Enterprise Architecture Documentation:
Empirical Analysis of Information Sources for Automation

Matthias Farwick, Ruth Breu
University of Innsbruck, Austria
Institute of Computer Science
{matthias.farwick, ruth.breu}@uibk.ac.at

Matheus Hauder, Sascha Roth, Florian Matthes
Technische Universität München, Germany
Lehrstuhl für Informatik 19
{matheus.hauder, roth, matthes}@tum.de

Abstract

Over the past decade, Enterprise Architecture (EA) management emerged to a mature discipline commonly applied to realize cost saving potentials while increasing effectiveness of IT in organizations. Typically, EA management starts by documenting the current state in an EA model and deriving future planned states heading towards an optimized EA. In practice, organizations struggle with the documentation of their current state due to the complexity of their enterprise architecture and its frequent changes. Current research activities seek to automate the data collection process by integrating existing EA information sources of operative systems. However, a comprehensive analysis of possible information sources and their appropriateness for EA is not yet provided. To build an empirical basis, this paper presents findings of a survey conducted on the key problems in EA documentation as well as the appropriateness of specific EA information sources for automation with respect to provided data types and data quality.

1. Introduction

Rapidly changing market requirements and globalization compel organizations to adapt their information technology (IT) management practices [1,20]. EA management (EAM) is promoted as an instrument to improve the alignment of business and IT, realize cost saving potentials, and increase availability and fault-tolerance [11,14,15]. While classical software engineering approaches focus on details, EAM conveys a holistic view of the entire organization including information about business, IT, and their interrelations.

In discussions with industry partners we diagnose that today's organizations easily end up with several thousand applications. Due to this complexity the documentation of EA information is often regarded as time consuming, cost intensive, and error-prone [6,7,9]. Despite the variety of available EA frameworks [18], the documentation of EA information, i.e. the collection and maintenance of data that makes up the model

of the EA, is not considered detailed enough in current EA literature to assist organizations with this challenge. E.g. in the most-widely applied framework, The Open Group Architecture Framework (TOGAF), the documentation is only briefly covered [19].

In 2004, ter Doest et al. stated in [5]: *"in 7 years from now, enterprise architecture will be a real-time tool for management and redesign of the enterprise for better performance, flexibility and agility.[...]There will no longer [be] a boundary between the descriptive nature of architecture models and the operational side comprising a.o. business process management and IT management"*. While this is partly fulfilled for systems operation management it is clearly not fulfilled in the EA context where manual EA documentation is still prevailing. Although researchers have identified this issue [6,9,12], only recently solution approaches have been undertaken to overcome EA-information collection issues.

In order to bring this vision of ter Doest et al. closer to reality we envision the use of external information sources, i.e. existing operative IT systems, to automate updates of EA models or produce events that trigger manual changes. Such information sources may deliver structured data from already existing data silos to an EA model and can thus increase the actuality of EA models.

Initial efforts [3,4] show the usage of Network Scanners or Enterprise Service Buses (ESB), as automatically integrated EA information sources. These efforts indicate improvements to the EA model data quality attributes such as actuality and completeness. However, the relevance of these data silos as EA information sources as well as the appropriateness of the provided data and the generalizability of the approaches is not well grounded.

In order to improve the methods of EA data collection we argue that future research in this direction requires profound analysis of possible information sources and their applicability for EA documentation in enterprises. In order to lay the ground for this research

we conducted an online survey among 123 EA professionals to target the following research questions:

- What is the status quo of EA documentation processes in organizations?
- Which productive systems contain relevant EA information?
- What data quality, w.r.t. actuality, completeness, correctness and correct granularity, can be expected from these systems?
- What are typical integration problems for the information sources found?

The main contribution of this paper is the discussion of the three latter research questions based on the results of our survey. To the best of our knowledge it presents the first empirical analysis of information sources for automated EA documentation.

The remainder of this paper is structured as follows. After revisiting related work, the paper continues with describing the details of the questionnaire. We continue with discussing our findings of the current practices and challenges in EA documentation. Subsequently, we present the information sources and their suitability for automated EA documentation. A discussion of the key findings is provided and important research directions towards improved methods for EA documentation are pointed out. The paper concludes with a brief summary and a pointer to our immediate future work.

2. Related work

Only little in-depth research has been conducted in the area of EA documentation and in particular its automation. However, recent approaches try to automate EA documentation seeking to reduce cost and effort while at the same time increasing data quality.

As outlined above, empirical groundings for EA documentation automation mechanism are scarce. The only empirical work containing information on this topic is a survey conducted by Winter et al. [21]. According to their data, the majority (72.6%) of their survey participants still fully rely on manual maintenance of their EA model.

In [5] ter Doest et al. present the ArchiMate Workbench and detail on the importance of a well-defined meta-model for EAM. The paper describes how an EA tool embeds in an application landscape. Thereby, the authors give a first impression on respective EA information sources, e.g. monitoring or reporting tools that could be used for automation. However, the authors did not intend to detail on automation.

