Collecting and Integrating Unstructured Information into Enterprise Architecture Management: A Systematic Literature Review

Robert Ehrensperger, Clemens Sauerwein and Ruth Breu
Institute of Computer Science, University of Innsbruck, Technikerstr. 21A, 6020 Innsbruck, Austria

Keywords: Unstructured Information, Information Collection, Enterprise Architecture Management.

Abstract: In the age of digital transformation, strategic IT alignment is becoming a primary driver for economic success. In this context, the optimization of strategic IT alignment plays a key role in enterprise architecture management (EAM). A successful EAM strategy depends on the quantity and quality of the available information within the enterprise architecture (EA) models. EA information about the functional scope of software solutions and its supported business processes is often available only in an unstructured form. Automatic acquisition of this information assists companies in the design of target architectures. In recent years, new technologies have been introduced that facilitate the use of unstructured information. The research at hand discusses these new technologies and emerging challenges. Furthermore, it provides a systematic literature review of the current state of research on collecting and integrating unstructured information into EAM.

1 INTRODUCTION

Through the ongoing digitalization, the importance of strategic IT alignment is increasing because it supports the ability of companies to transform themselves according to their business strategy (Valenduc and Vendramin, 2017; Luftman and Brier, 1999). The principal objective of enterprise architecture management (EAM) is to optimize this strategic IT alignment (Farwick et al., 2016). Therefore, EAM provides frameworks that allow alignment of information systems and underlying IT infrastructure with the business capabilities of an enterprise (Babar and Yu, 2015).

The success of EAM crucially depends on the quantity and quality of the available information within the enterprise architecture (EA) models (Fischer et al., 2007). These models represent EA dimensions such as business, application, technology, and information as well as the complex relationships between them (Buschle et al., 2011; Lankhorst, 2009). Moreover, these models allow us to gain a comprehensive understanding of an EA and are essential to communicate the required changes to stakeholders (Farwick et al., 2011). Enterprise architecture management is designed to manage necessary changes within companies' IT landscapes. It also allows us to improve the alignment of business and IT and to increase the availability of IT systems (Langenberg and Wegmann, 2004; Ross et al.,2006; Ross, 2003).

The collection of additional unstructured information from enterprise-external sources might introduce new opportunities in EAM (Becker et al., 2009; Becker et al., 2011). Enterprise-external information sources include all sources that originate outside of an internal enterprise environment and reflect changes coming from the real world that are relevant for enterprise architects. For example, enhancing EAM with vendor product information, such as license, product-life-cycle information, and the functional scope of software solutions, might allow improving the EA planning activities of enterprise architects (Farwick et al., 2013). Also, collecting announcements that appear in enterprise-external information sources such as online blogs about security vulnerabilities of IT systems might allow for faster implementation of required EA changes (Martin, 2008).

Technological improvements such as big data (Bakshi, 2012) and artificial intelligence (O'Leary, 2013) have unlocked new potential for examining unstructured information and drawing conclusions from this information. However, research and practice have not yet produced a comprehensive

Ehrensperger, R., Sauerwein, C. and Breu, R.
Collecting and Integrating Unstructured Information into Enterprise Architecture Management: A Systematic Literature Review.
DOI: 10.5220/0009339107280737
In *Proceedings of the 22nd International Conference on Enterprise Information Systems (ICEIS 2020) - Volume 2*, pages 728-737
ISBN: 978-989-758-423-7

understanding of how to collect unstructured information for EAM based on these technologies. Moreover, it is unclear to what extent these technologies are used within EAM. This unclearness motivated us to design the systematic literature review at hand. The main research goal of this paper is to provide a comprehensive overview of existing research that deals with collecting and integrating unstructured information into EAM and outlines potential research challenges. Therefore, we conducted a systematic literature review based on the snowballing methodology (Wohlin, 2014). First, we defined a start set of five papers, on which several iterations of forward and backward snowballing were performed. In doing so, we examined 384 papers. This procedure resulted in a final set of 20 papers, which we comprehensively analyzed.

The remainder of this paper is structured as follows. Section 2 provides a summary of related work. Section 3 outlines the applied research methodology. Section 4 describes the results of the snowballing procedure and the classification of identified papers. Section 5 discusses the key findings of the literature review concerning our research goal and outlines potential research limitations. Finally, Section 6 concludes our contribution and provides an outlook on future work.

2 RELATED WORK

A few researchers have addressed the topic of automated information collection for EAM. Farwick et al. (Farwick et al., 2013) stated that information sources such as network scanners and monitors, configuration management databases, project portfolio management tools, enterprise service buses, change management tools and license management tools can deliver valuable information for EAM. The automated collection of information with vulnerability scanners can also provide useful information (Buschle et al., 2011). However, according to multiple authors (Grunow et al., 2013; Buschle et al., 2011; Farwick et al., 2011), insufficient research has been conducted on the analysis of potentially relevant information sources.

