# Challenges for Automated Enterprise Architecture Documentation

Matheus Hauder, Florian Matthes, and Sascha Roth

Technische Universität München (TUM)
Chair for Informatics 19 (sebis)
Boltzmannstr. 3, 85748 Garching bei München, Germany
{matheus.hauder,matthes,roth}@tum.de

**Abstract.** Currently the documentation of an Enterprise Architecture (EA) is performed manually to a large extent. Due to the intrinsic complexity of today's organizations this task is challenging and often perceived as very time-consuming and error-prone. Recent efforts in research and industry seek to automate EA documentation by retrieving and maintaining relevant information from productive systems. In this paper major challenges for an automated EA documentation are presented based on 1) a practical example from a global acting enterprise of the German fashion industry, 2) a literature review, and 3) a survey among 123 EA practitioners. The identified challenges are synthesized to four categories and constitute the foundation for future research efforts and pose new questions not yet considered.

**Key words:** Enterprise Architecture (EA), automated EA documentation, challenges, literature survey, model transformation, practitioner survey

## 1 Motivation

Decision makers need to be supported with sound and up to date information about the EA [26]. This includes the organizational structure, processes, application systems, and technologies [15]. Existing EA documentation approaches struggle with the information volume and rapidly changing requirements within organizations. A study conducted by Winter et al. [27] reveals a high degree of manual work with very little automation during the documentation and maintenance of EA models. This high degree of manual work combined with the increasing information volume of organizations results in very time-consuming, error-prone and expensive maintenance of EA information. Next to meeting these information demands of organizations, the EA documentation also needs to achieve and sustain a high quality in the collected data [9].

Motivated by these problems recent research activities propose processes for automated EA documentation [8] and investigate possible information sources to retrieve relevant EA information from productive systems [6, 7]. These initial research efforts reveal a substantial amount of relevant EA information that can be gathered from productive systems and provide guidance for maintaining EA models using these information sources. While first steps towards an automated EA documentation in organizations were investigated, to the best of the authors' knowledge existing literature did not investigate major challenges for automated EA documentation. Literature regarding automated EA documentation is still very scarce, so that identifying current challenges for this field requires additional quantitative and qualitative analyses of current practices in organizations in order to receive a thorough list of relevant challenges. In this paper we illustrate model transformations to collect relevant EA information from three different information sources, conduct a survey among EA practitioners, and investigate current literature.

In the next section the applied research methodology to identify challenges for automated EA documentation is presented. In Section 3 a prototypical model transformation from an enterprise of the German fashion industry is provided to identify transformation challenges. The results from a survey among EA practitioners are shown in Section 4. Section 5 presents the identified challenges for automated EA documentation before the paper concludes with a summary.

## 2 Research Methodology

The research methodology to identify challenges for automated EA documentation is based on three different sources that are illustrated in Figure 1. The selection of these sources was performed to provide a thorough list of challenges. A prototypical model transformation for three potential EA information sources is provided to identify challenges regarding the transformation of the collected information into a central EA repository. Within a literature review major challenges for automated EA documentation are identified. Furthermore, a survey among EA practitioners is conducted including questions on EA documentation and automation in particular. In the following the individual parts of the approach are presented more detailed.

In previous work we have investigated an Enterprise Service Bus (ESB) from an enterprise of the German fashion industry as one particular information source [6]. With this information source entities of the ArchiMate meta-model could be covered with up to 50% on the infrastructure layer, 75% on the application layer and 20% on the organizational layer. In this paper we build on these findings and investigate further productive systems from this enterprise in order to integrate them into a central EA repository. We argue that considering several information sources is necessary in order to reach a high model coverage since these productive systems provide information on different layers of the EA. While an ESB can be used to retrieve information on the application level, a network monitor tool for instance might provide more technical information from the infrastructure layer. Therefore, an integration of several information sources is an essential challenge to achieve an automated EA documentation. Based on this prototypical implementation we identify challenges for the model transformation and integration of information sources.

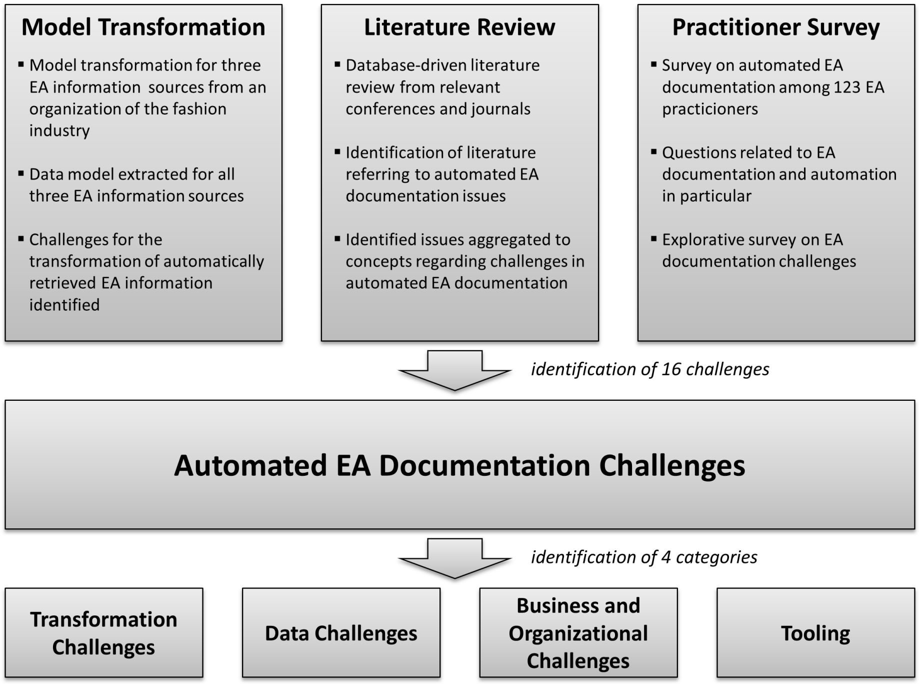


**Fig. 1.** Research methodology to identify automated EA documentation challenges

According to Glaser et al. [11] we performed a content analysis of relevant literature. This literature was identified with a database-driven review using the AIS Electronic Library and IEEE Xplore [25]. Regarding the relevance of this topic for organizations in practice the research efforts are still very scarce (cf. e.g. [1]). Nevertheless, we identified several publications dealing with different aspects of automated EA documentation. In previous work we have investigated a model transformation and data quality aspects of an ESB from an organization of the fashion industry [6, 12]. Another concrete implementation using a security network scanner can be found in [7]. A process and its requirements for automated EA documentation is presented in [8, 9]. Next to these recent publications on automated EA documentation we identified adjacent publications that deal with related research questions.

