that a user can have access to posts made by another user by becoming a follower (Livingstone et al, 2014). Using Facebook as an example, when two users become friends, the platform allows them to access the personal information uploaded on their profiles as well as other activity performed on profiles of their unique connections (Yang et al, 2011).
Malicious users have developed new techniques in making ‘friendship’ requests with the aid of automated scripts. This mechanism easily allows for stealing personal information by ‘befriending’ enormous users. For example, 75,000 out of 250,000 random Facebook users sent ‘friend requests’ using automated scripts unknowingly accepted the bogus request (Gao et al, 2011). Studies conducted by Bilge et al, (2009) demonstrated more sophisticated attacks. Firstly, the researchers identified same-site profile cloning, a process whereby an attacker duplicates a user’s profile in the within the same OSN and then uses the ‘duplicate’ to establish ‘friendship’ connection to unaware connections of the victim. Secondly, cross-site profile cloning was also identified; a situation whereby the malicious users or attacker finds a user registered on social network A. Consequently, the attacker then replicates the profile to social network B, where the victim is not registered and establishes a ‘friendship’ connection targeted at the victim’s already registered social network B.

One of the major implications of this type of threat faced by social network users’ is severe defamation of reputation in an irreparable manner. Bilge et al, (2009), added that to resolve this type of threat a system of user-awareness needs to be exclusively designed to improve the security awareness of social network users.

2.2 Threat 2: Breach from third-party applications

The expansion of third-party applications is due to many factors, such as the enormous end-user demand for more engaging features and desire of the platform providers to keep users on the platform. Facebook allows other developers integrate their applications on the platform either for free or a fee. Some research works have argued that third party applications integrated with Facebook do not undergo any form of security screening or testing which makes them relatively untrusted (Gao et al., 2011). Furthermore, it is mandatory for users to allow third party applications access to their personal information before they can install and use them. As a result of the inadequate measures to screen third party applications by social network providers, a huge opportunity is created for malicious users to develop third party applications in order to acquire personal information of unsuspecting users. (Singh et al, 2009). In a recent study on the interaction between 997 Facebook applications and ‘fourth parties’; the researchers found out that a staggering 22% of Facebook applications provide users' personal information to one or more fourth-party tracking entities illegitimately (e.g. trackers and advertisers) (Chaabane et al, 2014).

2.3 Malware attacks

The exponential growth of online social networks has led to upsurge in security issues such as malware attacks (Cosoi, 2011; Watters, 2011). Malicious users commonly referred to as ‘attackers’ often take advantage of security unaware users to distribute their malicious programs with various deceptive techniques (Abraham & Chengalur-Smith, 2010; Faghani, 2009; Leitch & Warren, 2009; Merrill et al, 2011; Sanzgiri et al, 2013; Tony Palmer, 2014). Attackers can spread worms and establish botnets more easily because of the amusing and high level of user interaction and engagement inherent with social networking activities. Malware can propagate over online social networks through profile, interaction & user engagement, and third-party applications. The Koobface worm is one of the most destructive worms reported on OSNs (Weir et al, 2011). The Koobface is the first malware to have a successful and continuous execution, proliferating through social networks. It executes an attack by automating Internet browsers to execute the following tasks:
• Register and activate a Facebook account
• Randomly Join Facebook groups
• Adding Facebook friends.
• Post messages on the friends’ walls that contain links to the Koobface loader module.

In 2009, a malicious worm targeted Facebook users, using cross-site request forgery to propagate itself by luring unsuspecting users to click on a malicious link (Doug, 2009). When a user falls for the lure, a web page is opened which triggers a sequence of other pages and pop-ups. Thereafter, the worm presents a form submission to the user, with a prompt to click on the “Share” button to confirm the posting. Consequently, all connections associated with the victim will see the malicious message as well as the link. Recent studies have reported the increased growth of characteristics of these types of malware attacks on Facebook (Kuma et al, 2013), and with the exponential expansion of this huge social networking platform, these form attacks are not likely to stop soon.

From the threats described, it is obvious that the structure of social networks is making it easier for malicious users to undertake their activities without the possibility of detection. It is unfortunate considering the efforts put into the development of anti-malware softwares to curb the excesses of malware attacks. Baskerville & Rowe (2012) argue that as functionality increases in IT systems, security threats facing users’ increases proportionately. However, it would be unrealistic to suggest the reduction of the features offered by OSNs in order to reduce malware attacks, because OSNs thrive by active user engagement which is also affected by the pleasure provided from engaging on the system via its multifarious functionalities. Hence, there is a need for a more proactive measure to undercut these threats facing the users of OSNs. Such measures should involve effective security awareness for end-users pointing out all the threat-based activities to the user and their likely implications. Nevertheless, before designing an effective security awareness system, it is vital to understand some theoretical and practical factors which need to be considered within such a system.

In section 3, key suggestions for designing effective security awareness systems from previous publications are described.