In line with ter Doest et al. [5], the more recent vision paper by Brückmann et al. [2] re-iterates the idea of the real time aspect EA models. Thereby, they apply the "control room" concept employed in power plants to EAM and detail the "business software control center" as a visionary sketch on how to present the results of an automated EA documentation to different stakeholders. However, Brückmann et al. do not detail how to achieve an automated alignment of EA models in practice.

The federated data collection approach by Fischer et al. [8] presents a more practical approach to EA data collection. Fischer et al. discuss the shortcomings of existing approaches to EA information maintenance and propose a federated approach thereof, i.e. collecting data in organizational units and pushing it to a central repository. Further they report on its implementation at a large financial service provider. Nevertheless, the work does not present implementation details for the data collection or which external information sources could be included for automation.

The work by Moser et al. [13] goes a step further in the direction of automating EA documentation by providing data collection processes that explicitly include automated information sources. Again, Moser et al. do not detail which information sources are envisioned and which data types could be provided.

In [6] we present requirements for an automated EA model maintenance method or process. Based on these requirements we propose automation processes for EA documentation [7] to reduce the amount of manual work when documenting the EA. Thereby, they propose (semi-) automated processes gathering information from both, human input and technical interfaces. However, so far we did not detail concrete information sources and respective data quality.

Buschle et al. [4] use a vulnerability scanner to automate EA documentation. Their approach is limited to the infrastructure layer, i.e. information like operating systems can be extracted with this approach. In [3] the authors show the coverage of the ArchiMate meta-model when using an ESB as an EA information source.

Summing up our findings in current literature, there is strong evidence for scientific relevance for automated EA documentation. However, to the best of our knowledge, no research exists on the appropriateness of specific information sources for EA with regard to their provided data types and the quality of the provided data.

3. Questionnaire and dataset

The final version of the questionnaire was published as an online survey available for 14 days. Prior to the opening of the survey it was filled out by three researchers in the field of EA not involved in this research and adapted according to their suggestions. Over 1100 survey invitations were sent by e-mail to

EA related experts. This list of experts was compiled during EA projects we performed with industry partners in recent years. In addition, the survey was announced in well-known online forums on Xing[1], LinkedIn[2] and Ning.com[3] that are related to EA or strategic IT management topics. The survey consisted of three parts. The first part included questions for the general categorization of the organizations in the survey, such as industry sector or job title of the participant. These are presented below. The second part aimed at gathering currently applied practices in collecting EA information in organizations and in particular to find out what type of automation is already applied for EA documentation. Respective findings of this part are summarized in Section 4. Target of the main and final part was to gather data on potential information sources that could be used to automatically collect EA-relevant data and integrate it in an EA model. The results of this part of the dataset are the core of this paper and are presented in Section 5 and discussed in Section 6.

We received 179 answers containing participants from, e.g., Canada, Germany, Great Britain, India, New Zealand, South Africa, Switzerland, and USA. During the survey 56 participants, i.e. 31%, dropped out during the questionnaire resulting in 123 completed answers for the evaluation. Table 1 illustrates the distribution of the industry sectors of the organizations in the survey with *Finance* as the largest sector followed by *IT, Technology* and *Government*.

Table 1. Organizations by industry sector

Industry sector	Count	Percentage
Finance	37	30.08%
IT, Technology	23	18.70%
Government	11	8.94%
Communications	8	6.50%
Transportation	8	6.50%
Education	7	5.69%
Manufacturing	6	4.88%
Services	5	4.07%
Retail	5	4.07%
Health Care	5	4.07%
Construction	2	1.63%
Agriculture	1	0.81%
Other	5	4.07%

Table 2 illustrates the distribution of the job titles of all participants. The largest groups in our survey are

Enterprise Architects with ~55% and Enterprise Architecture Consultants with ~17%. The consultants were asked to complete the survey with respect to one specific customer in order to receive coherent answers with the other participants. In addition the distribution shows ~4% in an upper management position (CxOs). The dataset also exposes an average EA experience of ~5.4 years of the participants, with only three outliers who have less than one year experience and 27 participants with more than 10 years of experience. The organizations of the participants apply EAM for an average of ~4.2 years.

Table 2. Participants by job title

Job title	Count	Percentage
Enterprise Architect	68	55.28%
Enterprise Architect Consultant	22	17.89%
Software Architect	8	6.50%
Project Manager	5	4.07%
IT Manager	5	4.07%
CTO	4	3.25%
CIO Software Developer	3	2.44%
CIO	1	0.81%
Software Development Manager	1	0.81%
Agriculture	1	0.81%
Other	5	4.07%

The results of our survey confirm the findings identified in literature [21] that most EA endeavors are IT-driven (51.22%), or both business and IT-driven (39.02%). Only 8.94% describe their EA endeavor as business-driven. However, 73.17% of these EA endeavors have upper management support whereas only 26.83% do not. This suggests that the benefits of EAM (e.g. [17,20]) are well-known in organizations.

In the following, we briefly discuss our findings on the current state of EA documentation.