In summary, there exists related work on the collection of structured information for EAM. However, to the best of our knowledge, only a few scientific investigations on the identification and collection of relevant unstructured information for EAM have been conducted. Additionally, the existing research did not focus on relevant enterprise-external information sources for EAM.

3 APPLIED RESEARCH METHODOLOGY

This systematic literature review was finished in December 2019. It is built on the snowballing methodology (Wohlin, 2014), which is based on the ideas of Webster and Watson (Webster and Watson, 2002). A comparison of the snowballing methodology with the conventional database search did not show any significant differences regarding the results (Jalali and Wohlin, 2012). Moreover, applying the snowballing methodology offers advantages in obtaining information (Hendriks et al., 1992), such as the ability to uncover hidden aspects (Atkinson and Flint, 2001).

Based on the guidelines by Wohlin (Wohlin, 2014), our systematic literature study can be divided into three steps. The first is the definition of start set. The second is execution of snowballing iterations, including both backward snowballing (i.e., looking at the references of a paper under investigation), and forward snowballing (i.e., identifying new papers citing the paper being investigated). The last step of our research methodology was analysis and classification of the final set of papers.

3.1 Definition of Start Set

Investigations showed that Google Scholar finds significantly more citations than other search engines such as Web of Science and Scopus (Martín-Martín et al., 2018). Accordingly, Google Scholar provides a reliable data source to extract citation information. Since the results of the snowballing methodology crucially depend on citation information, we used Google Scholar as search engine.

The first step of the snowballing methodology consists of the definition of a start set. According to Wohlin (Wohlin, 2014), there is no common methodology for defining a start set. Therefore, we searched for relevant publications focusing on "enterprise architecture management" and "unstructured information." This initial search was conducted systematically with the tool "Publish or Perish" (Harzing, 2019), which allows users to create complex and reproducible search queries by accessing the Google Scholar API (application programming interface). For our search, we specified that the keyword "enterprise architecture management" must appear in the title, and "unstructured information" somewhere in the content of the publication.

This procedure delivered a set of 11 papers. In order to obtain only relevant publications, we first analyzed the title, abstract, and keywords. If this preliminary analysis identified the paper as potentially relevant for our research goal, we conducted a partial reading of the papers. During reading of papers, we applied inclusion and exclusion criteria to select papers for the start set. We included papers that were accessible in full text, written in English or German, and focused on the collection of unstructured information for EAM. Any publication that did not match any of the inclusion criteria was excluded. We also excluded papers published before 2014, duplicates and publications dealing with big data analytics not relating to EAM. The exclusion criteria overrode the inclusion criteria; in other words, if a publication met an exclusion criterion, it was excluded even if it met one or more inclusion criteria. After applying the inclusion and exclusion criteria, we considered all remaining publications.

Ultimately, the first step of the applied research methodology yielded a start set of five papers. These papers are (Hacks and Saber, 2016; Rosina, 2015; Roth, 2014; Schmidt et al., 2014b) and (Vanauer et al., 2015).

3.2 Execution of Snowballing Iterations

In the second step of our systematic literature review, we executed several iterations of snowballing, each consisting of both forward and backward snowballing, until no new papers were found. In total four iterations of snowballing were executed, and 384 papers were examined. After each iteration, we again analyzed the papers' title, abstract, and keywords, conducted a partial reading of papers deemed potentially relevant, and applied the inclusion and exclusion criteria (cf. section 3.1). After four iterations no further relevant papers could be identified, which terminated the snowballing iterations. This procedure resulted in a final set of 20 papers.

3.3 Analysis and Classification of the Final Set of Papers

We analyzed the remaining 20 publications according to the following classification categories: (a) bibliographic information, (b) applied research methodology, (c) addressed research topics, and (d) identified challenges. Each of the analyzed papers was read at least by two authors of the publication at hand.

The bibliographic information category includes general information regarding the publication, such as title, author, date, venue, and type of publication. The research methodology category includes the description and categorization of the applied research methodologies (e.g., case study, survey, subjective/ argumentative study, descriptive/interpretative study, experiment). This categorization system is derived from (Galliers and Land, 1987). The addressed research topic category classifies publications according to the primary research topics they address. We followed an iterative process, meaning that we created a research topic category whenever at least two publications were focusing on the same research topic. Finally, we listed and sorted the research challenges identified by the authors in order to provide a comprehensive overview of addressed and open research challenges. In doing so, all the authors of the paper at hand discussed the mentioned challenges and drew different categories out of it.

4 RESULTS

The applied snowballing methodology resulted in a final set of 20 papers (Diefenthaler, 2016; Farwick et al., 2016; Fittkau et al., 2015; Hacks et al., 2016; Holm et al., 2014; Johnson et al.,2016; Kirschner and Roth, 2014; Möhring and Schmidt, 2015; Ortmann et al., 2014; Rosina, 2015; Roth, 2014; Roth and Matthes, 2014; Schmidt et al., 2014a; Schmidt et al., 2014b; Välja et al., 2016; Välja et al., 2015; Vanauer et al., 2015; Zimmermann et al., 2016; Zimmermann et al., 2017; Zimmermann et al., 2015). In this section, the content of these papers and the results of the classification are discussed.

