ABSTRACT

Organizational purpose or function changes usually concern with information systems that have to be modernized. However, there are difficulties of making information system modernization decisions because of suppliers and clients may have different notions about modernization. In this paper, we present a review of statements for information system modernization. The objective of this review is to describe current reported knowledge in terms of what kind of modernization is defined and what the domain of modernization is. We found 42 statements for modernization from 36 papers. These findings show that the concept of modernization is unestablished. However, if we want to have similar notions about information system modernizations then we have to be transparent where the changes have effect. The analysis of relationships between the entities of enterprise architecture is presented to help in making information system modernization decisions. The case-based lessons concern the semantic assets of Finnish social welfare. The suppliers of three client information systems (CIS) were evaluated and they answered that 2-58 % of the semantic assets are unknown. Furthermore, there are described the main data groups and logical data stores which are not allowed in the TOGAF content metamodel.

KEYWORDS

software modernization, system modernization, architecture-driven modernization, literature review, enterprise architecture, semantic asset

1 INTRODUCTION

When something is becoming modern or something is made modern, it might mean that something is more suitable for present styles or needs, or the newer information, methods, or technology is used [1]. Nowadays, we are dependent on various information systems. The information systems can be seen as a collection of people, processes, data, models, technology and partly formalized language, which serves an organizational purpose or function [2]. If the organizational purpose or function changes, then we have to make something modern and quite often we are concerned with information system modernization.

Many information systems can be called legacy systems. Legacy systems are typically described as mission critical systems, systems that are in an important role in an organization, and systems which are vital but unmanageable [3, 4]. Bennett [4] has further described that legacy systems may contain years of accumulated experience and knowledge - the system may be the only place where an organization’s business rules exist. Wallace et al. [5] cited that system “modernization is a specialized application of system reengineering, which is the disciplined evolution of a system from its current state to a new one”. Weiderman et al. [6] present five activities which can be used with legacy systems: assessment, maintenance, transformation (can be seen as reengineering), replacement and combined strategies. Comella-Dorda et al. [7] divided the system evolution activities into the following categories: maintenance, modernization and
replacement. Furthermore, they define that modernization involves more extensive changes than maintenance and the system must be replaced when it can no longer be evolved.

In this paper, we are interested in information system modernization. Our earlier studies have discussed decision-making in modernizations [8-10], but during our research we have been doubtful about the concept of modernization. Clients and suppliers may have different notions about modernization [11], which causes difficulties of making modernization decisions.

Although modernization has been studied from different perspectives, the concept of modernization is vague. Our motivation is to find what modernization is. A definition can be described as “a statement expressing the essential nature of something” [12]. Suonuuti [13] states: “Definitions are statements describing a concept”. In this study, we review and analyze the statements of information system modernization. We collect the statements by using a literature review. In our review, a statement is something which tells us what modernization is. Our search returned altogether 1056 hits and 36 papers included some kind of a statement or statements of modernization.

This paper is organized as follows: Section 2 introduces literature-based modernization; Section 3 presents the architecture-based modernization efforts in Finnish social welfare. Section 4 discusses our results.

2 LITERATURE-BASED MODERNIZATIONS

We are interested in the statements of information system modernization. Therefore, we need to find out how the concept of modernization is defined in scientific papers. The data collection method in this study is literature review. Qualitative methods (content analysis and creating themes) are used for analyzing the collected material.

Research literature can be reviewed for different purposes: to provide a theoretical background for research, to learn the breadth of the research field or to answer practical questions by finding out what is said in existing research literature [14]. In our case, the purpose for the literature review is to learn what is said in the existing literature. Our review has been conducted by adapting two systematic review guidelines [14, 15]. Okoli and Schabram [14] describe that a literature review is systematic, when it 1) follows methodological approach systematically, 2) explains explicitly how review has been conducted, 3) includes all relevant material comprehensively, and 4) can be reproduced by other researchers. Kitchenham and Charters [15] define systematic literature review as “a form of secondary study that uses a well-defined methodology to identify, analyse and interpret all available evidence related to a specific research question in a way that is unbiased and (to a degree) repeatable”. In our review, we have decided not to assess the quality of the studies and thus, our review is not systematic in the means of Kitchenham and Charters [15]. Moreover, in our case, we look for statements of modernization and we use all the scientific articles, not only primary studies, from the computing field which can be retrieved from the electronic databases we have chosen. However, we have defined our review process as follows:

1. Specifying the search terms
2. Selecting the databases
3. Creating the selection criteria
4. Searching the papers
5. Appraising the hits and selecting the papers
6. Citing the definitions from the papers

Our research strategy is a mix of exploratory and descriptive strategies. Exploratory strategy tries to find out what is happening, seeks new insights and generates new ideas and
hypotheses for future research and descriptive strategy describes the current status of a phenomenon or a situation [16]. A systematic literature review aims at identifying, evaluating and interpreting all available research material to answer a research question [15]. In our study, we collected all scientific papers from the field of computing using the databases we have access to. We did not use books or grey material or manual searches. However, we used all the databases, which were available to us.