3 Methodology

Since security awareness in online social networks is a relatively new research area, there were no published articles on the development of any security awareness program for online social network users. Therefore, the authors embarked on a systematic literature review of publications on the development of cyber-security awareness programs between 2005 and 2015. The authors selected this range because between these periods, cyber-attacks experienced a sudden rise as well as concerns for security awareness (Ramim & Levy, 2006). The review was conducted between November 2014 and February 2015.

3.1 Inclusion criteria

The inclusion criteria used to select the publications analysed where based on the following elements

1. The publication mentions cyber-security threats which are relatively similar to threats faced by online social network users. Some of the threats mentioned are not exclusive to OSNs, but their method of execution exploits the structure and context of OSNs in order to be successful.
2. The publication explains the working process of the artefacts developed for end-user security awareness. This is important because in order to evaluate a security awareness program, the operational procedures must be known.
3. The targeted end-user group for which the security awareness systems were developed are clearly known. The authors consider this important because there are diverse groups of
internet users with varying interests. It is important to correlate the delivery method of an awareness program with the target users in order to evaluate its potential effectiveness.

4. The publication must either be a journal or conference article.

3.2 Databases

The following databases were searched:

- **IEEE Digital Library:** This is amongst the leading publishers of computer science and information systems research articles with top quality. This database was selected because it contains high quality technical literature in engineering and technology which have been published since 1998.

- **ScienceDirect:** This database is amongst the trusted source for journal articles by millions of researchers with over 13 million content pieces available.

- **Google Scholar:** Considered as the ‘database’ of all other databases, Google Scholar, provides a comprehensive and scalable access to conferences, journals, whitepapers, books in academia and the industry. It often assists researchers in accessing papers from the original source.

3.3 The Search Results

Research on the development of cyber-security awareness programs is relatively emerging, therefore, the authors were more specific in using an initial search term; ‘cyber-security awareness programs’.

Since cyber-security implies the collection of tools, technologies guidelines, risk management techniques and training that can be used to protect the assets of individuals and business organizations (Wamala, 2011). It was important to include the terms cyber-security, awareness and programs as it best describes the goal of the study and describes part of the definition of cyber-security.

The initial search using the search term ‘cyber-security awareness programs’ returned a total number of 13041 results (11,300 from Google Scholar, 1741 from ScienceDirect and 12 from IEEE Explore). Four publications appeared in the three databases and the duplicates were deleted. Out of the 11,300 results from Google Scholar, 15 were selected for review and 12 were included.

While from the 1,741 results found in ScienceDirect, 8 were selected for review and only 2 were included. The authors found 12 results using the same search term on IEEE Explore but none of them were found to be relevant. Therefore, the search term was modified to 'designing security awareness' which returned 386 results, from which 5 were reviewed and 4 were excluded based on the inclusion criteria.

Similarly the authors applied the search term 'designing security awareness' on Google Scholar and ScienceDirect databases which returned 64,300 and 7,212 results respectively. However, all the relevant results identified were previously found using the search term 'cyber-security awareness'. Hence a total number of 15 articles were included for the systematic review.

In section 5 below, a structured analysis of the key findings from the systematic literature review is presented.
This section discusses, some of the suggested factors that needs to be considered when designing effective cyber-security awareness systems (Abawajy, 2014; Aloul, 2012; Kirlappos & Sasse, 2012; Wu, Guynes, & Windsor, 2012). From the theoretical and practical suggested factors for implementing an effective cyber-security security awareness program found from previous academic publications, five factors have been considered based on the context of online social networks which are; end-user learning preference, time efficient, multi-media delivery mechanisms, non-technical means of communication, and contextual based approach.

4.1 End-user learning preference

According to the security standards council (2014), emphasis must be made on the content as well as the technique for delivering security awareness for different user groups within a technological setting. This factor has also been previously suggested by Peltier & Thomas (2005), who stated that the needs of the end-user must be considered to aid the effectiveness of security awareness programs.

4.2 Time efficiency

In their recent publication, Kim & Eyong (2014), recommended time efficiency for developing security awareness programs for college students. However, most organisations place more emphasis on delivering gigantic curriculums about security awareness, which often takes hours to disseminate (Bada, 2014).

4.3 Multi-media mechanisms

In today’s technological setting, the role of multi-media in enhancing education cannot be underestimated. Previous research works have evaluated the significance of multi-media systems in persuading the interests of targeted end-users have clearly proved the use of this method over other conventional educational methods of learning (Shaw et al, 2009).

4.4 Non-technical Communication

Rowe et al (2011), highlights that it is the role of information technology practitioners to ensure that users effectively and securely utilize systems through the design of usable user interface for providing security awareness. In addition, the authors supported the need to educate end-users on how to avoid malicious social engineering threats. One of the important issues raised by the authors is that cyber-security professionals do not necessarily need to educate end-users using technical security terminologies such as (cross-site scripting and man-in-the-middle attacks). These terminologies are best suited for system administrators. End-users only need to be aware of the characteristics of malicious threats using non-technical security terminologies.