Furthermore, we conducted a global explorative survey to analyze the status-quo of EA documentation and investigate quality aspects of possible information sources within organization. Within this survey over 1100 invitations were sent by e-mail to EA experts for an online questionnaire. We received 123 answers in total with organizations from, e.g., Canada, Germany, Great Britain, India, New Zealand, South Africa, Switzerland, and USA. Among the participants were 68 Enterprise Architects (55.28%), 22 Enterprise Architecture Consultants (17.89%), as well as 8 Software Architects (6.50%). The Enterprise Architecture Consultants in this survey were asked to answer on behalf of one specific organization. Largest industry sectors of the participating organizations are Finance with 37 (30.08%), IT and Technology with 23 (18.70%), and Government with 11 (8.94%). Main goal of this survey is to answer research questions on the status-quo of EA documentation, relevant productive systems containing EA information, data quality attributes of these systems, and typical integration problems for the identified information sources. First findings of the survey show that documentation of EA information is a major challenge for organizations since it is regarded as very time consuming and the achieved data quality is not sufficient. Furthermore, some organizations have already implemented automation in their EA documentation processes. In this paper we summarize the questions as well as free text answers on automated EA documentation challenges organizations are currently faced with or are considered to be relevant for the future.

## 3 Model Transformation

In this section we exemplify the combination of three models, namely 1) Iteraplan that can be assigned to the business layer 2) SAP Process Integration (PI) which is an Enterprise Service Bus (ESB) as a representative for the application layer, and 3) Nagios, an infrastructure monitoring tool that gathers data from the technology layer. Presented models have been reverse engineered from the respective information source whereas semantics of the entities therein are inferred through exegesis of respective documentation [14, 22, 20].

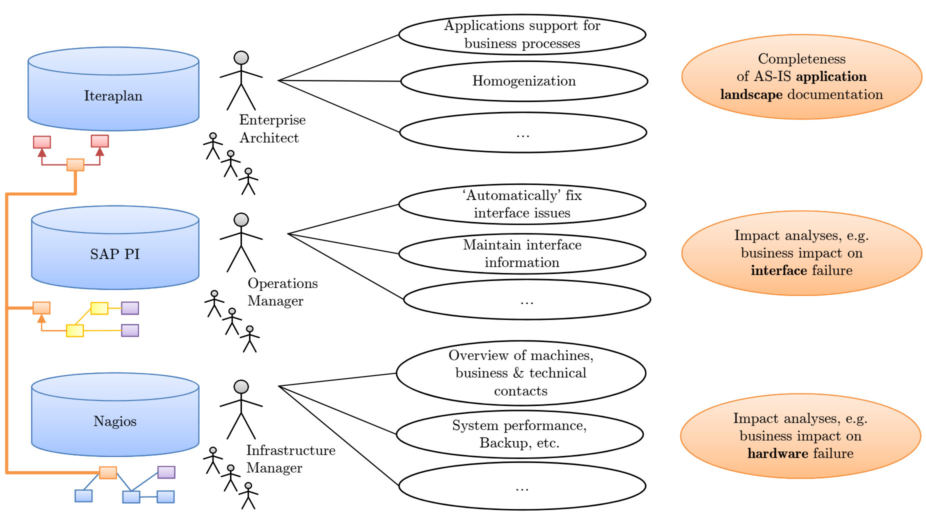


**Fig. 2.** Data model integration, use cases and sample concerns of stakeholders

In practice, semantic concepts of those systems strongly depend on concrete instance data. As reference to an existing and established standard, they are compared the ArchiMate 2.0 specification [24] where appropriate. Our industry partner's vision of an automated EA documentation, data should not only be imported to a common repository, but also (vertically) integrated. As shown in Figure 2, the systems we utilize as an illustrating example serve different stakeholders and, thus, are especially suited for respective use cases. However, when integrating these information sources vertically, i.e. by connecting these information silos, impact analyses from top-down (e.g. 'Which parts of my infrastructure is business critical?') and bottom-up (e.g. 'Which business process are influenced by server downtimes?') are facilitated and each individual stakeholder gets a more holistic view (for viewpoints see e.g. [2, 5, 3, 23, 13]). Moreover, connecting an EA tool with operative systems also can be utilized to double-check manually collected data, i.e. facilitate data correctness, completeness, or detect white-spots.

### 3.1 Iteraplan

At the Business Layer Iteraplan covers concepts like *Business Domains* that group *Business Processes*, *Business Functions*, *Business Objects*, *Business Units* and, via the *Business Mapping*, also *Products*. *Business Processes* of Iteraplan "have a Name and a Description, and may also have Attributes [...]. You can also specify one or more subordinate Business Processes and the sequence of these subordinate processes" [14]. In contrast, ArchiMate defines it as "[...] a behavior element that groups behavior based on an ordering of activities. It is intended to produce a defined set of products or business services" [24]. In addition, we found that a *Business Function* of Iteraplan has a respective entity in the ArchiMate specification, namely *Business Function*. Thereby, the former is documented as "Business Functions have a Name and a Description, and may also have attributes[...]" [14] whereas the later separates the meaning of Business Functions and Business Processes and describes a *Business Function* as "[...] a behavior element that groups behavior based on a chosen set of criteria (typically required business resources and/or competences) [...and] while a business process groups behavior is based on a sequence or 'flow' of activities that is needed to realize a product or service, a business function typically groups behavior based on required business resources, skills, competences, knowledge, etc." [24].