4.1 Bibliographic Information

In the first step, the bibliographic information was investigated. The number of publications remained roughly constant between 2014 and 2016, averaging more than six publications per year. A total of 19 publications between 2014 and 2016 were identified, including three journal papers. Only one publication fulfilling the criteria was found for 2017, and none for 2018 or 2019.

4.2 Research Methodology

In some publications, multiple research methodologies are applied. The most theoretical work in this field has used interpretive and argumentative methods. Fourteen publications used a descriptive research method, and nine of these focused on describing frameworks for EAM. The second most commonly used methodology was the subjective approach, used in six publications. Three authors used the research methodology experiment whereby they tested approaches to automate EA modeling and its visualization. Furthermore, two case studies were identified evaluating the benefits of methods and tools in an industrial setting. Moreover, surveys were not used at all as a research methodology. However, surveys provide a better description of the general population (Rea and Parker, 2014) compared to the other described research methodologies.

4.3 Research Topics

The identified papers address several research topics regarding the collection of unstructured information for EAM. Our investigations yielded a comprehensive overview of the current state of research in this field. Ultimately, it resulted in six different categories, which are shown in Figure 1. In this section, we will discuss each of these categories in detail.

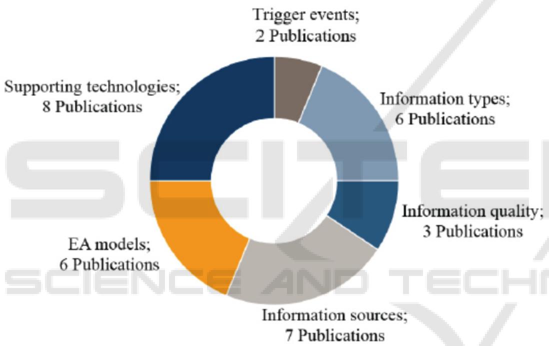

Figure 1: Main research topics found in the review.

4.3.1 Trigger Events

The first category includes trigger events, which are required to start the information collection process. For frequently changing information objects, these triggers play a crucial role in the timeliness of the collected information. Farwick et al. (Farwick et al., 2016) and Roth (Roth, 2014) identified several relevant trigger events. These events are classified into three groups: (i) manual trigger events, (ii) trigger events from enterprise-internal tools, and (iii) trigger events from the EA repository. However, existing research does not take into account enterprise-external trigger events.

4.3.2 Information Types

The second category, which contained six papers, addresses the different types of unstructured information for EAM. The publications of Vanauer et al., Rosina, Hacks, and Saber (Vanauer et al., 2015; Hacks et al., 2016; Rosina, 2015) describe different relevant information types such as documents, spreadsheets, and presentations as well as general applications and processes. Schmidt et al. (Schmidt et al., 2014a) focused on these information types from two different perspectives, the (i) static and (ii) dynamic perspectives. These are shown in Figure 2. The static perspective operates mainly on structured information and supports strategic decision making. The dynamic perspective focuses on highly volatile semi- and unstructured information such as log files. Leveraging both perspectives would allow us to run a descriptive, prescriptive, and predictive analysis for EAM. These analysis opportunities support EAM not only in long-term strategic decisions but also in imminent tactical or even operational decisions.

Figure 2: A static and dynamic perspective on EAM information (Schmidt et al., 2014a).

Moreover, the research discusses linked information. For example, the authors (Zimmermann et al., 2017; Ortmann et al., 2014) highlight this type of information. Linking different information such as servers, applications, interfaces, and their supported business processes enables enterprises to reach value-added conclusions, such as the identification of redundant IT systems in the EA.

4.3.3 Information Quality

The third category addresses the quality of the collected information. Hacks and Saber (Hacks and Saber, 2016) argued that better quality can be achieved through preprocessing and preparation of the information. Ortmann et al. (Ortmann et al., 2014) raised the possibility of increasing information quality by conducting analysis directly on the original enterprise-external information sources in their diverse formats and terminologies. The advantage of this approach is that the information can be made available in a timely fashion and thereby is of higher quality. An automated information collection process is less error-prone than a manual process. Thus, the automation of the information collection process leads to better information quality, according to Holm et al. (Holm et al., 2014).

4.3.4 Information Sources

The fourth category concerns relevant information sources for EAM. In total, seven publications were identified that discuss these information sources, which can be divided into enterprise-internal and enterprise-external sources. First, many researchers such as Farwick et al. (Farwick et al., 2016), Johnson et al. (Johnson et al.,2016), Välja et al. (Välja et al., 2016), Diefenthaler (Diefenthaler, 2016), Schmidt et al. (Schmidt et al., 2014a), and Fittkau et al. (Fittkau et al., 2015) identified and described different enterprise-internal information sources. These sources can be categorized according to the components of IT systems: hardware (e.g., systems, infrastructure), software (e.g., runtime information, vulnerability), databases (e.g., wiki's, change-, license-, and portfolio-management tools), networks (e.g., network monitors and scanners), and procedures (e.g., process state information, enterprise service bus). This structure is derived from Rainer and Cegielski (Rainer and Cegielski, 2013).