4.5 Awareness That Addresses Threats Based on Context of Each Organisations’ Context (Contextual)

In a recent empirical study conducted by Abawajy (2014), the author investigated the preference of computer users concerning the delivery techniques of cyber-security awareness. The findings of the study show that despite the buzz around game based security awareness; most users prefer the use of videos for cyber-security awareness. In addition, the research recommended that further investigation needs to be conducted regarding the effects of security awareness through videos for different organisational context. This is due to the fact that the videos created for the study were relatively
generic and does not address the unique threats faced by different organisations and their technological settings.

According to (Abawajy, 2014), “Although many of the concepts included in cyber security awareness training are universal, such training often must be tailored to address the policies and requirements of a particular organisation.”

Figure 1. The results from the empirical study about preference for cyber-security awareness delivery methods. It shows that out of 60 voluntary participants, the percentage of participants who preferred games were 5%, 50% preferred videos, 33% preferred texts while 12% were undecided (Abawajy, 2014).

In their design of an embedded email security awareness system for employees, Kumaraguru et al (2007) employed an experimental approach by sending their subjects’ periodic ‘phishing’ emails in order to lure them to visit a supposed ‘malicious’ website. When a victim clicks an email link sent from the system administrator of the organisation, they are being redirected to an intervention that gives them a text and image based information about the malicious links they clicked.

In the following below, a brief analysis of why the security awareness system designed by Kumaraguru et al (2007) did not meet the needed requirements are presented.

**End-user learning preference:**

The authors assumed that all the users would be interested in reading textual information or viewing images on security awareness information. This is one of the reasons the end-users found the system confusing (Kumaraguru et al., 2007). Some employees would have reacted better to audio information and may have impacted their awareness better.

**Time efficiency:**

Secondly, the intervention had no time specified for its completion. This means that users are allowed to spend more than the time needed to comprehend the message, thus boring them quickly.

**Multi-media mechanisms:**

As suggested by security researchers (Shaw et al, 2009), the impact of persuading the interests of targeted end-users, to learn about security using multi-media mechanisms (i.e. combination of videos, images, audio, animations) is far more significant than the use of texts or images alone. Again this factor was missing in the intervention designed for the experimental subjects.

**Non-technical Communication:**

The intervention developed by the authors attempted to inform the experimental subjects about the technical knowledge of malicious URLs which for most people is a daunting process to learn.

**Awareness That Addresses Threats Based on Context of Each Organisation’s Process (Contextual)**

From the design and experiments detailed by the author, the security awareness information presented to users did not address the threats which may occur through the unique business processes of the
organisation. This work argues that when users can relate security awareness to a specific online activity, their comprehension of the risks and mitigation techniques would be better enhanced.

5 **EXISTING SECURITY AWARENESS SYSTEMS AGAINST MALWARE ATTACKS**

This section examines the existing measures developed to enable cyber-security awareness. The publications reviewed and analysed are presented in table 2 below which are evaluated with the guidelines recommended for designing effective cyber-security awareness programs as previously highlighted in section 3. As shown in table 1, the letter Y, represents YES, i.e. the reviewed paper complied with one or more recommended guidelines, while the letter N represents NO, i.e. there is no evidence that one or any of the recommended guidelines were followed. The strengths and weaknesses of these awareness programs are discussed in section 5.1.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Program Developed</th>
<th>Objectives Of Program</th>
<th>NeedsUserEndEfficiency</th>
<th>TimeMechanisms</th>
<th>Media</th>
<th>Contextual Method of Evaluation</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kamaraguru et al.</td>
<td>2007</td>
<td>Embedded training email system</td>
<td>1. To teach people about phishing during their normal use of email</td>
<td>N N N N N</td>
<td>1. Text and graphics 2. Comic strip format</td>
<td>Lab Experiments</td>
<td>1. Embed the training into users’ regular activities.</td>
<td></td>
</tr>
<tr>
<td>Sheng, Steve Magnien , Bryant</td>
<td>2007</td>
<td>Online game</td>
<td>1. Online game that teaches users good habits to help them avoid phishing attacks</td>
<td>N N N N N</td>
<td>Interactive game</td>
<td>Lab Experiments</td>
<td>learning science principles to training materials can stimulate effective learning</td>
<td></td>
</tr>
<tr>
<td>Olusegun, Oj Ithnin, Nf</td>
<td>2013</td>
<td>ISAT</td>
<td>Change the perceptions of people’s thinking and reactions when it comes to information security issues</td>
<td>N N N N N</td>
<td>Mass e-mail, newsletter articles,</td>
<td>None</td>
<td>ISAT program does not address all the needs required by the users</td>
<td></td>
</tr>
<tr>
<td>Cone et al.</td>
<td>2007</td>
<td>Video game</td>
<td>To provide basic information training programs for general computer users</td>
<td>N N Y Y Y</td>
<td>Video game</td>
<td>None</td>
<td>CyberCIEG E engages typical users in an engaging security adventure</td>
<td></td>
</tr>
<tr>
<td>Denning et al.</td>
<td>2013</td>
<td>Card Game</td>
<td>To increase people’s awareness of</td>
<td>N N N Y N</td>
<td>Card Game</td>
<td>None</td>
<td>the graphic design, illustration,</td>
<td></td>
</tr>
</tbody>
</table>