*Business Objects* of Iteraplan are also defined as an entity with name and description whereas the ArchiMate specification defines a *Business Object* "as a passive element that has relevance from a business perspective" [24]. Moreover, the ArchiMate specification details Business Objects "represent the important 'informational' or 'conceptual' elements in which the business thinks about a domain. [...] Business objects are passive in the sense that they do not trigger or perform processes" [24]. *Product* refers to the ArchiMate concept *Product*, where a "[...] product is defined as a coherent collection of services, accompanied by a contract/set of agreements, which is offered as a whole to (internal or external) customers" [24]. In Iteraplan, the entity *Product* does not cover contracts or agreements, but may include several *Business Functions* for this *Product* in a *Business Domain*.

At the Application Layer, detailed in Figure 3, Iteraplan contains data about *Information Systems* whereby “most work with Information Systems is done by creating or modifying their releases” [14]. Thereby, each Information System Release is a version of a particular Information System. Besides name and description, an *Information System Release* has two timestamps to indicate the period in which a release is productive.

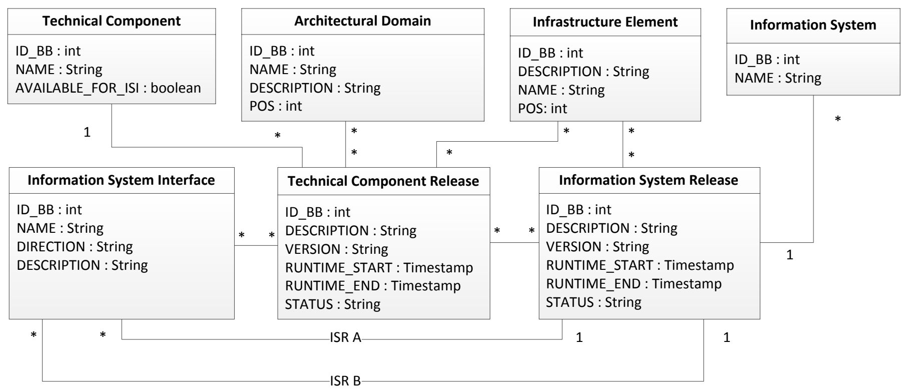


**Fig. 3.** Simplified excerpt of the Iteraplan data model at the application layer

*Information System* in the sense of Iteraplan fall close to the Application Component of ArchiMate. “An application component is defined as a modular, deployable, and replaceable part of a software system that encapsulates its behavior and data and exposes these through a set of interfaces” [24]. Iteraplan also contains information about *Information System Interfaces* which can be directly mapped to *Application Interface* of ArchiMate “defined as a point of access where an application service is made available to a user or another application component” [24]. In Iteraplan, an *Information System Interface* “has moreover relationships with the Business Objects it is transporting, and with the Technical Components on which it is based” [14].

At the Infrastructure Layer Iteraplan uses a *Technical Component* to describe for instance programming languages or frameworks, databases, or application servers use by an *Information System*. A *Technical Component* can be compared with a *Node* of ArchiMate “[...] defined as a computational resource upon which artifacts may be stored or deployed for execution” [24]. Such a Node can be a device, system software, or even a network element. Thereby a device is “a hardware resource upon which artifacts may be stored or deployed for execution” [24]. In this vein, Iteraplan also uses *Infrastructure Elements* that “describe the operating platform (servers etc.) on which the Information System Release is running” [14].

The Iteraplan documentation describes the remaining entities from a technical view, i.e. they all have a name and a description field and may or may not be hierarchically organized. As a consequence, it strongly depends on a concrete instance of Iteraplan whether its data refers to above outlined concepts. Iteraplan uses *Attributes* and *Attributes Values* to extend concepts by means of key value pairs.

### 3.2 SAP PI

Figure 4 details the data model of SAP PI utilized as information source to map a tool entirely used as knowledge management tool, namely Iteraplan, to the real world, i.e. operative IT.

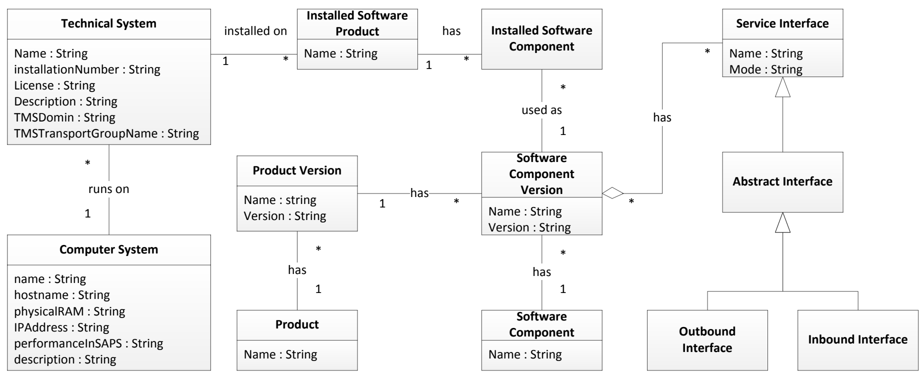


**Fig. 4.** Simplified excerpt of the SAP PI data model

At the Business Layer, SAP PI only implicitly contains relevant information. Considering the entire SAP PI data model (see Appendix), information about underlying, pursued goals is completely absent. Even though business objects, i.e. "a unit of information relevant from a business perspective" [24], are not directly included in SAP PI, data types may indicate their existence (cf. SAP PI best practice data naming conventions [21]). As a consequence, it strongly depends on a concrete instance [6, 12] whether the SAP PI system contains information about business objects.