4. Current practices and challenges in enterprise architecture documentation

This section gives a brief summary of the second part of the survey with the current practices in EA documentation and the type of already applied automation.

When asked about the challenges in EA information collection 67.48% of all participants confirmed that EA information collection is very time consuming, and the EA model is hard to keep up-to-date (45.53%). This confirms the findings of [9] that the maintenance of EA models is a major problem in the EA field.

Our dataset also shows that the information collection for modeling the AS-IS state of the EA is mainly a

manual process that is triggered by organizational events, such as project completion or releases of new application versions. The participants stated that the data is, among others, collected by conducting interviews (62.60%), workshops (47.97%) and questionnaires (32.52%). However, only a minority (22.76%) of the participants reported that this data collection is following a dedicated and specified process. Hence, it can be stated that the maturity of the quality assurance processes for EA models is at a low level in most organizations.

To find out the current state in automated EA documentation mechanisms we asked the participants whether they apply automation for their EA documentation. Only 23 participants (18.70%) stated that they have implemented some form of automated EA documentation mechanisms for their EA tool, 79 have no automation implemented and 21 do not know or do not use an EA tool. In most of the 23 cases the automation relies on the manually triggered import from CSV, XML, or MS Excel files (65.21%). In significantly less cases the data is pushed from an external information source to the EA repository (30.43%). In only 6 cases (26.08%) full automation is implemented with automated data pushes from external sources to the EA tool or pulls by the EA tool from an external tool.

We then asked the 79 participants who have no automation implemented yet, about their preferences on automated EA documentation and to describe their intentions. 31.24% of these organizations plan to implement automated EA documentation mechanisms in the future and 38.75% have not yet considered automated EA documentation. 30% rule out the usage of automation techniques. This leads to 39.02% of all the participants who either already use automation or plan the implementation in the future.

Our findings from this part of the survey are:

- Documentation of EA information is a major challenge for organizations.
- EA documentation is regarded as very time consuming and the data quality is not sufficient.
- The majority of organizations have no dedicated process for EA documentation defined.
- Some organizations have already implemented a form of automated EA documentation. However, this documentation is mostly limited to simple file import mechanisms that are manually triggered.
- The direct data integration between other information systems and the EA tool is only considered by few organizations.

These findings further motivate the analysis of automated EA documentation from external information sources.

5. Information sources for enterprise architecture documentation

As stated in Section 2 several research groups recently started to investigate methods for automated EA documentation from existing data silos. Research efforts in this direction so far presented approaches for integrating a Network Scanner [4] and an ESB [3].

In the main part of the survey, participants were asked to answer whether their organization uses specific information sources and their judgments on data quality attributes of these sources. In particular, we asked for the usage of Network Scanners, Configuration Management Databases (CMDBs), Project Portfolio Management tools, Change Management tools and License Management tools. This list has been compiled from a literature review and our practical experience in the field. Additionally, participants were asked to add details of potential sources that they find missing in order to elicit more potential sources in an explorative manner.

Figure 1 shows the number of organizations using a specific tool (dark grey bar) and those who deem the tool as a relevant source for EA information (light grey bar).
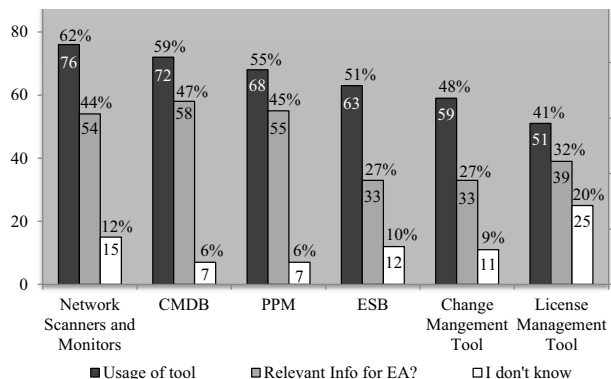

Figure 1. Usage and relevance as EA information sources (n=123).

The white bars represent participants that do not know whether the respective tool is used in their organization. Except Change Management and License Management tools, all others are being widely used by the participating organizations, i.e. by more than 50%. The small discrepancy between the two bars shows that in the majority of all cases the usage of a tool implies the existence of a data silo containing relevant EA data. Researchers can use this data to prioritize their efforts to devise mechanisms for the integration, whereas practitioners may be able to use this information to evaluate their respective information sources (application landscape) for automating their EA documenta-

tion. In the following we further elaborate the findings of each information source in detail.

5.1 Information sources

For each information source we asked the participants which tool they use and which types of EA relevant data the specific tool could potentially deliver. In addition, we questioned on the data quality attributes actuality, completeness, correctness and appropriate granularity in the EA context. With granularity we refer to how difficult the mapping between the information source data format and the EA-model is, as well as how large the semantic gap is.