8
<table>
<thead>
<tr>
<th>Year</th>
<th>Participants</th>
<th>Model/Tool</th>
<th>Objective</th>
<th>Y/N Y/N Y/N Y/N Y/N</th>
<th>Learning Method</th>
<th>Additional Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>Kritzinger, E. von Solms, S.H.</td>
<td>E-Awareness Model</td>
<td>To make users understand the risks of using Internet, the importance of securing their personal information and the consequences if this is not done properly</td>
<td>Y N N N N</td>
<td>E-Awareness Model</td>
<td>None</td>
</tr>
<tr>
<td>2011</td>
<td>Labuschagne et al.</td>
<td>Interactive game</td>
<td>To show the effectiveness of using a virtual tool in cyber awareness creation</td>
<td>Y N N Y N</td>
<td>Facebook game App</td>
<td>None</td>
</tr>
<tr>
<td>2012</td>
<td>Arachchi lage, Nag Love, S Scott, Mj</td>
<td>Mobile game</td>
<td>This paper focused on a design that develops the conceptual knowledge that is necessary to combat phishing threats</td>
<td>N Y N N N</td>
<td>Google App Inventor Emulator</td>
<td>Lab Experiments</td>
</tr>
<tr>
<td>2013</td>
<td>Lehrfeld et al</td>
<td>Video creation tool</td>
<td>East Tennessee State need is a security Awareness program to decrease the number and severity of computer virus outbreaks across campus.</td>
<td>Y Y Y Y N</td>
<td>Videos</td>
<td>None</td>
</tr>
<tr>
<td>2013</td>
<td>Potgieter et al</td>
<td>Browser Plugins</td>
<td>To promote security values and provide security suggestions based on a</td>
<td>Y N N N Y</td>
<td></td>
<td>Texts</td>
</tr>
</tbody>
</table>
Table 1. The reviewed security awareness programs.

<table>
<thead>
<tr>
<th>Speciﬁc users</th>
<th>Targeted information security awareness content to the user on possible information security dangers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavioural pattern</td>
<td></td>
</tr>
</tbody>
</table>

5.1 Limitations of existing security awareness programs

The findings from the systematic literature review shows that although the exploits of malware attacks by deception are exponentially growing, there seems to be very little and slow advances in the development of security awareness systems for various groups of end-users. Secondly, the findings reveal the adoption of conceptual and procedural methods for creating security awareness and other mechanisms such as contextual awareness have not been explored. Thirdly, the analyses show that most security awareness systems developed, do not take into consideration the interests or needs of the end-users while designing their programs which makes reduces its efficiency. As suggested by Peltier & Thomas (2005), effective security awareness programs must consider the needs of the intended users in order to communicate the security information in a more effective manner.

Furnell et al, (2006), argue that the method in which security awareness are presented can often obfuscate the process, such that users cannot have any access to the information which are needed to avoid malicious threats. Furthermore, Albrechtsen (2007) conducted a qualitative study of users’ opinion on information security and their main finding was that users consider a user-involving method to be much more effective for influencing user awareness and behaviour.

Unfortunately, most awareness programs overwhelm end-users with extensive monumental training, with very little thought on how security information should be disseminated to distinct categories of users. For example, the security awareness program developed for a business organisation by (Kumaraguru et al, 2007) utilized texts and comic strips to implement an embedded email security awareness system for employees. The researchers, did not consider the categories of employees who utilize the system, neither were the organisational functionalities considered. These non-contextual or non-situational systems have huge limitations especially when the programs designed are intended to be a continuous engaging process of the end-users.

As a result, organisations are spending a huge amount of time teaching people generic technical topics which are not necessary to avoid malicious threats. Security awareness programs needs to be on an incessant life-cycle where users must constantly be updated. But, most awareness programs are simply a monotonous onetime event while others do not consider the online activities which engage the end-user. Hence, the methods of deploying security awareness to a regular user of online social networks cannot be the same process for a regular user of electronic commerce systems. Although the threats may be similar in both situations, the technique of malware deployment is significantly different.

In section 6, a framework for designing security awareness programs for online social network users is proposed and described.
6 A SITUATION-AWARE (CONTEXTUAL) SECURITY AWARENESS SYSTEM FOR SOCIAL NETWORK USERS

In this section, the need for a contextual security awareness program for online social network users to avoid malicious threats is discussed. Moreover, design factors such as time boxing, end-user engagement, integration, activity specific and knowledge testing are suggested for the practical development of an effective security awareness system for OSN users.