Central to the Application Layer is the *Application Component* specified as a "[...] modular, deployable, and replaceable part of a system [...]" [24]. While SAP PI introduces two similar concepts, software components and software products whereas the former are not deployable. Products involved in message exchange processes form an application collaboration. Access to the underlying services provided by application components as well as their groupings is modeled by application interfaces, semantically equivalent to SAP PIs enterprise service interface. Which application component invokes which interface is implicitly included in SAP PIs routing information (receiver determination and interface determination) defining the message exchange between enterprise service interfaces and software products. While internal functionality of application components remains invisible to SAP PI, first indications on external visible functionality (application services) exist. Behavioral information in SAP PI is available rather indirectly in interfaces and the included descriptions of operations. However, even though the operations contain all service information, it is questionable whether the operations can be automatically aggregated to specify the service forming the combined functionality.

At the Technology Layer, SAP PI comprises information about the underlying infrastructure. This begins with ArchiMate's *node*, modeling a computational resource which corresponds to SAP PI's computer system. As the provided and needed interfaces of infrastructure components are not essential for the coordination of applications, none of this information appears in SAP PI's data. While the underlying physical mediums are abstracted in SAP PI each invocation of a service comprises two communication paths, one between the service client and SAP PI and the other between SAP PI and the service provider. In ArchiMate, system software ("software environment for specific types of components and objects" [24]) belongs to the behavioral concepts. In SAP PI, a subset of installed system software is registered at the System Landscape Directory including the following elements: operating systems, database systems, and technical systems. In ArchiMate, artifact, the sole informational element, represents "a physical piece of information" [24]. With the exception of files imported by the SAP PI components such as WSDL files for specifying interfaces, information about existing artifacts, especially artifacts application components are realized by, is not available.

### 3.3 Nagios

A simplified version of the data model for Nagios is shown in 5. As an infrastructure monitoring tool, Nagios does not contain any data referring to Business or Application Layer. Nagios is able to actively and passively monitor infrastructure elements and thus contains manifold information about hosts, services, and network elements. At the Infrastructure Layer Nagios uses the *Downtime History* to manually store (planned) downtimes of hosts or services. If downtimes are defined assigned hosts and services are not checked anymore and no notifications are sent to the contact person during defined periods, because the downtime is scheduled.

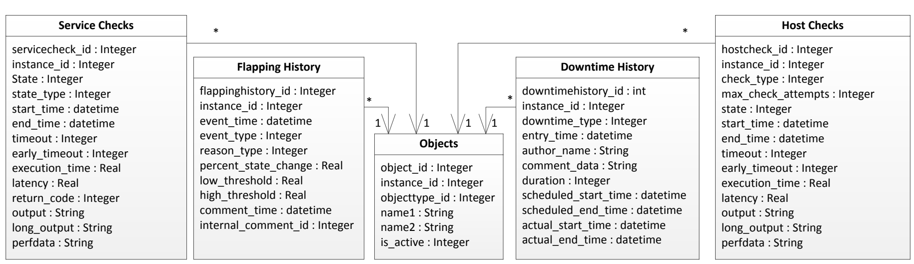


**Fig. 5.** Simplified excerpt of the Nagios data model

Nagios uses a client/server architecture and saves responses of acknowledgment requests in the *Acknowledgments* class. For hosts, these acknowledgments are up, down, or unreachable whereas for services, they can be either ok, warning, critical, or unknown. This data is stored together with a timestamp. The *Flappinghistory* stores the flapping data of services or hosts, i.e. it saves the event when one state of a service or host is changed. Nagios saves the periodical checks of hosts in *Host Checks* whereas periodical checks of services are stored as *Service Checks* whereby the state (up, down, or unreachable) is also captured. Via the *start_time* and *end_time* attribute, the period of a certain state can be calculated.

Since ArchiMate does not include such fine grained information, a mapping may embrace nodes or devices (cf. above). However, monitoring tools can be employed to map the 'real world' to an EA model to facilitate its completeness and correctness. Mapping such fine grained information to an EA model refers to the challenge of Data Granularity detailed below.

### 3.4 Vertical Model Integration

Figure 6 shows an example for a model mapping of the three above introduced data models. As illustrated, transformation rule $\varphi$ is required to perform a semantically and syntactically correct mapping.

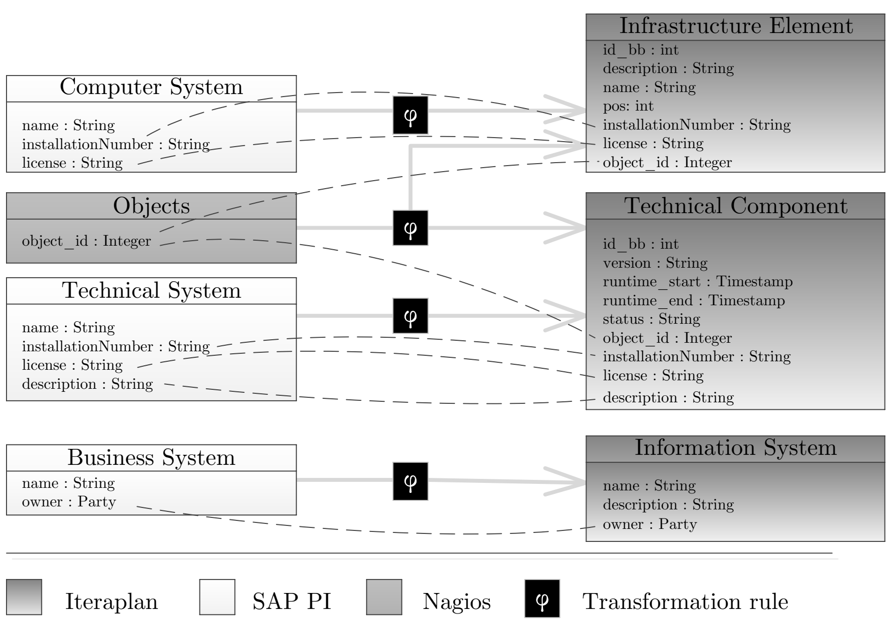


**Fig. 6.** Data model mapping of SAP PI and Nagios to Iteraplan

As illustrated, SAP PI's *Computer Systems* can be mapped to *Infrastructure Elements* of Iteraplan. In our vertical integration scenario, Iteraplan's Infrastructure Elements gets licensing information and installation numbers. In this vein, the Nagios model element *Objects* contains an enumeration of `objecttype_id` which can either indicate that the object is a host or an infrastructure service. In this case, a mapping table has to be provided that maps *Objects* in Nagios to *Infrastructure Elements* or *Technical Components* of Iteraplan by applying a filter (`objecttype_id == 1`) ensuring only hosts are mapped. Another filter (`objecttype_id == 0`) can be used to identify services which subsequently could be mapped to *Technical Components*. Thereby, $\varphi$ first has to search for a matching object (e.g. similar or even equal hostname or IP address). If a matching object has been found, $\varphi$ has to align data if a source attribute already corresponds to a target attribute, e.g. validating or invalidating data.