Six of the seven publications focused on enterprise-internal information sources. Only one publication discussed enterprise-external information sources; Zimmermann et al. (Zimmermann et al., 2017) underlined that in addition to enterprise-internal sources, enterprise-external sources might also improve EAM. However, the authors did not mention any specific information sources.

4.3.5 EA Models

The fifth category discusses EA models and the use of unstructured information sources to create or maintain existing models. These models allow us to visualize EA information, which supports us in understanding and working with EAs (Fischer et al., 2007). In our analysis, we distinguish between the following two categories: (i) creation of EA models, and (ii) maintenance of EA models. In total, we identified six publications that addressed EA models.

First, in order to automate the creation of EA models, Johnson et al. (Johnson et al.,2016) proposed the use of machine learning techniques. They considered the use of dynamic Bayesian networks because such networks can capture fuzzy information that describes EAs. Välja et al. (Välja et al., 2016) automatized the creation of EA models from multiple heterogeneous information sources. To do so, they used the Joint Directors of Laboratories (JDL) framework (Välja et al., 2016; Liggins, 2008; Steinberg and Bowman, 2017), which facilitates the fusion of information on different granular levels.

Second, for the maintenance of EA models, Kirschner and Roth (Kirschner and Roth, 2014) merged different EA models for a single EA repository. In order to reach this goal, they described merge algorithms, which detect EA model conflicts and generate resolution tasks. Roth and Matthes (Roth and Matthes, 2014) presented a concept that allows the differences between EA models to be analyzed and visualized. Fittkau et al. (Fittkau et al., 2015) introduced an approach that utilizes information system monitoring to improve consistency between EA models and real information systems. Furthermore, (Zimmermann et al., 2016) detailed a process that makes it possible for information sources to be integrated continuously with an EA model.

4.3.6 Supporting Technologies

Finally, a remarkable number of publications discuss technologies that support the collection of unstructured information for EAM. This topic will attract more attention in the future because many new technologies are on the rise (Bakshi, 2012; O'Leary, 2013). Generally speaking, to make it more comprehensible, these technologies can be categorized according to their nature into (i) big data technologies, (ii) semantic technologies, and (iii) service-oriented technologies. Eight publications were included in the following overview.

First, big data technologies allow us to collect large quantities of EAM-relevant information with varying structure and to conduct near-real-time analysis on, for example, architectural information contained in many infrastructure components (Zimmermann et al., 2017; Schmidt et al., 2014a; Provost and Fawcett, 2013). Thus, a considerable number of researchers have focused on the supportive role of big data in the collection of unstructured information for EAM (Möhring and Schmidt, 2015; Hacks and Saber, 2016). Hacks and Saber (Hacks and Saber, 2016) explain the state of the art of the usage of big data technologies for EAM. They evaluated how different big data frameworks (e.g., Hadoop (Ghazi and Gangodkar, 2015)) can be used to support specific EAM requirements. A methodology to deploy big data for information collection was introduced by Vanauer et al. (Vanauer et al., 2015). This publication provides guidelines for implementing big data technologies in the EAM context.

Second, the use of semantic technologies for information collection in EAM was proposed by Ortmann et al. (Ortmann et al., 2014) and Rosina

(Rosina, 2015). The idea behind semantic technologies is to leverage the semantic value of the collected information. The responsibility to keep the original information up to date remains with the information's owner. This concept allows accessing current information from diverse information sources independently of the original format and terminology. Rosina showed how, based on the use of these semantic technologies, EA information could be collected, formalized, and integrated into EAM.

Third, in order to enable the collection and integration of a growing diversity of information for EAM, a concept that uses service-oriented technologies was presented by Zimmermann et al. (Zimmermann et al., 2015). The goal of this concept is to foster digital transformation based on a holistic EAM approach. This approach integrates information from the Internet of Things into EAM. Zimmermann et al. focus on the extension of a static enterprise-internal architecture to accommodate the flexible and adaptive digitization of new information sources that come from an enterprise-external environment (Zimmermann et al., 2016; Zimmermann et al., 2017). This extension is realized with the help of microservice technologies that enable the integration of information sources into an EA.

4.4 Challenges

The authors of the publications examined in this literature review identified a variety of challenges. In order to create a systematic overview on these challenges, we defined the following categories: (i) exploitation of information, (ii) information quality, and (iii) EA governance. They are briefly discussed in the following section.

4.4.1 Exploitation of Information

The work of Hacks and Saber (Hacks and Saber, 2016; Katal et al., 2013) identified challenges regarding the exploitation of information for EAM, including the difficulty of identifying relevant, accurate information to support EAM decision making. For example, enterprise architects have to make decisions influencing the software release management process of an EA. Therefore, it is essential to have accurate information about the restrictions of the different versions of software applications to be released.