6.1 Contextual or Situational learning

Situated learning stresses that effective learning is achieved by specification to the situation in which it is learned. It is grounded on the proposition that; (a) action is founded within the situation of occurrence, (b) knowledge does not transfer between tasks; (c) abstraction learning is ineffective and (d) learning is more effective in a complex social environment. Although, Anderson et al (1996), criticised these claims of situated learning theory based on their review of empirical studies conducted by Lave (1988). Anderson et al (1996) argue that, how learning will be bound to context depends on the type of knowledge that is being acquired. From the review of existing security awareness programs, findings show that most information about cyber-security are provided in abstraction, it does not effectively explicate the threat faced in real time.(i.e.) Moreover, recent findings from Jeng et al (2010), who explored the learning outcome of vocational special education students using situated learning theory with e-learning shows that using teaching strategies for combining situated learning with e-learning yields a better retention for students than other traditional learning methods. It is logical to mention that based on the complexity of issues surrounding malware attacks through online social networks, the need for a contextual awareness program which would visually show users how malicious threats are executed cannot be undervalued.

In the following subsections, factors needed to be considered when developing the proposed system are briefly discussed.

6.2 Factors needed in developing a situation-aware (contextual) security awareness program for online social network users.

6.2.1 Time boxing

Timeboxing simply means the act of setting firm time limits around an activity. Timeboxing a security awareness program results in the tendency for practitioners to focus on the most important issues first (Jalote et al, 2004). In addition, it is a constraint used by system developers to focus on value. For online social networks, delivering multi-media contextual security awareness to end-users needs to be considerably brief as possible. Timeboxing also helps to avoid a common challenge in system design and development known as “feature creep”. “Feature creep” is arises when developers incrementally add features to system without examining its relevance. Feature creep results in wasted effort in both the development and maintenance of systems.

6.2.2 End-user engagement

A very effective technique of making end-users open to listening to the security information is to provide them with awareness training on topics that are in the news (Merrill et al, 2011). It is important for practitioners to investigate existing security vulnerabilities in the news which are relevant to the user community they intend to inform about malicious threats. An effective security awareness program must be appealing to the consciousness of the intended users. Having examined
the ploys utilized by malicious users to lure victims into downloading malware, it has been discovered that most of these lures relatively appeals to the emotions of most victims which explains the reasons for the successful attacks. Moreover, an engaging social network security awareness program should also be enjoyable to engage with. By designing and implementing such a system, security awareness will be embedded into the culture of social network users; hence making the knowledge gained by using such programs easier to understand and transfer.

6.2.3 Integration

According to a report by Marketing Charts, as of 2013, Americans aged 18-64 who use social networks say they spend an average of 3.2 hours per day (marketingcharts.com, 2013). With such high time devotion to social networks on a daily basis, it is important for practitioners to start considering ways to integrate their security awareness systems with online social networking platforms. As earlier described in section 2.2, most social networking platforms (e.g. Facebook), allow the installation of third party applications in order to create a more pleasurable and useful social networking ecosystem for their users. Therefore, such huge opportunity can be utilized to deploy a contextual multi-media security awareness programs.

6.2.4 Activity specific

An effective security awareness program for social network users needs to be presented in a manner that explains the specific activities and corresponding threats within a given instance. For example, an OSN user should get security information on clicking links, playing games or accepting friends request at different instances. This will ensure that users will engage in security training relative to their most engaged social networking activity and optimal learning will be consequently achieved. Again this factor simply addresses the complexity and vagueness of existing security awareness programs which deploys enormous information to the end-user within a very space of time. Activity specific security awareness information essentially means presenting end-users information which is proportional to a risk-related social networking activity - one instance at a time.

6.2.5 Knowledge testing

One of the significance of an effective security awareness program is to allow practitioners assess the practicality and learnability of their programs on the end-user (Kim, 2014). Without this factor, it would be tough to create better upgrades to reinforce the usability of their security awareness programs. Knowledge testing should not only be conducted in a classroom or with the use of survey questionnaires as seen with existing awareness systems. Nevertheless, knowledge testing should be integrated with the security awareness program deployed to the end-user. One of the gaps identified with existing security awareness systems is that their knowledge testing processes is normally a one-time or once in week endeavour. With online social network users, this method of knowledge testing is totally inadequate because the process of gathering the general users in a room or laboratory in order to test their knowledge on the security awareness information is impracticable.
Animated videos, audio illustration

Story telling of previous cases

Notify the friends of users

The use of current news trends based on the profile interests of the user

Security awareness information should be presented within 3-5 mins

Time Boxed

Users should find the system engaging

The system should be installed on Facebook Platform using Facebook APP functionality

Engaging

The system should inform users about the activity related potential attacks of social networking e.g. Using apps and clicking links

Integrated

Activity Specific

The system should present a set of queries to test the retention of users

Knowledge Testing

Putting strict time boundaries around an activity.