Otherwise, $\varphi$ could add/fill non-existing attributes or append values. Again, attributes contained in the source model (Nagios) are transfered to the target model (Iteraplan) to enable vertical integration. Thereby, fields contained in both models need to be synchronized, e.g. the field description. *Technical Systems* of SAP PI can be mapped by $\varphi$ to *Technical Components* of Iteraplan. In this vein, naming conventions [21] are essential since $\varphi$ needs an identifier for each *Technical System* or *Technical Component* and a respective mapping table. Commonly, IP addresses for instance are not maintained in an EA tool, possibly in an ESB, but definitely in an infrastructure monitoring tool or Configuration Management Database (CMDB). This becomes more critical when harmonizing or vertically integrating three different data sources. Finally, SAP PI's *Business Systems* can be directly mapped to Iteraplan's *Information Systems*. Thereby, $\varphi$ has to find the relevant *Information Systems* in Iteraplan first, or insert new data, if the system does not exist.

## 4 Survey Results

Next to the exemplified model transformation from an enterprise of the German fashion industry, a survey among 123 EA practitioners was conducted in order to identify challenges for automated EA documentation. For this purpose we asked the organizations what their current challenges in this context are using a predefined set of challenges. In addition, we asked the organizations to provide challenges not covered in our selection by using a free text field in the survey. About 20 organizations utilized this option and provided information on further challenges.

These challenges hindering automated updates in the documentation of the EA are summarized in Table 1, whereas only a very small minority of 5 organizations (4.07%) mentioned that they have no specific challenge in their organization. A total number of 123 organizations answered this question with at least one of the provided answers. 91 (73.98%) of the organizations have mentioned the abstraction gap between the EA and the information source as challenge. This

| Answer | Count | Percentage |
|---|---|---|
| Abstraction gap to EA model | 91 | 73.98% |
| Cost of integration for EA tool | 74 | 60.16% |
| Low data quality at information sources | 55 | 44.72% |
| Low return of investment | 35 | 28.46% |
| Security when using network scanners | 17 | 13.82% |
| Other | 20 | 16.26% |
| Nothing specific | 5 | 4.07% |

**Table 1.** *What hinders updates in the context of EA?*

challenge has also been identified within the model transformation from Section 3. The cost of integration for the EA tool was stated by 74 (60.16%) of the organizations indicating a missing support of existing solutions. Almost half of the organizations (44.72%) also highlighted low data quality at the information sources as challenging. Since organizations typically have multiple information sources containing relevant EA information, further research is necessary to identify possible information sources and their data quality attributes. Automating the EA documentation requires large initial investments in the organizations due to the missing tool support. In our survey 35 (28.46%) organizations stated a low return of investment as an obstacle. Around 17 (13.82%) organizations foresee security when using network scanners as challenging. Usually these tools require administrative rights since that have to be executed on the machines to monitor.

| Answer | Count | Percentage |
|---|---|---|
| Yes | 25 | 20.33% |
| No | 24 | 19.51% |
| Not yet considered | 31 | 25.20% |

**Table 2.** *Do you plan to use automated EA model updates in the future?*

| Answer | Count | Percentage |
|---|---|---|
| Too difficult | 11 | 8.94% |
| Too expensive | 9 | 7.32% |
| Not enough ROI | 9 | 7.32% |
| Not enough tool support | 8 | 6.50% |
| Other | 6 | 4.88% |

**Table 3.** *Why do you not plan to use automation?*

One organization stated within the free text field the definition of roles and responsibilities for the collected data as challenging. Since a complete automation of the EA documentation is probably not achievable, manual activities will be necessary in future. Therefore, appropriate roles and responsibilities are necessary to coordinate the data collection and ensure a high data quality of the imported information. Another organization mentioned the effort to transform data from the information sources to the EA repository as challenging. Similarly, the lack of standardization and the inclusion of all appropriate information sources are related to this challenge. Many of these issues already have been investigated in the exemplified model transformation in Section 3. Several organizations also mentioned the general acceptance and a low degree of upper management awareness of EA management as challenge. This is a critical challenge although it is not directly related to automated EA documentation.

The organizations were also asked if they plan to use automated EA model updates in the future. The results are shown in Table 2 with 25 (20.33%) organizations planing to use automation. At the same time 31 (25.30%) organizations have not yet considered to use automated EA model updates. Therefore, almost half of the organizations might apply techniques for automated EA documentation in the future. The 24 (19.51%) organizations not planing to use automation were also asked give a reason for this decision that are shown in Table 3. 11 (8.94%) organizations envision automation as too expensive to implement in their organization, while 9 (7.32%) organizations think it is too difficult to acquire and it provides not enough return on investment. Around 8 (6.50%) of the organizations mentioned that enough tool support for automated EA documentation is available.

## 5 Challenges for Automated Enterprise Architecture Documentation

In this Section challenges from the above presented model transformation from an enterprise of the German fashion industry as well as the practitioner survey are identified and grouped into three high-level categories. In addition, a literature review was performed to identify new challenges and align them with the findings from this paper. An overview of all categorized challenges found in this paper is shown in Table 4 containing a reference to the identified sources for every challenge.