Besides identifying information, it was also highlighted that unstructured information cannot be uniformly analyzed and are more challenging to process for big data technologies. More research is required to investigate how big data technologies may be used in the analysis of EAM-relevant unstructured information. As objects of further investigation, information about the functional scope of software solutions and its supported business processes may be used.

Moreover, Zimmermann et al. (Zimmermann et al., 2015) noted that the integration of a vast number of dynamically growing systems and services, such as Microservices and the Internet of Things, presents a considerable challenge for the scalability, extension, and evolution of EA models. In this context, the processing of raw information objects, such as the output of network scanners, to fulfill their specific purposes within EAM (e.g., into an EA model) remains a challenge (Holm et al., 2014). One obstacle is the difficulty of defining clear rules for this processing. This was also emphasized by Välja et al. (Välja et al., 2015) and Farwick et al. (Farwick et al., 2016).

Furthermore, Ortmann et al. (Ortmann et al., 2014) highlighted the challenge of identifying relationships between different information objects in order to derive new knowledge regarding the EA. For example, business partners might be modeled as information objects. For enterprise architects, it is vital to know at what date the object business partner is stored by which software application within the EA. Therefore, the identification of relationships between objects is essential.

4.4.2 Information Quality

Information quality plays a crucial role when integrating unstructured information into EAM. Unstructured information has to be reliable to be able to draw the right conclusions. Holm et al. (Holm et al., 2014) used network scanners to identify system software, applications, and interfaces and assessed the quality of the resulting information. They discussed the challenges of tracking the quality of this information over time in order to create a coherent view.

Furthermore, Fittkau et al. (Fittkau et al., 2015) and Roth (Roth, 2014) outlined the challenge of maintaining consistency between EA models and information sources. This problem might be traced back to manual information collection processes, which are still in place in many companies.

Finally, Farwick et al. (Farwick et al., 2016) characterized the research into the information quality issue as only a small island within the literature, indicating a profound lack of research in this field.

4.4.3 EA Governance

In regard to managing the changes required for the integration of unstructured information into an EA, several challenges on the EA governance level can be identified. The involvement of stakeholders with differing or even contradicting but interrelated areas-of-interest was mentioned as a challenge by Rosina (Rosina, 2015) and Schmidt et al. (Schmidt et al., 2014b). For example, the integration of publicly available unstructured information about customers such as product feedback, opinions, or interests can reveal the differing goals of stakeholders within a company. A sales department might be interested in improving its customer care activities, while the IT department might want to reduce its operational support resources.

Moreover, various stakeholders often use different vocabulary to describe the same information object. Thus, it is difficult to share, exchange, or consolidate information about integration approaches for EAM. Further research should focus on the topic of how to facilitate the communication of enterprise architecture transformations and the involvement of different stakeholders.

5 DISCUSSION

In the following section, we discuss the key findings and limitations of the research at hand.

5.1 Key Findings

Based on the results of our systematic literature review and the subsequent analysis, we derived the following five key findings.

Key Finding 1: There is a profound lack of surveys and case studies regarding requirements and sources for collecting unstructured information for EAM.

Our systematic literature review and the related work indicated a lack of surveys and case studies investigating the collection for unstructured information for EAM (cf. section 4.2). This lack suggests that researchers are not leveraging the advantages of surveys and case studies in this field. Case studies hold the potential to gain insights into many details that would not usually be easily obtained by other research methodologies. The results of case studies are usually richer and of greater depth than can be obtained through other experimental designs. For example, a case study might clarify the potential use cases that arise from collecting unstructured information for EAM. The advantage of a survey is the ability to gather qualitative feedback about the need for collecting and integrating unstructured information into EAM. Surveys may help to assess the expected utility value of collecting unstructured information for EAM. In summary, more research, such as a survey and a case study, is needed to determine the requirements for the implementation of tools and frameworks for automated collection of relevant unstructured information.

Key Finding 2: Leveraging dynamically changing unstructured information holds the potential to predict required EA changes in the future.

Our investigations showed that existing research does not leverage dynamically changing unstructured information (cf. section 4.3.2). To identify the required EA adjustments, tracing of dynamically changing information is helpful. For example, rapid growth in the size of log file entries might lead to a lack of storage resources. The footprint of how users interact with a GUI (graphical user interface) can point out the popularity of different functionalities. This dynamic un- and semistructured information might pave the way to predict required EA changes in the future. Moreover, leveraging this information allows us to react more quickly to changing EA vulnerabilities, thus giving us more time to make the needed adjustments to the EA. Linking different information types (e.g., correlating highly volatile information with static information) would also enable identification of legacy IT systems. This identification can be made by linking real-time operational information and static EA model information of a particular IT system and outlining potential mismatches between them.

Key Finding 3: Investigations regarding enterprise-external information sources or trigger events for EAM and their relationships are missing.