The participants were asked to judge the quality attributes on a Likert scale from 1 (very bad) to 5 (very good). For each information source we visualize the dataset with box-whisker plots. In descriptive statistics this type of chart is used to show 5 data characteristics, namely the minimum and maximum (the whiskers), the 2nd quartile (light grey box) and 3rd quartile (dark grey box), the median and the arithmetic average. Note that the sum of all answers in the second and third quartile contains 50% of the answers. Figures also show raw data where '-' stands for the answer 'I don't know'.

5.1.1. Network Monitors and Scanners.
First, we discuss the results for the potential use of Network Scanners as automated information sources for an EA model such as described in [4]. Operations teams often use these tools to monitor the infrastructure and its performance. As shown above in Figure 1, 62% of all organizations use Network Scanners. The most commonly mentioned tools are IBM Tivoli Network Manager (22.37%), Compuware dynaTrace (6.58%), and Wireshark (5.26%).

For each source we first asked which types of relevant EA data could be gathered from this source. As expected the most frequent answers for Network Scanners were *servers*, *applications* and *databases*. However, some participants raised the question whether this type of instance data is actually relevant in an EA context. This can be exemplified by this quote of one of the participants: *"I would clearly distinguish between - Enterprise Architecture modeling - Solution or System Architecture and CMDB based data architecture modeling"*. Others, while considering the data relevant, raised the concern that the granularity of the collected data is not in-line with the granularity of their EA model, as this answer exemplifies: *"The scope of the infrastructure monitoring tools is too granular and focused on the operations team"*.

This is also reflected by the questions regarding respective EA data quality attributes of the Network Scanner summarized in Figure 2.
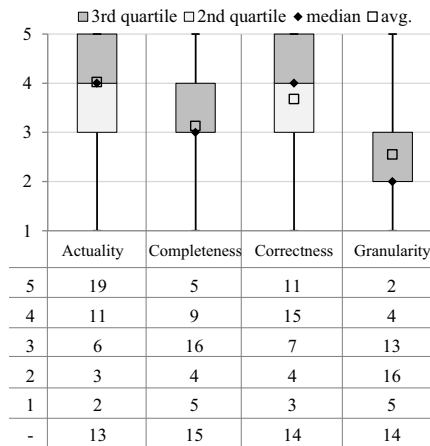

	Actuality	Completeness	Correctness	Granularity
5	19	5	11	2
4	11	9	15	4
3	6	16	7	13
2	3	4	4	16
1	2	5	3	5
-	13	15	14	14

Figure 2. Information source quality attributes of Network Scanners (n=54)

With today's sophisticated tools, the quality attributes actuality and correctness are considered positive with a median of 4. Since these tools cannot cover all technologies, it is reasonable that the median of the completeness attributes is 3. In contrast, the granularity attribute only has a median of 2 that is the worst result for all information sources covered in this survey. This shows that although 44% of all participants consider the data of this source as EA-relevant, there need to be mechanisms in place in order to raise the level of abstraction of collected data to an appropriate level. It is, however, a difficult task to automate this step, and it is likely that human judgment is indispensable in most cases.

5.1.2. Configuration Management Databases.
Configuration Management Databases (CMDBs) according to ITIL [10] are used in organizations to collect operations data such as hardware instances and related incidents. As such, CMDBs can potentially suffer similar data quality challenges as EA repositories, since the data often needs to be manually collected and cleansed.

Similar to the EA-relevant data that can be provided by a Network Scanner the participants answered that CMDBs could provide data about *server*, *database* and *application* instances. The answers were also similar concerning the free text answers where concerns on the data granularity were raised: *"Again, too fine-grained for use in an architecture context; lack of abstraction"*. Here again the different usages of the EA function became apparent. Some organizations also collect operations data, whereas others focus on the strategic planning aspect of EA as described in [16]. The latter one can be exemplified by this free text answer of a survey participant: *"CMDBs contain all instances, EA should be interested in type of devices. CMDB does not reveal the function of a device, only its existence"*

In Figure 1 one can see that of the 72 participants that make use of a CMDB in their organization, 58 think that the CMDB can deliver EA-relevant information. The most common CMDB tools used in these organizations are HP Universal (18.06%), BMC Atrium (15.28%), and IBM Tivoli CMDB (9.72%). Figure 3 shows the results for the EA-relevant data quality attributes of CMDB of those 58 participants.
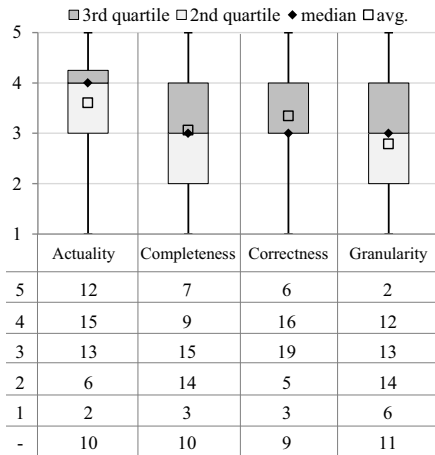

	Actuality	Completeness	Correctness	Granularity
5	12	7	6	2
4	15	9	16	12
3	13	15	19	13
2	6	14	5	14
1	2	3	3	6
-	10	10	9	11

Figure 3. CMDB information quality (n=58)

It is interesting to notice that although CMDBs have similar data collection issues as EA repositories, the actuality attribute still has a median of 4. A reason might be that CMDBs can better rely on infrastructure scanning mechanisms since the contained data types are more hardware related and at a lower level of abstraction. The other quality attributes of completeness, correctness and granularity all have a median of 3.