### 5.1 Data Challenges

The collection of appropriate data is the foundation to enable automated EA documentation in organizations. Data challenges result from collecting data utilizing productive systems that contain relevant EA information within organizations. Main reasons for these challenges are the multitude of possible productive systems in organizations and the quality as well as actuality of the retrieved information.

*DC 1 - Overload of productive systems due to large volume of transactions for automated data collection.* Productive systems may be influenced in the daily operation when the entire data store or parts thereof are collected during the data collection step. As a result these these outlined mechanisms could lead to unexpected peak loads in the productive systems. This can be quite a challenge due to causal relationships in the infrastructure of the organization, especially if the productive system used as information source is essential for the business. In our illustrating example (cf. Section 3), the ESB can be considered as the nervous system of an enterprise interconnecting business applications and processes [6].

*DC 2 - Selection of the right productive systems as information sources for EA documentation.* Automated EA documentation requires the integration of several information sources in the organization. The selection of the information sources need to assessed according to several categories as already identified by Farwick et al. [8]. For this purpose the selection has to consider the content of the information sources with respect to the relevance for EA. Another issue in this context is the necessary effort to build an interface for exporting the data from the information source. In many cases the productive systems have no interface provided and their meta model needs to be reverse engineered in an additional step [6]. Further examples are the data quality attributes and the level of security that can be achieved for the exchange of the data.

*DC 3 - Detection of changes in the real world EA and their propagation to the EA model in the repository.* Automated EA documentation consists of two major steps, which are the documentation of the existing EA as well as the maintenance of an appropriate data actuality of the repository. Maintaining the EA repository requires an automatic detection of changes in the real world EA from different sources. This includes for instance the detection of new information systems, infrastructure elements, projects, as well as changes of these elements as highlighted by Farwick et al. [9]. Furthermore solutions are necessary to propagate these changes to the EA repository.

*DC 4 - Data quality in the productive systems not sufficient for the documentation of EA information.* Automated EA documentation requires sufficient data quality at the productive systems. However, 55 (44.72%) of the participating organizations in the survey envision the data quality of the information sources as too low for EA documentation. At this point further research is necessary to evaluate possible information sources in organizations and their quality attributes. Possible quality attributes for information sources that need to be investigated are for instance actuality, completeness, correctness, and granularity. Next to this evaluation of possible information sources in organizations, quality assurance mechanisms are necessary to ensure the data quality using manual checks.

### 5.2 Transformation Challenges

Once the data could be collected from information sources of the organization a transformation step needs to align this information with the target model of the EA repository. To achieve this goal the transformation has to deal with several challenges resulting from different models between information source and the target repository.

*TC 1 - Model transformation for the exchange of EA information necessary due to missing interfaces and standards.* As exemplified above, customized model transformations are necessary to map the different information sources to a target model. Major reason for these individual model transformations are missing standards or non-conformance to standards of enterprises. Conformance to standards, e.g. to ArchiMate 2.0 [24], could simplify such a mapping for instance when a semantically and syntactically correct mapping is required. The transformation rules become more complex when adding new information sources. In our case, additional mapping tables where provided. Thereby, a mapping function commonly first has to search for a matching object and, if found, align data only if a source attribute already corresponds to a target attribute, e.g. validating or invalidating data. Otherwise, strategies like adding new attributes, filling unset attributes, or appending values have to be chosen individually for each transformation rule.

*TC 2 - Ambiguous concepts imported from the productive systems in the organization require a consolidation.* Our examples already indicate that rigorous data migration mechanisms like table merging and data cleansing (see [17]) must be provided for such a vertical integration. We conclude that frequent model changes of the target model are necessary when adding new information sources. Thus, a non-rigid typed model could be beneficial to some extent. Moser et al. [18] address the challenge of inaccurate data in the EA repository. Data gathered from different information sources tends to be inhomogeneous [17], i.e. different data formats or simply different lengths of fixed-character fields. Regardless data is entered manually or automatically via import mechanisms, data has to be consolidated. For instance, synonyms and homonyms have to be cleared. In the worst case, ambiguous concepts are imported and have to be cleaned afterwards. For manual data collection, Fischer et al. propose data quality contracts [10] between different parties. However, for automation, this remains a challenge after all and data migration mechanisms like a staging area (see [17]) might be necessary.

*TC 3 - Administration of collected data from the productive systems is required to ensure actuality and consistency.* A meta-model as mentioned by [8] is necessary to automatically trigger activities to increase the quality of the collected information. Such a meta-model needs to consider attributes for expiry time of imported data elements, the date of last change, data responsibilities, data sources, etc.

*TC 4 - Duplicate EA elements imported from different productive systems of the organization.* Once imported in an EA repository, data can be analyzed. During the analysis process, the actual source of an information piece could matter [8, 9], e.g. if information is wrong or bad data quality is detected. Identity reconciliation also is a necessity to synchronize changes in the EA repository with the original source.

*TC 5 - Abstraction between the EA model and the imported information from productive systems of the organization.* Major challenge for organizations is the abstraction gap between the EA model and the provided elements from the information sources. 91 (73.98%) organizations rated this as the most important challenge for automation. This confirms our findings from Section 3. If our industry partner did not chose an integrative approach, elements imported from