This review confirmed that the majority of EAM-relevant information originates from enterprise-internal sources (cf. section 4.3.4). The authors of existing research did not mention any enterprise-external information sources that are used within EAM. However, the need to collect enterprise-external sources is explicitly expressed in literature (Zimmermann et al., 2017). This fact reveals the need for more research focusing on enterprise-external information sources for EAM. Our investigations showed that currently, information collection processes are triggered mainly by enterprise-internal events. However, for the collection of enterprise-external information sources, it can also be necessary to leverage trigger events from an enterprise-external environment. The question of which trigger events for enterprise-external information sources are adequate to trigger information collection processes automatically remains open.

Key Finding 4: Technology is on the rise that tackles long-standing challenges regarding the collection of unstructured information for EAM.

In summary, this review discusses three general technologies that enable the collection of unstructured information for EAM (cf. section 4.3.6). First, big data technologies can store and analyze massive amounts of unstructured information. This might allow us to handle the growing amount of unstructured information from an enterprise-external environment. Second, semantic technologies can integrate highly dynamic information from diverse information sources by keeping their timeliness. The advantage of using these semantic technologies is that the responsibility to maintain the source information remains with the information's owner, and EAM could collect the information without needing to spend effort on managing mass storages and maintaining information quality. Moreover, service-oriented technologies seem to be adequate to manage a growing diversity of EAM information. It is possible to easily reuse and adjust existing services to change and extend the EA repository.

Key Finding 5: Information quality remains a crucial issue for the collection of unstructured information for EAM.

Only a few researchers highlighted and considered the quality of the collected information (cf. section 4.3.3). It is challenging to derive specific further actions without adequate information quality. However, there has not yet been sufficient research into the required quality of unstructured information for EAM. Accordingly, further investigations are necessary to develop appropriate measures to guarantee the necessary information quality for EAM.

5.2 Limitations

The research at hand might be limited by a (i) selection bias of papers, (ii) false classification and analysis, (iii) missing papers and (iv) limited generalizability of results. In order to overcome (i), this review is based on the well-established guidelines by Wohlin et al. (Wohlin, 2014). A detailed description of the implementation of these guidelines can be found in Section 3. In order to overcome (ii), we provided definitions of the classification criteria and analyzed the papers based on them. Moreover, each of the observed papers was read at least by two authors of the publication at hand. Moreover, there is the possibility that we (ii) missed relevant papers. Since we applied the snowballing methodology, and recent studies have shown that it delivers comparable results to other research methodologies for conducting systematic literature studies (Jalali and Wohlin, 2012; Badampudi et al., 2015), the risk of (iii) is at an acceptable level. Finally, there might be the risk of (iv). We reduced this risk by applying the snowballing methodology until no new papers were found. In doing so, we provided a comprehensive overview of existing research in the field.

6 CONCLUSION AND FUTURE WORK

This paper presents a systematic literature review that provides a comprehensive overview of existing research into the collection and use of unstructured information for EAM. It uses the snowballing methodology to identify relevant literature and classifies it according to the following criteria: bibliographic information, research methodology, research topics, and research challenges. In total, we identified and classified 20 relevant publications. Our investigations showed that there is a profound lack of research regarding the requirements and sources for collecting unstructured information for EAM. We also determined that leveraging the dynamic perspective of unstructured information enables the prediction of required EA changes in the future. However, we found that there has been little research regarding enterprise-external information sources or trigger events for EAM and their relationships. Our investigations also showed that there is technology on the rise that can help tackle long-standing challenges regarding the collection of unstructured information. Finally, we determined that there is a lack of research regarding the required information quality of unstructured information for use in EAM. In summary, the amount of unstructured information is continuously growing, and some early steps have been undertaken to leverage this information for EAM. However, the results of this review suggest that enterprise-external information sources have not yet been investigated in detail. Further research is required to identify relevant enterprise-external information that supports EAM. Because researchers have not undertaken any surveys in this field, we plan to conduct an exploratory survey. The goal of this survey is to provide a comprehensive view of the relevant enterprise-external information sources from a research and practice perspective.

REFERENCES

Atkinson, R., & Flint, J. (2001). Accessing hidden and hard-to-reach populations: Snowball research strategies. *Social research update*, *33*(1), 1-4.

Babar, Z., & Yu, E. (2015, June). Enterprise architecture in the age of digital transformation. In *International Conference on Advanced Information Systems Engineering* (pp. 438-443). Springer, Cham.

Badampudi, D., Wohlin, C., & Petersen, K. (2015, April). Experiences from using snowballing and database searches in systematic literature studies. In *Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering* (p. 17). ACM.

Bakshi, K. (2012, March). Considerations for big data: Architecture and approach. In *2012 IEEE Aerospace Conference* (pp. 1-7). IEEE.

Becker, H., Naaman, M., & Gravano, L. (2009, June). Event Identification in Social Media. In *WebDB*.