5.1.3. Project Portfolio Management tools. The concept of change as part of projects is a central aspect of EAM. Hence, EA relevant projects are often modeled as part of the EA model. As pointed out by one of the survey participants, Project Portfolio Management (PPM) can be an *"important [...] trigger point in [the] EA maintenance process"*. A different participant pointed out the strong relationship between the two disciplines by arguing that *"PM and EA should be integrated at process level"*. The participants identified the general concept of *architecture change project* including its *start and end date*, *budget information*, as well as the *artifacts affected by the projects* as data that could be automatically collected from this tool. However, some participants added that it is often not possible to automatically decide if a project is architecture relevant. The participants also mentioned that changes in the PPM tool could be used as triggers for events in an EAM tool.

As illustrated in Figure 1, 68 (55%) of the survey participants make use of a PPM tool in their organization. 55 of those believe that the PPM tool could automatically provide EA-relevant data. The most common tools are Microsoft Enterprise Project Management Solution (13.24%), CA Clarity PPM (10.29%), Primavera Enterprise PPM, and MindJet MindManager (both 5.88%). Figure 4 details results for data quality attributes of PPM tools.
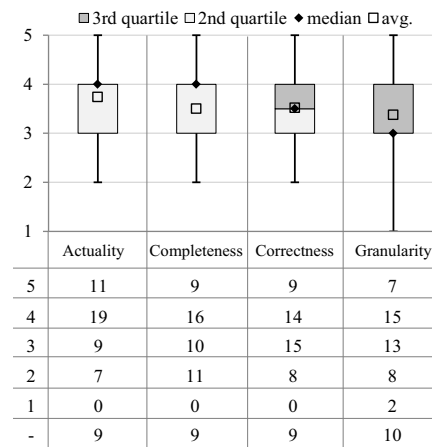

	Actuality	Completeness	Correctness	Granularity
5	11	9	9	7
4	19	16	14	15
3	9	10	15	13
2	7	11	8	8
1	0	0	0	2
-	9	9	9	10

Figure 4. Project Portfolio Management tool information quality attributes (n=55)

One can see that both, actuality and completeness have a median of 4 and correctness has a median of 3.5. This implies that in most cases the content of PPM tools is maintained regularly as part of project management. Also notice that actuality, completeness, and correctness received a minimum rating of 2, indicated by the whiskers. However, the granularity only has a median of 3. This can be attributed to the difficulty of deciding which project actually is architecture relevant, and hence is important to include in the EA model.

5.1.4. Enterprise Service Bus. The analysis and optimization of dependencies between applications is another central aspect of the EAM discipline. With the recent restructuring of the application landscape of many organizations towards a Service Oriented Architecture (SOA), ESBs are often used as the central mediating entity for inter-application communication. Since ESBs are configured to allow the communication between applications these configurations can be potentially collected as input for the EA repository [3]. This is also reflected by the corresponding question regarding the EA-relevant data that can potentially be collected from an ESB. The typically mentioned data types were *services*, *interfaces* and *applications*. However, again some participants remark that their *"ESB is too technical [...] for EA"*.

In Figure 1 one can see that of the 123 participants 63 (51%) run an ESB in their organization. Among the most popular tools are IBM WebSphere (28.57%), Oracle ESB (22.22%), and SAP PI (20.63%). Of these 63, 52 (42% of all participants) believe that their ESB potentially contains EA-relevant data. Figure 5 shows the respective data quality evaluation.
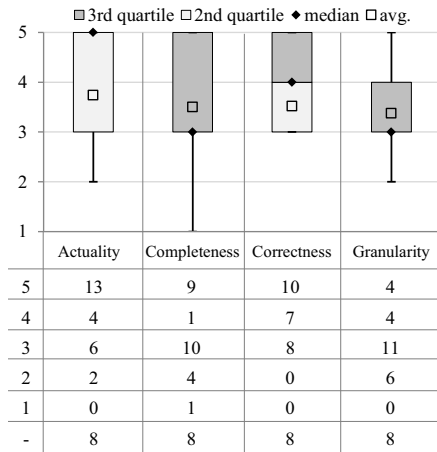

	3rd quartile	2nd quartile ◆ median	avg.	
	Actuality	Completeness	Correctness	Granularity
5	13	9	10	4
4	4	1	7	4
3	6	10	8	11
2	2	4	0	6
1	0	1	0	0
-	8	8	8	8

Figure 5. ESB information quality (n=58)

As an ESB is a runtime artifact that needs to be properly configured in order to function, it is obvious that the actuality and correctness data quality attributes of an ESB receive very positive evaluations, with a median of 5 for actuality and 4 for correctness; completeness receives a median of 3. The reason is very likely that many organizations are still in the transition phase between traditional point-to-point communication and a Service Oriented Architecture (SOA) based on an ESB. However, it is remarkable that 9 participants gave their ESB the best rating for completeness, which indicates that their inter-application communication almost fully relies on their ESB. For these organizations it could be a major benefit to integrate their ESB with their EA tooling in order to keep their model up-to-date. As mentioned before, the mapping between the data types provided by the ESB and the EA data types is a difficult task. This is also reflected by the answers regarding the difference between EA and ESB data granularity. Here the median of the answers is 3.