After we piloted our search, we decided that the best way to search for the answer to our question is to use the search term “software modernization” OR “system modernization”. We came up this search string, because we were interested in what is modernized: the software, the information system, the software system, the legacy system or the legacy software. Our search string finds all of these terms.

We used the following databases to conduct our search:

- ACM Digital Library
- IEEE Xplore
- SpringerLink
- Elsevier Science Direct
- Wiley Online Library
- Emerald
- EBSCOhost

We did not use manual searches. We included the papers which are scientific conference or journal papers, come from the field of computing, are available in full versions, are written in English, and include a statement of (software/system) modernization.

We excluded papers which are written by the authors of this paper or are found an earlier search from other databases. Altogether, we received 1056 hits from the selected databases. Finally, after going through all the databases, we had 36 articles where modernization was defined (Table 1).

<table>
<thead>
<tr>
<th>Database</th>
<th>Number of hits</th>
<th>Selected articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACM Digital Library</td>
<td>108</td>
<td>23</td>
</tr>
<tr>
<td>IEEE Xplore</td>
<td>296</td>
<td>9</td>
</tr>
<tr>
<td>SpringerLink</td>
<td>228</td>
<td>-</td>
</tr>
<tr>
<td>Elsevier Science Direct</td>
<td>220</td>
<td>2</td>
</tr>
<tr>
<td>Wiley Online Library</td>
<td>122</td>
<td>2</td>
</tr>
<tr>
<td>Emerald</td>
<td>18</td>
<td>-</td>
</tr>
<tr>
<td>EBSCOhost</td>
<td>64</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>1056</td>
<td>36</td>
</tr>
</tbody>
</table>

With selected articles we started the content analysis that is the quantification of qualitative text material [17]. In content analysis the codes (i.e., sets of words and phrases) are occurrences of those codes are searched [18]. We found 42 statements that we compared with the modernization domains (Table 2). In some papers, there was more than one statement.

In software modernization [19-30] the domain can be some kind of system and the software artifacts or assets. In ADM (architecture-driven modernization) [28, 31-36], the possible domains are the system, application, software or software asset. In eight cases modernization is named divergently: LIS modernization [45], modernization of enterprise systems [44, 46], modernizing a legacy system [47, 48], model-driven modernization [49, 50] and software system modernization [51].

1 http://dl.acm.org/
2 http://ieeexplore.ieee.org/Xplore/home.jsp
3 http://link.springer.com/
4 http://www.sciencedirect.com/
5 http://www.ebscohost.com/academic/academic-search-elite
6 http://onlinelibrary.wiley.com/
7 http://www.emeraldinsight.com/
Table 2. Comparison of modernization domains and defined modernizations

<table>
<thead>
<tr>
<th>Domain</th>
<th>ADM</th>
<th>LIS modernization</th>
<th>model-driven modernization</th>
<th>modernization of enterprise</th>
<th>modernizing a legacy system</th>
<th>software modernization</th>
<th>software system modernization</th>
</tr>
</thead>
<tbody>
<tr>
<td>aged legacy system</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aging software system</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>enterprise system</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>existing applications</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>existing IT assets</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>existing software</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>existing software artifacts</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>existing software assets</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>existing system</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>legacy information system</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>legacy software</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>legacy system</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIS</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>software system</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>system</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>2</td>
<td>13</td>
<td>1</td>
</tr>
</tbody>
</table>

3 ARCHITECTURE-BASED MODERNIZATION

Governments are adopting Enterprise Architecture (EA) frameworks for improving the interoperability of the information systems that are used in the production of services. In Finland, the EA development framework is called JHS 179 the entities of which are implicitly mapped the entities of the TOGAF content metamodel [52].

There seem to be several challenges on the adoptions of the EA frameworks [53, 54]. The case-based lessons concern the semantic assets of Finnish social welfare. “The data model is an important part of Finnish social welfare’s information architecture” and “it contains over 200 different social care client records, over 150 core components and over 100 classifications” [55]. Actually, the spreadsheet-based core components and social care client records are nowadays the main semantics assets of the social welfare in Finland.