| ID | Challenge | Source |
|---|---|---|
| | Data Challenges | |
| DC 1 | Overload of productive systems due to large volume of transactions for automated data collection. | Model Transformation |
| DC 2 | Selection of the right productive systems as information sources for EA documentation. | [8, 6] |
| DC 3 | Detection of changes in the real world EA and their propagation to the EA model in the repository. | [9] |
| DC 4 | Data quality in the productive systems not sufficient for the documentation of EA information.. | Survey |
| | Transformation Challenges | |
| TC 1 | Model transformation for the exchange of EA information necessary due to missing interfaces and standards. | Model Transformation, [18] |
| TC 2 | Ambiguous concepts imported from the productive systems in the organization require a consolidation. | [10, 17, 18], Model Transformation |
| TC 3 | Administration of collected data from the productive systems is required to ensure actuality and consistency. | [8] |
| TC 4 | Duplicate EA elements imported from different productive systems of the organization. | [8, 9] |
| TC 5 | Abstraction between the EA model and the imported information from productive systems of the organization. | [9], Survey, Model Transformation |
| | Business and Organizational Challenges | |
| BC 1 | Security vulnerability through monitoring tools in the infrastructure of the organization. | [7], Survey |
| BC 2 | Not enough return on investment due to large initial investment efforts. | Survey |
| BC 3 | Involvement of data owners for the maintenance of imported EA information. | Survey, [18] |
| | Tooling | |
| T 1 | Synchronization of changes in the EA model to the underlying productive systems. | [4, 19], Model Transformation |
| T 2 | Collection of information not relevant or too fine-grained for decision makers in the EA. | [8] |
| T 3 | Analyses have to be decoupled from the meta-model. | [13, 16], Survey |
| T 4 | Not enough tool support for automated EA documentation available. | Survey, Model Transformation |

**Table 4.** Categorization of automated EA documentation challenges

productive systems (SAP PI and Nagios) are too fine-grained for mere EA purposes. To overcome the abstraction gap, EA documentation may be facilitated by human tasks in a semi-automated manner.

### 5.3 Business and Organizational Challenges

Next to rather technical issues resulting from the information extraction and transformation of this information, there are further challenges regarding the added business value of automation as well as the organization. Since it requires large initial investments in organizations and proven solutions are missing in industry, automation is not feasible in some situations.

*BC 1 - Security vulnerability through monitoring tools in the infrastructure of the organization.* Network scanners provide information about the network architecture of an organization regarding all devices that are communication over TCP or UDP. This includes computers, firewalls, printers, and application information [7]. In the survey conducted in this paper 17 (13.82%) organizations foresee security as a critical challenge when using these network scanners for EA documentation. Since applications are actively observed within the machines, these tools usually need to be executed directly on the observing infrastructure with privileged access rights. As a result, tools monitoring infrastructure information about the EA pose security vulnerabilities for an organization.

*BC 2 - Not enough return on investment due to large initial investment efforts.* The initial effort to develop interfaces for the considered information sources and the cost for adapting existing EA tools to support automated documentation is regarded as very high. As a result 35 (28.46%) of all organizations that participated in the survey mentioned concerns about a low return of investment for automated EA documentation. Among the 24 (19.51%) organizations not planning to use automated EA model updates this issue was also raised as one of the main reasons. 9 (7.32%) do not plan to use automation since it is too expensive and does not guarantee any ROI.

*BC 3 - Involvement of data owners for the maintenance of imported EA information.* Within the survey another organization stated the definition of roles and responsible persons for the collected data as challenging. Defined roles are necessary to coordinate the maintenance of EA models on a coarse-grained level. Further responsibilities on the detail level of single applications, infrastructure elements, and processes are required to maintain the EA model information. These responsibilities are necessary since a complete automation of the EA documentation is not possible and manual quality assurance is necessary.

### 5.4 Tooling

Automated EA documentation is only feasible with the appropriate support of tool vendors. However, available tools are not capable to support importing, editing, and validating model data for automated EA documentation [16]. Existing solutions only support simple import mechanisms that are mainly limited to Excel or CSV files.

*T 1 - Synchronization of changes in the EA model to the underlying productive systems.* The exemplified model transformation presented in Section 3 processes data from the information sources to a target EA model by mapping the concepts and attributes. The managed evolution of an EA requires architects to adapt certain parts of the model, e.g., remove unused interfaces [4, 19]. These changes in the EA model need to be synchronized with the underlying information sources that were used to import the information automatically. Ideally, this information could be directly applied to a CMDB for instance to avoid multiple updates in several applications that might create inconsistencies.

*T 2 - Collection of information not relevant or too fine-grained for decision makers in the EA.* One of the main goals of automated EA documentation is to provide as many as possible concepts of the EA model by gathering the information from productive systems to avoid time consuming and error prone manual data collection. At the same time the EA model needs to omit information that are too fine-grained for decision makers in order to keep the model as lean as possible. Therefore mechanisms are necessary to tailor the target EA model and define concepts that should not be automatically imported [8]. A tool for automated EA documentation needs to support these requirements sufficiently.

*T3 - Analyses have to be decoupled from the meta-model.* An automated EA documentation endeavor is an ongoing process and, thus, it is not very likely to be realized with a big-bang strategy. Consequently, it is very likely that the meta-model of the target model (EA repository) has to be extended over time. Current EA tools [16] offer analysis mechanisms to analyze the EA meta-model with respect to some extension mechanisms. Thus, it might happen these analyses have to be altered when the EA model changes. As discussed in [13] by Hauder et al. analyses of a frequently changing meta-model is a challenging task. We conclude that analyses cannot be directly bound to the meta-model but the subject to be analyzed (models) must be interchangeable.

*T4 - Not enough tool support for automated EA documentation available.* A majority of 74 (60.16%) organizations stated that the necessary tool integration is very expensive to extend existing tools for EA management. Among the 24 (19.51%) organizations not planing to use automated EA model updates in future around 8 (6.50%) organizations mentioned not enough tool support as a reason for this. Due to the lack of available solutions for automated EA documentation, existing tools require a customized adaption to import EA information from productive systems. In our example, we implemented the model transformations individually.

## 6 Conclusion

In this paper, we have identified challenges for automated EA documentation. Therefore we have investigated model transformations from three information sources, presented our findings from a survey among 123 EA practitioners, and combined it with a literature study. Major challenges identified in this paper are synthesized and grouped along the categories *data*, *transformation*, *business and organization* as well as *tooling*. Within data challenges main aspects are the data quality and the selection of appropriate information sources. Transformation challenges deal with the mapping from different information sources to a central repository and the maintenance of this repository. Business and organization challenges address the added value of automation and the impact on the organizational structure. As the last category, tooling contains challenges for tool aided realization of automated EA documentation and the integration with existing EA repositories. The present paper is the first contribution elaborating challenges for automated EA documentation. These challenges constitute the foundation for future research efforts dealing with the applicability and effectiveness of automated EA documentation in organizations. We intent to discuss identified challenges as well as solutions at TEAR and to critically reflect automated EA documentation when put into practice with the audience.