5.1.5. Change Management tools. Change Management tools are used in the context of ITIL in order to optimize the processes of implementing changes in the IT-landscape. The concept of this kind of change is at the heart of EAM so it is reasonable that integrating such a tool in the EA documentation process can be of value. The participants mentioned *change projects*, and the *statistics of tickets* for applications as the data types that could be collected. However, the granularity and

the mapping were again of concern. An interesting comment was given by one of the participants regarding the provided data: *"Change Management does not really deliver EA data, but it can deliver events to update EA data"*. Hence, such a tool could be used to automatically trigger manual update tasks to the EA tool, which would increase the timeliness of the EA data.

In Figure 1 one can see that of the 123 participants 59 (48%) use a Change Management tool in their organization. The most popular tool is IBM Rational ClearCase (23.73%). Of these 59, 33 participants believe that their tool can potentially provide EA-relevant data. Figure 6 shows the evaluation of the EA-relevant data quality attributes.
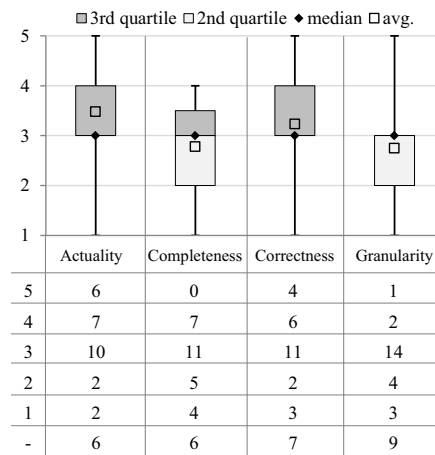

	3rd quartile	2nd quartile ◆ median	avg.	
	Actuality	Completeness	Correctness	Granularity
5	6	0	4	1
4	7	7	6	2
3	10	11	11	14
2	2	5	2	4
1	2	4	3	3
-	6	6	7	9

Figure 6. Change Management tool information quality attributes (n=33)

The data regarding the actuality of Change Management tools has a median of 3, implying that in most cases the manual maintenance of the tools is not optimal. Also the completeness of the data in the Change Management tools received a median of 3 with no selection of the most positive value. This implies that not all changes and change projects are covered by Change Management tools. Correctness and granularity also have a median of 3 with the former having a positive and the latter having negative tendency. The negative result for the granularity also indicates that the potentially provided data is not easy to map to EA-model elements. Hence, we argue that in the context of EA documentation automation, Change Management tools are more likely to be useful for producing events that trigger manual action, rather than actually directly providing data to an EA model.

5.1.6. License Management tools. License Management tools are used to keep track of the diverse acquired software licenses. They contain information

about the number of running installations, age as well as the type of license. As shown in Figure 1, 51 organizations currently have a License Management tool in use. The EA-relevant data types reported by the participants can mostly be summarized as calculated figures, such as the *number of installations*, *number of allowed users*, *costs*, *application age* as well as *type and duration of licenses*. In addition, the participants reported that *application types* could be gathered from such a tool. This seems to be the only data type that could potentially directly be mapped to an EA model from the License Management tool.

As depicted in Figure 1, 51 of the 123 participants use a License Management tool in their organization. Of those 51, 39 believe that their License Management tool could deliver EA-relevant data. Figure 7 shows the evaluation of the responses regarding the data quality of the EA-relevant data items of the License Management tools.
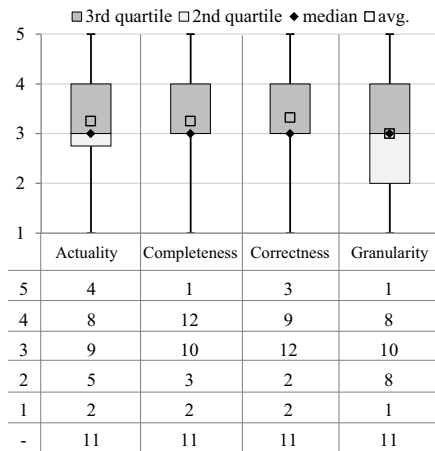

□3rd quartile □2nd quartile ◆ median □avg.			
	Actuality	Completeness	Correctness	Granularity

	Actuality	Completeness	Correctness	Granularity
5	4	1	3	1
4	8	12	9	8
3	9	10	12	10
2	5	3	2	8
1	2	2	2	1
-	11	11	11	11

Figure 7. License Management tool information quality attributes (n=39)

Actuality, completeness, and correctness all receive a median of 3 with a positive tendency, whereas the results for the granularity are centered at the median of 3. The tool could also be used to produce events that trigger manual changes based on license expiry events. Notice that a relatively large percentage of the participants (28%) state that they cannot make a statement regarding the data quality attributes of the tool. This might be an indicator that the stakeholders with EA roles, do not access the License Management tool on a regular basis. For this information source it is obvious that the effectiveness of its usage for collecting EA-data depends on the percentage of licensed software in an organization.