Becker, H., Naaman, M., & Gravano, L. (2011, July). Beyond trending topics: Real-world event identification on twitter. In *Fifth international AAAI conference on weblogs and social media*.

Budgen, D., & Brereton, P. (2006, May). Performing systematic literature reviews in software engineering. In *Proceedings of the 28th international conference on Software engineering* (pp. 1051-1052). ACM.

Buschle, M., Holm, H., Sommestad, T., Ekstedt, M., & Shahzad, K. (2011, June). A Tool for automatic Enterprise Architecture modeling. In *International Conference on Advanced Information Systems Engineering* (pp. 1-15). Springer, Berlin, Heidelberg.

Diefenthaler, P. I. (2016). Interactive Transformation Path Generation in Enterprise Architecture Planning.

Farwick, M., Agreiter, B., Breu, R., Ryll, S., Voges, K., & Hanschke, I. (2011). Requirements for automated enterprise architecture model maintenance. In *13th International Conference on Enterprise Information Systems (ICEIS), Beijing*.

Farwick, M., Breu, R., Hauder, M., Roth, S., & Matthes, F. (2013, January). Enterprise architecture documentation: Empirical analysis of information sources for automation. In *2013 46th Hawaii International Conference on System Sciences* (pp. 3868-3877). IEEE.

Farwick, M., Schweda, C. M., Breu, R., & Hanschke, I. (2016). A situational method for semi-automated Enterprise Architecture Documentation. *Software & Systems Modeling*, *15*(2), 397-426.

Fischer, R., Aier, S., & Winter, R. (2007). A federated approach to enterprise architecture model maintenance. *Enterprise Modelling and Information Systems Architectures (EMISAJ)*, *2*(2), 14-22.

Fittkau, F., Roth, S., & Hasselbring, W. (2015, May). ExplorViz: Visual runtime behavior analysis of enterprise application landscapes. AIS.

Galliers, R. D., & Land, F. F. (1987). Choosing appropriate information systems research methodologies. *Communications of the ACM*, *30*(11), 901-902.

Ghazi, M. R., & Gangodkar, D. (2015). Hadoop, MapReduce and HDFS: a developers perspective. *Procedia Computer Science*, *48*, 45-50.

Grunow, S., Matthes, F., & Roth, S. (2013). Towards automated enterprise architecture documentation: Data quality aspects of SAP PI. In *Advances in Databases and Information Systems* (pp. 103-113). Springer, Berlin, Heidelberg.

Hacks, S., & Saber, M. (2016). Supportive Role of Big Data Technologies in Enterprise Architecture Management. *Full-scale Software Engineering/Current Trends in Release Engineering*, *13*, 37.

Hendriks, V. M., Blanken, P., Adriaans, N. F. P., & Hartnoll, R. (1992). *Snowball sampling: A pilot study on cocaine use*. IVO, Instituut voor Verslavingsonderzoek, Erasmus Universiteit Rotterdam.

Holm, H., Buschle, M., Lagerström, R., & Ekstedt, M. (2014). Automatic data collection for enterprise architecture models. *Software & Systems Modeling*, *13*(2), 825-841.

Harzing, A.-W. (2019). *Publish or Perish* (Version 6.34.6288.6798): Tarma Software Research. Available under https://harzing.com/resources/publish-or-perish [29.10.2019].

Jalali, S., & Wohlin, C. (2012, September). Systematic literature studies: database searches vs. backward snowballing. In *Proceedings of the 2012 ACM-IEEE international symposium on empirical software engineering and measurement* (pp. 29-38). IEEE.

Johnson, P., Ekstedt, M., & Lagerstrom, R. (2016, September). Automatic probabilistic enterprise IT architecture modeling: a dynamic bayesian network approach. In *2016 IEEE 20th International Enterprise Distributed Object Computing Workshop (EDOCW)* (pp. 1-8). IEEE.

Katal, A., Wazid, M., & Goudar, R. H. (2013, August). Big data: issues, challenges, tools and good practices. In *2013 Sixth international conference on contemporary computing (IC3)* (pp. 404-409). IEEE.

Kirschner, B., & Roth, S. (2014). Federated enterprise architecture model management: collaborative model merging for repositories with loosely coupled schema and data. In *Multikonferenz Wirtschaftsinformatik*.

Langenberg, K., & Wegmann, A. (2004). *Enterprise architecture: What aspects is current research targeting* (No. REP_WORK).

Lankhorst, M. (2009). *Enterprise architecture at work* (Vol. 352). Berlin: Springer.

Liggins, M. E. (2008). Handbook of Multi-Sensor Data Fusion, ME Liggins, DL Hall, and J. Llinas, Eds.

Google LLC. Google scholar - bibliographic database. https://scholar.google.de/, 09 2018. Accessed: 30-Oct-2019.

Luftman, J., & Brier, T. (1999). Achieving and sustaining business-IT alignment. *California management review*, *42*(1), 109-122.

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & López-Cózar, E. D. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, *12*(4), 1160-1177.