Tendency to focus on the most important issues.

The system should motivate users to be engaged.

There should be an element of humour and amusement.

The system must not be a standalone system, it should be integrated with the system were the threats occur

The system should evaluate users retention ability on awareness received

Figure 2. The proposed framework for developing effective security awareness programs for online social network users.

7 CONCLUSION AND FURTHER WORK

Based on the “trust” based relationship of OSNs, it has become relatively easy for malicious attackers to execute deceptive mechanisms on unsuspecting users to spread malicious softwares or malwares. The findings of this work show that very little attention has been given to develop contextual security awareness systems for diverse groups of Web users, e.g. users of online social networks. Consequently, due to lack of effective security awareness, many online social network users are now unknowingly becoming active partakers in the distribution of various malicious softwares growing exponentially. We have seen some commendable efforts to create cyber-security awareness programs by platform providers and practitioners; however, from our systematic literature review, a huge gap was identified regarding designing security awareness programs for distinct user groups. Our proposed contextual or situation-aware security awareness program could eventually be the next innovative breakthrough in developing security awareness systems within OSNs domain. Further research, needs to consider strategies for implementing the factors we highlighted that are needed to develop an effective security awareness program to enable users of OSNs avoid malicious threats.
References


Williams, P. A. H. (2008). In a “trusting” environment, everyone is responsible for information security.pdf.


TRAINING USERS ON INFORMATION SECURITY: EVIDENCE FROM JAVA APPLETS

(Research in progress)

Norilyz Figueroa, University of Massachusetts Boston, USA, email: norilyz.figueroa@gmail.com
Ramakrishna Ayyagari, University of Massachusetts Boston, USA, email: r.ayyagari@umb.edu

Abstract

Information Security issues are one of the top concerns that worry CEOs (Plant, 2014). Accordingly, information systems research has addressed security issues. One of the main areas of research has been the behavioural issues in Information Security, mainly focusing on users’ compliance to information security policies. We contribute to this literature by arguing that users adhere to security policies better if they know how serious and easy the security attacks can be – in other words, ‘seeing is believing’. We argue that proper implementation of security policies requires effective training. Specifically, we argue that training in which written policies are ‘shown’, leads to better adherence to security policies. We use a scenario users face often when browsing – installation of java applets, to test our assertion. Based on previous literature we identified key drivers of compliance and test their effectiveness in our experimental setting – i.e. one group got guidance from written policy, whereas the other group got training from a video. Our contribution is simple yet powerful – in addition to written policies, it is beneficial to actually ‘show’ what the policies accomplish to enhance compliance to security policies.

Keywords: Information security, policies, awareness, training, compliance, behaviour.

1 INTRODUCTION

In this digital age, information has become an important component to any type of organization. From big corporations to small businesses, non-profit organizations and governments, organizations need to safeguard and secure their information. To safeguard the critical information, organizations spend valuable resources on technology tools like intrusion detection systems, firewalls, anti-virus and similar technologies (Lee and Larsen, 2009). However, purely technological solution to security is not going to work (Mitnick, 2003).

Organizations are socio-technical systems and a holistic approach to security needs to involve a socio-technical solution. Individuals are an integral part of organizations and their interactions with technology can be a weak link. Researchers argue that employees are the weakest link in the security chain of an organization (Mitnick, 2003; Warkentin and Willison, 2009). Employees can become an asset to information security, rather than a liability if they choose pro-security behaviours. These behaviours are based on organizational policies and their adherence towards these policies. Accordingly, research in information systems has studied why employees comply or don’t comply to information security policies (Siponen and Willison, 2009; Bulgurcu et al., 2010; Vroom and von Solms, 2004; Guo, 2013). Security compliance issues can be due to intentional (malicious) and unintentional behaviours. Our paper focuses on unintentional behaviours due to lack of awareness or inappropriate assessment of risk.

Although the previous compliance research focuses on the reasons for not complying or how to improve compliance (Crossler et al., 2013; Herath and Rao, 2009; Johnston and Warkentein, 2010), why the issues of non-compliance to security policies exist in the first place? We argue that one of the reasons is due to the gap between the message of the policies and users’ understanding of those policies. Written policies are long and typically full of technical jargon. For an average user, it is difficult to understand what the
policy is conveying. We suggest that organizations can convey policy content in various ways - simulation games, training videos etc. Accordingly, we suggest that ‘seeing is believing’ i.e. rather than telling the users what to do or what not to do (typical wording of security policies), show them why. We test the above assertion by studying user reactions to a security decision involving installation of java applets. Specifically, we test user awareness and compliance to java applet warnings for two groups – group A has to choose behaviour based on applet warning (written policy), group B has to choose behaviour after seeing a video about the meaning of the applet warning.