5.1.7. Additionally mentioned information sources. As already mentioned we also asked the participants to point out the potential information sources they found missing in the survey. As many organizations still heavily rely on keeping data in *Microsoft Excel* sheets, it is obvious that it was mentioned several times as a potential information source. Importing Excel sheets is already possible in many EA tools. With the recent trends towards the usage of virtualization techniques, the usage of cloud services was mentioned frequently for collecting infrastructure data. Additionally mentioned tools were also directory services to collect organizational structure, business process engines to collect process information, and release management tools to trigger release events.

Finally, the participants mentioned *business analysis* tools as another potential information source. Several participants also argue for standard data exchange formats in the EA domain.

5.2 Integration problems

As a final question regarding the data sources we asked the participants to select the main problems they see with using external data sources for automating EA documentation. The results can be seen in Table 3. As may be expected from the previous discussions the data granularity mismatch between data source and the EA information model is seen as the biggest problem.

Table 3. Problems in EA information source integration (n=123)

Problem	Count	Percentage
Granularity	91	73.98%
Cost	74	60.16%
Data quality	55	44.72%
Low ROI	35	28.46%
Security	17	13.82%
Other	20	16.26%
Nothing	5	4.07%

The cost of implementing the integration is the second most mentioned problem. Low data quality at the source is regarded as a problem by almost half of the organizations and 35 generally see a low return on investment (ROI) for integrating external information sources. Surprisingly, only a small percentage sees security as an issue when integrating various information sources. Additionally mentioned problems are lack of management support, lack of standardization and problems assigning responsibility to stakeholders that should be responsible for automatically collected data.

6. Discussion and future research directions

Our survey showed that maintaining the EA model is one of the biggest problems in EA practice. However, only few organizations make use of automation to reduce the amount of manual data collection. A main finding of this survey is that all discussed information sources potentially contain EA-relevant information, or at least can be used as sources for change events.

To improve the process and method of EA documentation we see several directions of future research. These are:

(1) Research of methods to (automatically) raise the level of abstraction of collected structured information to be useful in an EA model.
(2) Research of methods to utilize events from information systems to trigger and drive EA data collection processes.
(3) Research of information demands based on the type of EA endeavor.
(4) Research towards a cost-benefit analysis model for the implementation of automated EA information collection.

In the following we discuss these potential areas of future research in more detail.

Referring to research direction (1), the results of this survey show that the largest problem for automated EA documentation is the abstraction gap between the source model and the EA model. However, the actuality appears to be better for more technical aspects such as can be seen in the results for the CMDB, ESB, and the Network Scanner tools. Nevertheless, for some of these tools, especially the Network Monitors and Scanners, a much larger abstraction gap can be expected. Hence, research approaches need to identify ways to bridge this abstraction gap, possibly with the help of manual tasks in the integration step as presented in [7]. Another approach that is worth mentioning is the adaption/extension of the EA meta-model to match the information source models. However, research has to find out in which cases this is appropriate.

With regard to research direction (2), the information sources that cover the upper EA-layers, such as the project or process layers, suffer a similar problem as the EA tools. Their data is mostly manually maintained. Hence, actuality, correctness and completeness heavily depend on the maintenance process in a given organization. However, these sources, such as project Portfolio Management tools, may be used as providers for events that trigger manual data collection processes and can thus also participate in the automation. The usage of events for documentation is an under researched topic in the EA field.

Generally speaking, organizations whose EA endeavor is more IT-centric, i.e. also used to supported systems operations, may benefit more from automation; since more information sources exist that provide relevant technical data, than sources providing business data. More business centric EA endeavors can, however, still benefit of collecting events from business oriented information sources. In line with findings of Aier et al. [1], the free text answers of the survey confirm that there is no general consensus on which technical level data needs to be collected in an EA tool. Based on the EA endeavor classification of Aier et al., the information demands of each type should be researched. This constitutes research direction (3).

As a result of different information demands and differences in available sources to fulfill these demands, we argue that each organization needs to weigh the cost of implementing automation against the cost of manual data collection. Since the tool support and standards for automation are limited, as our survey shows, the cost of implementing automation is still high. With regard to research direction (4) cost-benefit analysis methods have to be researched to be able to predict in which cases automation provides enough benefits to justify the cost of implementation.

7. Conclusion

The documentation of the EA is a challenging task for today's organizations since it is very time consuming. In this paper we detailed these challenges in practice with a survey and summarized related findings from literature. While initial efforts in research on automated EA documentation were identified, we provide a concrete list of possible EA information sources found in organizations. The appropriateness of these information sources is discussed with respect to their provided data types and data quality.