Martin, R. A. (2008, November). Making security measurable and manageable. In *MILCOM 2008-2008 IEEE Military Communications Conference* (pp. 1-9). IEEE.

Möhring, M., & Schmidt, R. (2015). Daten-getriebene Unternehmensarchitekturen im E-Commerce für das präventive Retourenmanagement. *INFORMATIK 2015*.

O'Leary, D. E. (2013). Artificial intelligence and big data. *IEEE Intelligent Systems*, *28*(2), 96-99.

Ortmann, J., Diefenthaler, P., Lautenbacher, F., Hess, C., & Chen, W. (2014). Unternehmensarchitekturen mit semantischen technologien. *HMD Praxis der Wirtschaftsinformatik*, *51*(5), 616-626.

Provost, F., & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. " O'Reilly Media, Inc.".

Rainer, R. K., Cegielski, C. G., Splettstoesser-Hogeterp, I., & Sanchez-Rodriguez, C. (2013). *Introduction to information systems: Supporting and transforming business*. John Wiley & Sons.

Rea, L. M., & Parker, R. A. (2014). *Designing and conducting survey research: A comprehensive guide*. John Wiley & Sons.

Rosina, P. C. (2015). Semantic Web Technologies for Method Management: Model, Methodology and Enterprise Architecture Integration.

Ross, J. W. (2003). Creating a strategic IT architecture competency: Learning in stages.

Ross, J. W., Weill, P., & Robertson, D. (2006). *Enterprise architecture as strategy: Creating a foundation for business execution*. Harvard Business Press.

Roth, S. (2014). *Federated Enterprise Architecture Model Management* (Doctoral dissertation, Technische Universität München).

Roth, S., & Matthes, F. (2014). Visualizing differences of enterprise architecture models. In *International Workshop on Comparison and Versioning of Software Models (CVSM) at Software Engineering (SE). Kiel, Germany*.

Schmidt, R., Wißotzki, M., Jugel, D., Möhring, M., Sandkuhl, K., & Zimmermann, A. (2014a, September). Towards a framework for enterprise architecture analytics. In *2014 IEEE 18th International Enterprise Distributed Object Computing Conference Workshops and Demonstrations* (pp. 266-275). IEEE.

Schmidt, R., Zimmermann, A., Möhring, M., Jugel, D., Bär, F., & Schweda, C. M. (2014b, September). Social-software-based support for enterprise architecture management processes. In *International Conference on Business Process Management* (pp. 452-462). Springer, Cham.

Steinberg, A. N., & Bowman, C. L. (2017). Revisions to the JDL data fusion model. In *Handbook of multisensor data fusion* (pp. 65-88). CRC Press.

Valenduc, G., & Vendramin, P. (2017). Digitalisation, between disruption and evolution. *Transfer: European Review of Labour and Research*, *23*(2), 121-134.

Välja, M., Korman, M., Lagerström, R., Franke, U., & Ekstedt, M. (2016, September). Automated architecture modeling for enterprise technology manageme using principles from data fusion: A security analysis case. In *2016 Portland international conference on management of engineering and technology (PICMET)* (pp. 14-22). IEEE.

Välja, M., Lagerström, R., Ekstedt, M., & Korman, M. (2015, September). A requirements based approach for automating enterprise it architecture modeling using multiple data sources. In *2015 IEEE 19th International Enterprise Distributed Object Computing Workshop* (pp. 79-87). IEEE.

Vanauer, M., Böhle, C., & Hellingrath, B. (2015, January). Guiding the introduction of big data in organizations: A methodology with business-and data-driven ideation and enterprise architecture management-based implementation. In *2015 48th Hawaii International Conference on System Sciences* (pp. 908-917). IEEE.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly*, xiii-xxiii.

Wieringa, R., Maiden, N., Mead, N., & Rolland, C. (2006). Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. *Requirements engineering*, *11*(1), 102-107.

Wißotzki, M., & Sonnenberger, A. (2012). Enterprise architecture management-state of research analysis & a comparison of selected approaches. In *5th IFIP WG 8.1 Working Conference, PoEM 2012* (pp. 37-48).

Wohlin, C. (2014, May). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th international conference on evaluation and assessment in software engineering* (p. 38). ACM.

Zimmermann, A., Bogner, J., Schmidt, R., Jugel, D., Schweda, C., & Möhring, M. (2016). Digital enterprise architecture with micro-granular systems and services.

Zimmermann, A., Schmidt, R., Sandkuhl, K., Jugel, D., Bogner, J., & Möhring, M. (2017, October). Decision management for micro-granular digital architecture. In *2017 IEEE 21st International Enterprise Distributed Object Computing Workshop (EDOCW)* (pp. 29-38). IEEE.

Zimmermann, A., Schmidt, R., Sandkuhl, K., Wißotzki, M., Jugel, D., & Möhring, M. (2015, September). Digital enterprise architecture-transformation for the internet of things. In *2015 IEEE 19th International Enterprise Distributed Object Computing Workshop* (pp. 130-138). IEEE.