The following sections are organized as follows. First, we provide a brief literature review of research in information security behaviours. Next, the research model and hypotheses are presented. Third, we discuss our sample and proposed methodology. Finally, we discuss our plan for completing this research-in-progress.

2 LITERATURE REVIEW

Users have to make security decisions as part of their interactions with computer systems. For example, whether to update software, whether to install a plugin or applet, whether to click on links etc. are all decisions that are not directly part of work tasks. For typical users, these actions add to the mental overload and can lead to irrational decisions (West, 2008). It is organization’s responsibility to enable pro-security behaviours without overloading users’ daily activities. Organizations provide guidance for expected behaviour through security policies. However, compliance to such policies is difficult to achieve. How can better compliance be achieved with established policies? This has been a theme of information security studies (Crossler et al., 2013).

Since our goal is to enhance user compliance to security policies, we focused on research that studied user behaviours and compliance to security policies. The goal of this literature review is to identify the key variables that impact compliance behaviours.

What motivates users to practice pro-security behaviours? Anderson and Agarwal (2010) have addressed this question by using modified protection motivation theory. This theory “predicts individual response when faced with a threat” (2010, pg. 615). Based on a multimethod study, the research found that cognitive variables like self-efficacy are an important driver determining pro-security behaviours. Once the user encounters a security decision like a message from an applet, if the user is unsure about the consequences, the users will not be confident in their response to the security scenario. In other words, their self-efficacy in response to the security will be low. Previous research has shown that self-efficacy influences security behaviour (Anderson and Agarwal, 2010; Lee and Larson, 2009).

Researchers have also used variants of theory of planned behaviour to explain behavioral intention to comply to security policies. Studies using this approach suggest that attitude towards the behvoior is a critical variable in explaining user behaviors (Bulgurcu et al., 2010; Anderson and Agarwal, 2007).

One of the ways to counter security threats is to use protective technologies. Dinev and Qing (2007) examined the factors that influence user’s intentions to use protective technologies. Protective technologies are “information technologies that protect data and systems from disturbances such as viruses, unauthorized access, disruptions, spyware, and others” (Dinev & Qing, 2007, Pg. 386). Drawing from theory of planned behaviour, they found that awareness of threats is a strong predictor for making a pro-security decision. Bulgurcu et al. (2010) further delineate awareness into general information security awareness and information security policy awareness. Their research also found support for awareness as a key driver for intent to comply with security policies.

Since our study specifically focuses on java applets, we also reviewed research on browsing behaviour. To practice safe web browsing individuals must be aware of how to configure their security settings and understand web browser alerts. Experiments have also been conducted to understand what actions users
take when they are asked to make security decisions. Web browsers may alert users if they are trying to access a website that is a known phishing website, has a security issue, or is trying to install a Java Applet. Some prior studies have focused their research on the interaction between users and the internet.

Research suggests that users are willing to learn about safer security practices, but might be unsuccessful if not provided guidance (Finn and Lumsden, 2005). Previous studies have shown that users are not adept at setting even security options on browsers (Furnell et al., 2006). It suggests that unless proper training is provided, users might not be able to make pro-security decisions when presented with security scenarios like java applet messages.

Users are often unaware of the impact of their security decision (Zurko et al., 2002). In a study of lotus client users, Zurko et al. (2002) found that when presented with a security decision during users’ work, normally security conscious users allowed potentially insecure applications to run. This alludes to the mental workload of security decisions on top of the work demands i.e. users see these security decisions as additional workload (West, 2008).

One of the ways to assess the effectiveness of a security intervention is to see if users’ understand the impact of their security decision. In literature this is reflected in the outcome construct ‘vulnerability of resources’ (Bulgurcu et al., 2010). This construct captures the users’ belief that organisational resources are at risk if they don’t follow security recommendations. Based on the above review, the key variables studied in this paper are awareness, self-efficacy, attitude and vulnerability of resources.

3 OUR STUDY

One of the activities that is far reaching is browsing the web. Typical policies that govern users’ behaviours regarding browsing can be found in ‘Internet use policy’, ‘acceptable user policy’ etc. While browsing the Internet many users will encounter mobile codes (like applets, active x controls, plugins) that enhance user experience, but at the same time pose security risk.

Mobile codes are executable software that is transferred between systems. Common mobile codes are Java Applets, ActiveX controls and Plugins. This study will focus specifically on how users behave towards Java Applets. Java Applets will run on a variety of platforms and browsers, unlike ActiveX controls that will only run on Microsoft applications and platforms (Microsoft, 2015). A Java Applet is a program written in Java Programming Language which is transferred to a system and then executed by a web browser (Oracle, 2015). The mobile code dialog boxes typically require users to make a security decision, and users might override security protections (for example, running an untrusted applet). This issue is compounded because the messages provided by the dialog boxes are conflicting, for example see the active x control warning message in Figure 1 – the message tells the users that Active content can be useful but at the same time it might also harm their computer. Once a user downloads a malicious ActiveX control it can gain full access of the computer and will endanger user privacy (Schneier, 2004).