Our survey reveals that in current organizations automation of the EA documentation is hindered by the abstraction gap between the collected data and the EA model. The information is often too technical and cannot be directly used in the EA model.

In addition, we found that the percentage of the EA model that can be automatically collected mostly depends on whether the model is focused on technical or more on abstract business related information. The analysis of the provided data types and their quality showed that data that can be automatically collected is mainly located on lower technical layers in EA models. However, project and process related information sources might be used as providers for events that trigger manual data collection processes and can thus also participate in the automation.

EA practitioners can use our results to analyze their application landscape for automation opportunities. For researchers the results highlight the main problem in the domain that needs to be tackled: bridging the abstraction gap between existing information silos and EA tools.

As the next step in our future work we will further analyze the data set to research if the identified challenges mentioned by the organizations correlate with solution approaches by the participating organizations in order to measure their effectiveness.

10. References

[1] Aier, S., Gleichauf, B., and Winter, R. Understanding Enterprise Architecture Management Design - An Empirical Analysis. *Proceedings of the 10th International Conference on Wirtschaftsinformatik WI 2.011 (Zurich)*, (2011).

[2] Brückmann, Tobias and Gruhn, Volker and Pfeiffer, M. *Towards real-time monitoring and controlling of enterprise architectures using business software control centers*. Springer Berlin / Heidelberg, 2011.

[3] Buschle, M., Ekstedt, M., Grunow, S., Hauder, M., Matthes, F., and Roth, S. Automated Enterprise Architecture Documentation using an Enterprise Service Bus. *Americas conference on Information Systems*, (2012).

[4] Buschle, M., Holm, H., Sommestad, T., Ekstedt, M., and Shahzad, K. A Tool for automatic Enterprise Architecture modeling. *CAISE11 Forum*, (2011), 25-32.

[5] ter Doest, H. and Lankhorst, M. *Tool Support for Enterprise Architecture-A Vision*. 2004.

[6] Farwick, M., Agreiter, B., Breu, R., Ryll, S., Voges, K., and Hanschke, I. Requirements For Automated Enterprise Architecture Model Maintenance. *ICEIS 2011 - 13th International Conference on Enterprise Information Systems*, SciTePress (2011), 325-337.

[7] Farwick, M., Agreiter, B., Breu, R., Ryll, S., Voges, K., and Hanschke, I. Automation Processes for Enterprise Architecture Management. In *2011 IEEE 15th International Enterprise Distributed Object Computing Conference Workshops*. Ieee, 2011, 340-349.

[8] Fischer, R., Aier, S., and Winter, R. A Federated Approach to Enterprise Architecture Model Maintenance. *Enterprise Modelling and Information Systems Architectures 2*, 2 (2007), 14–22.

[9] Kaisler, S.H., Armour, F., and Valivullah, M. Enterprise Architecting: Critical Problems. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences 00*, C (2005), 224b-224b.

[10] Klosterboer, L. *Implementing ITIL Configuration Management*. IBM Press, 2008.

[11] Langenberg, K. Enterprise architecture: What aspects is current research targeting. *Laboratory of Systemic Modeling*, (2004).

[12] Lucke, C., München, U.D.B., Krell, S., and Lechner, U. Critical Issues in Enterprise Architecting – A Literature Review Critical Issues in Enterprise Architecting – A Literature Review. *AMCIS 2010 Proceedings*, (2010).

[13] Moser, C., Junginger, S., Br\"uckmann, M., and Schöne, K.M. Some process patterns for enterprise architecture management. *Proceedings, Workshop on Patterns in Enterprise Architecture Management (PEAM2009), Bonn*, (2009), 19-30.

[14] Ross, J.W., Weill, P., and Robertson, D. *Enterprise architecture as strategy: Creating a foundation for business execution*. Harvard Business Press, 2006.

[15] Ross, J.W. *Creating a Strategic IT Architecture Competency: Learning in Stages*. Boston, MA, USA, 2003.

[16] Saat, J., Winter, R., Franke, U., Lagerstroem, R., and Ekstedt, M. Analysis of it/business alignment situations as a precondition for the design and engineering of situated it/business alignment solutions. *Proc. Hawaii International Conference on System Sciences (HICSS-44).(January 2011) To appear*, (2010).

[17] Schekkerman, J. *The Economic Benefits of Enterprise Architecture*. Trafford Publishing, 2005.

[18] Shah, H. Frameworks for enterprise architecture. *It Professional*, October (2007), 36-41.

[19] The Open Group. *TOGAF Version 9.1*. Van Haren Publishing, 2011.

[20] Weill, P. and Ross, J.W. *IT Savvy: What Top Executives Must Know to Go from Pain to Gain*. Harvard Business Press, 2009.

[21] Winter, K., Buckl, S., Matthes, F., and Schweda, C.M. Investigating the state-of-the-art in enterprise architecture management method in literature and practice. *5th Mediterranean Conference on Information Systems MCIS2010*, (2010).