## References


1. M. Berneaud, S. Buckl, A. Diaz-Fuentes, F. Matthes, I. Monahov, A. Nowobliska, S. Roth, C. M. Schweda, U. Weber, and M. Zeiner. Trends for enterprise architecture management tools survey. Technical report, Technische Universität München, 2012.
2. S. Buckl, A. M. Ernst, J. Lankes, and F. Matthes. Enterprise Architecture Management Pattern Catalog (Version 1.0). Technical report, Chair for Informatics 19 (sebis), Technische Universität München, Munich, Germany, 2008.
3. S. Buckl, F. Matthes, I. Monahov, S. Roth, C. Schulz, and C. M. Schweda. Enterprise architecture management patterns for enterprise-wide access views on business objects. In *European Conference on Pattern Languages of Programs (EuroPLoP) 2011*, Irsee Monastery, Bavaria, Germany, 2011.
4. S. Buckl, F. Matthes, S. Roth, C. Schulz, and C. M. Schweda. A conceptual framework for enterprise architecture design. In *Workshop Trends in Enterprise Architecture Research (TEAR)*, Delft, The Netherlands, 2010.
5. S. Buckl, F. Matthes, S. Roth, C. Schulz, and C. M. Schweda. A method for constructing enterprise-wide access views on business objects. In *Informatik 2010: IT-Governance in verteilten Systemen (GVS 2010)*, Leipzig, Germany, 2010.
6. M. Buschle, M. Ekstedt, S. Grunow, M. Hauder, F. Matthes, and S. Roth. *Automated Enterprise Architecture Documentation using an Enterprise Service Bus*. 2012.
7. M. Buschle, H. Holm, T. Sommestad, M. Ekstedt, and K. Shahzad. *A Tool for automatic Enterprise Architecture modeling*, pages 25–32. 2011.
8. M. Farwick, B. Agreiter, R. Breu, S. Ryll, K. Voges, and I. Hanschke. Automation processes for enterprise architecture management. In *Proceedings of the 2011 IEEE 15th International Enterprise Distributed Object Computing Conference Workshops*, EDOCW '11, pages 340–349, Washington, DC, USA, 2011. IEEE Computer Society.

9. M. Farwick, B. Agreiter, R. Breu, S. Ryll, K. Voges, and I. Hanschke. Requirements for automated enterprise architecture model maintenance - a requirements analysis based on a literature review and an exploratory survey. In *ICEIS (4)*, pages 325–337, 2011.
10. R. Fischer, S. Aier, and R. Winter. A federated approach to enterprise architecture model maintenance. In M. Reichert, S. Strecker, and K. Turowski, editors, *EMISA*, volume P-119 of *LNI*, pages 9–22. GI, 2007.
11. B. G. Glaser and A. L. Strauss. *The Discovery of Grounded Theory*, volume 20. Aldine, 1967.
12. S. Grunow, F. Matthes, and S. Roth. Towards automated enterprise architecture documentation: Data quality aspects of SAP PI. In *16th East-European Conference on Advances in Databases and Information Systems*, 2012.
13. M. Hauder, F. Matthes, S. Roth, and C. Schulz. Generating dynamic cross-organizational process visualizations through abstract view model pattern matching. In *Architecture Modeling for Future Internet enabled Enterprise (AMFInE)*, Valencia, Spain, 2012.
14. iteratec GmbH. iteraplan documentation home, 2012. online: `http://www.iteraplan.de/wiki/display/iteraplan/iteraplan+Documentation+Home`.
15. M. Lankhorst. *Enterprise Architecture at Work: Modelling, Communication and Analysis*. Springer, Berlin, 2. edition, 2009.
16. F. Matthes, S. Buckl, J. Leitel, and C. M. Schweda. *Enterprise Architecture Management Tool Survey 2008*. 2008.
17. F. Matthes and C. Schulz. Towards an integrated data migration process model. Technical report, Technische Universität München, München, Germany, 2012.
18. C. Moser, S. Junginger, M. Brückmann, and K. Schöne. Some process patterns for enterprise architecture management. In *Proceedings, Workshop on Patterns in Enterprise Architecture Management (PEAM2009), Bonn*, pages 19–30, 2009.
19. S. Murer, B. Bonati, F. J. Furrer, S. Murer, B. Bonati, and F. J. Furrer. *Managed Evolution*. Springer Berlin Heidelberg, 2011.
20. Nagios Enterprises. Nagios documentation, 2012. online: `http://www.nagios.org/documentation`.
21. SAP AG. PI Best Practices Naming Conventions, 2012. online: `http://www.sdn.sap.com/irj/scn/go/portal/prtroot/docs/library/uuid/40a66d0e-fe5e-2c10-8a85-e418b59ab36a`.
22. SAP AG. SAP help portal, 2012. online: `http://help.sap.com`.
23. M. Schaub, F. Matthes, and S. Roth. Towards a conceptual framework for interactive enterprise architecture management visualizations. In *Modellierung*, Bamberg, Germany, 2012.
24. The Open Group. *Archimate 2.0 Specification*. Van Haren Publishing, 2012.
25. J. Vom Brocke, A. Simons, B. Niehaves, K. Riemer, R. Plattfaut, and A. Cleven. *Reconstructing the giant: on the importance of rigour in documenting the literature search process*, pages 1–13. Hampton Press, 2009.
26. P. Weill and J. Ross. *IT Savvy: What Top Executives Must Know to Go from Pain to Gain*. Harvard Business Press, 2009.
27. K. Winter, S. Buckl, F. Matthes, and C. M. Schweda. Investigating the state-of-the-art in enterprise architecture management method in literature and practice. In *5th Mediterranean Conference on Information Systems MCIS2010*, 2010.