![Figure 1. Sample Active X control dialogue message](image_url)
While browsing the Internet, users can encounter two different types of Java Applet warning messages. Users can encounter an Applet with a verified digital signature. A user can also come across an Applet with a digital signature that cannot be verified. If the signature is verified it is coming from a trusted source, and if the Applet is executed it will have greater access over users computing resources (Oracle, 2015). It should be noted that trusted source doesn’t imply safe source, if your computer crashes after installing an applet from a trusted source, then at least you know who caused the crash (because the source of the applet is known). On the other hand, if the signature cannot be verified then the Applet is originating from an untrusted source. Users can be easily confused when deciding to run an Applet if they are not clearly informed on the distinction between these two types of Applets. Further, users are so used to seeing mobile codes (like applets, active x controls); users might not think twice and may run the codes.

If a user mistakenly allows a malicious Applet to run on their computer, the Applet can gain full control over the users computing resources. A malicious Applet has the ability to capture images of users computing environment. It can also capture keystrokes which can compromise users’ sensitive information (i.e. passwords). They are also capable of executing new programs onto a user computer.

This study will allow a better understanding of how users behave towards these applet warnings. Our assertion is that if the users are shown the severity of these threats, it leads to users who are better prepared to handle security decisions.

When presented with a security decision like applet warning, users might be unaware of the meaning of the message and also understand the options it presents (Anderson and Agarwal, 2010). This concept is captured through self-efficacy which reflects users’ confidence in dealing with security scenarios (in this instance, java applets). If the users are actually shown the impact of the applet options, they will be better prepared. Better understanding of the applet options also increases users’ awareness about the security message and also the consequences of their actions. In general, technology or information security awareness captures the raised consciousness or understanding (Dinev and Qing, 2007; Bulgurcu et al., 2010). Armed with the knowledge and understanding of options provided in the applet messages, users will be more responsible and understand the risk posed by their actions to organizational resources (Bulgurcu et al., 2010; Zurko et al., 2002). Understanding the severity of their actions (i.e. clicking on one button can lead to complete ownership of a machine by attacker) leads to changed attitudes towards java applets. Therefore, we propose that

H1: Users’ perception of self-efficacy will be higher for the group that watches the video compared to group that relies on text warning only.

H2: Users’ perception of awareness will be higher for the group that watches the video compared to group that relies on text warning only.

H3: Users’ perception of vulnerability of resources will be higher for the group that watches the video compared to group that relies on text warning only.

H4: Users’ perception of attitude will be higher for the group that watches the video compared to group that relies on text warning only.

4 METHODOLOGY AND RESULTS (IN PROGRESS)

Surveys were used to collect data for this research and test the hypotheses. The survey population consisted of 141 undergraduate students from a large public university in northeast United States. The participants belonged to College of Management and were enrolled in either introductory business or information technology courses. There was no incentive for students to complete the survey and participation was strictly voluntary. No personally identifiable information about the respondents was collected and respondents were assured of anonymity of their responses.
Surveys were administered by paper and contained questions measured on a 5 point Likert scale. The survey contained a captured image of a typical Java Applet warning “The application’s digital signature…Do you want to run the application?” The scales used in the present study have been adapted from previous research (Bulgurcu et al., 2010). For example, information security awareness scale was adapted to reflect java applet awareness. The variables used in this study are self-efficacy (Anderson and Agarwarl, 2010), Awareness, Attitude, Vulnerability of resources (Bulgurcu et al., 2010).

The survey was used to determine users’ awareness of Java Applet Security decisions and how they responded towards them. In order to capture awareness the survey was administered to two separate groups. Group A consisted of 65 students and Group B contained 76 students. Both of the groups responded to same questionnaire, however; Group B took the survey after watching a video.

The three minute video demonstrated the risks associated with downloading Java Applets. The video started by demonstrating a user being prompted to install a Java Applet during a web browsing session. Then, the user installs the applet and continues the browsing session. However, unknown to the user, the act of installing the applet provides a backdoor access to the attacker. The video then shows how easy it is for the attacker to capture screenshots of users’ desktop, execute programs, capture keystrokes etc. This video was intended to stimulate awareness and educate Group B on Java Applet security risk. After Group B watched the video, they were then asked to complete the survey.

4.1 Current status and future work
We have finished our data collection and initial coding of the surveys. Our goal is to finish the data analysis by conference time. Since we are measuring differences among the two groups, we will be using SPSS to conduct our analysis. If our results support our assertion, we contribute a simple yet powerful message to behavioural information security literature. Showing users’ the reason behind security policies rather than just telling them what to do (typical policy language) is more effective. This will have implications on how users are trained on security policies of an organization.

References