One-fifth of the public social welfare organizations in Finland use more than one client information system (CIS), i.e., they have different CIS for different services (e.g. child welfare or disability services) [56]. It is concluded that “the client information systems do not yet support casework processes”, as well as, “[the substantial and functional development of CIS is the most important challenge in the near future” [56]. The casework processes can be reached by implementing the core components and client records into different CIS. Therefore, three CIS, the suppliers of which are the leaders of the market9 and one other representative10, were evaluated [57]. The representatives of the CIS suppliers were interviewed and their answers were analyzed. The results are presented numerically, whereas, the preliminary results were present perceptually (Table 3).

Table 3. Analyzed answers

<table>
<thead>
<tr>
<th>Semantic assets</th>
<th>Number</th>
<th>Known</th>
<th>Unsure</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client records</td>
<td>217</td>
<td>120-195</td>
<td>2-11</td>
<td>2-63</td>
</tr>
<tr>
<td>Core components</td>
<td>148</td>
<td>112-148</td>
<td>0-7</td>
<td>35-86</td>
</tr>
<tr>
<td>Core components in CIS</td>
<td>148</td>
<td>60-103</td>
<td>1-13</td>
<td>35-85</td>
</tr>
</tbody>
</table>

9 http://www.jhs-suositukset.fi/web/guest/jhs/recommendations/179
10 http://www.cgi.fi/
11 http://www.tieto.fi/
11 http://www.abilita.fi/index.html
The client records have the name and description and they are mainly known by the suppliers. However, the core components are better to known than implemented. Furthermore, all suppliers recognized the same 25 core components, the implementations of which are not in CIS.

If casework processes are going to be changed, it is reasonable to be familiar with data entities such as the core components and client records. Not only the name and description of the client record because the client records contain the specified core components and specific field. Furthermore, the eWelfare survey [56] did not take into consideration the challenges of the core components and client records implementations. There are only the results of the implementation of different classifications in the client information systems used by public social service providers [56].

The integrity of the data model is crucial both in the definition phase and in the implementation phase. Adaptations and explorations of the data model were improved by the Resource Description Framework (RDF) for storing and utilizing the Core Component Technical Specification (CCTS) model [58].

During the years 2013–2015 the data model is under the inspection, in where the contents of documents such as the necessity, adequacy, notation and recurrence of the information are estimated. However, the RDF-based semantic assets are not used in the inspection phase because the versatile know-how for the interpretation of the RDF-based assets, not to mention, for the understanding and adapting of the RDF-based assets is needed [59].

There are other semantic assets of the social welfare including the information architecture as follows:
- main data groups (päättietoryhmien kuvaus in Finnish)
- dataflow diagram (tietovirtakuvaus in Finnish)
- logical data stores (loogiset tietovarannot in Finnish)
- vocabularies (sanastot ja koodistot in Finnish)

The conceptual model does not contain any terms that can be mapped into data entities (i.e., social care client records, core components and classifications). The main data groups and logical data stores are not related to the data entities. However, the names of 37 social care client records are related to processes in the dataflow diagram. The vocabularies contain the descriptions of the social care client records, core components and classifications.

4 DISCUSSIONS

We noticed that the defined modernization (system, software, architecture-driven) does not reveal what the domain of modernization is. System and software modernization can be defined similarly as well as software and architecture-driven modernization. Our findings reveal the vagueness of the term modernization [60].

If we want to have similar notions about information system modernizations, then the changes have to be transparent. Enterprise architecture frameworks make the entities and relationships of the enterprise transparent, and the information system modernization decisions can be based on the analysis of relationships between the entities. We have identified that at least four entities (i.e., services, data entities, logical application components, logical

---

technology components) should be taken into account when making the decision. Services provide governed interfaces to access organizational functions, as well as, they provide and consume data entities. Therefore, if organizational purposes or functions are going to be changed, it is reasonable to list all data entities which are provided and consumed by services to be changed. Furthermore, we have to be familiar with the logical components of the information system because the services are realized through logical application components and implemented in logical technology components.

There are several challenges on the adoptions of the EA frameworks at the government level. The main target is semantic interoperability. However, are we closer with semantic interoperability with numerous architectural descriptions? The main aim of the paper was questioning the meaning of several architectural descriptions. Do they help us to be closer with semantic interoperability? Furthermore, we have to questioning the frameworks of the enterprise architecture. For example, the attributes of the data entity in the TOGAF content metamodel do not support the semantics interoperability because they do not specify the precise meaning or precise semantics of data elements that are atomic units of data.

5 REFERENCES


[52] The Open Group, "TOGAF® Version 9.1, an Open Group Standard".


