

Towards Automation of Enterprise Architecture Model Maintenance

Matthias Farwick* (Supervisor: Ruth Breu)

University of Innsbruck,
Institute of Computer Science
`matthias.farwick@uibk.ac.at`
`http://www.q-e.at`

Abstract. Enterprise Architecture Management (EAM) is a common practice in organizations that need to have a model of how their business relates to the supporting IT-landscape in order to make informed decisions to enhance the enterprise architecture (EA). Creating and maintaining such an EA model is an expensive and time consuming but crucial task in today's organizations. This has been recognised by both researchers and practitioners. However, only little research literature and practical approaches can be identified that target the automation of the EA model maintenance and to reduce the manual work. In this thesis we elaborate means to increase the data quality attributes of actuality and consistency of EA models via semi-automated data collection processes from external data sources and events. It is our main goal to better synchronize EA models with what they represent in the real world and thus reduce the manual labor of model maintenance.

Keywords: enterprise architecture management, maintenance, automation, meta-model, EAM, modelling

1 Introduction

Enterprise Architecture Management (EAM) is a practice used in mid-sized to large organizations that aims at modelling the relationships between business, its supporting information systems and the underlying IT infrastructure. This effort is done to be able to make informed decisions to better align business and IT, enable strategic planning of IT-changes, assess risks, and check compliance with legal regulations.

Several EAM frameworks are applied in practice today, such as TOGAF [1] and the Zachmann framework [2]. These frameworks prescribe enterprise architecture (EA) meta-models, processes, best practices and EA principles. It is common that EA specific applications are used in order to collect the EA data and build a model of the current state of the EA as well as to elaborate the

* This thesis is partially supported by the Austrian Federal Ministry of Economy as part of the Laura-Bassi – Living Models for Open Systems – project FFG 822740/QE LaB and iteratec GmbH, Munich.

planned future state of the EA. This data collection is often a very time consuming task since the relevant EA data is distributed among different departments and is mostly collected via interviews or surveys with stakeholders [3]. Due to this time consuming nature of EA data collection, frequent changes in the EA and the immense size as well as complexity of EA models, it is a difficult but crucial task to keep these models up-to-date with the reality.

This problem has been recognized by both researchers [4–7] and practitioners [8]. However, only little tool support and research literature can be identified that actually provides approaches for solving the problem of EA model maintenance in practice.

To tackle this problem we are elaborating means for better keeping EA models in-sync with what they represent in the real world in the course of this thesis. The aim is to decrease manual work in the EA practice and increase the EA data quality attributes actuality and consistency. We are aiming to achieve this by better supporting the manual data collection processes with tool support. This includes automating the process of integrating EA data from external data sources and making use of change events. These events are either generated by data quality checks or the environment in order to trigger manual and automated updates of the EA repository.

The thesis is conducted in collaboration with the company *iteratec GmbH*, Munich, which produces the open-source EA tool *iteraplan*. The work on this thesis is at beginning of its third year with the expected completion in summer 2013.

This paper is structured as follows. In the next section we detail on the research questions we want to tackle in this thesis. After that we point out the preliminary work, such as empirical studies and implementation work, that was already conducted. We then introduce our approach to automated EA model maintenance in Section 3. From the introduction of the preliminary and future work we then derive the expected contributions in Section 4. After that we introduce the means with which we plan to evaluate the artifacts produced in the course of this thesis. We then show in Section 6 that, although the problem of automated EA model maintenance is highly relevant in practice, related work on this topic is scarce. Finally, we summarize and point to the direction of future work in Section 7.

1.1 Research Questions

The general research question of this thesis can be stated as follows:

How can EA models be better synchronized with the real world?

In order to holistically address this research problem of automated EA model maintenance, we approach the problem space not only from the technical side, but also from the point of view of how the EA model maintenance is embedded in organizations. Hence, in our research we first analyzed the context of enterprise architecture initiatives in typical organizations, i.e. the roles of EA

stakeholders, (neighboring) processes and common data sources for automated EA data collection.

This constitutes the first specific research question of this thesis as follows:

Q1 *What are the typical environmental factors for the automated maintenance of EA models, such as people, processes and available data sources for automation?*

These environmental factors also determine in which way events from the environment can be utilized to trigger changes to the EA model. Examples of such events are, e.g. a new release of an application that was developed in-house, the end of an architecture change project, or simply the scheduled execution of a manual architecture review process. This leads to the second research question:

Q2 *What are typical architecture change events and how can an EA tool collect these in order to update the EA model at the right time?*

In our preliminary work we noticed that in the context of automated EA data collection it is unrealistic to assume that every step of the collection processes can be automated. One of the main reasons is the high abstraction level of EA models. When data is collected from external sources of EA relevant information it is, e.g. very likely that duplicates are introduced or the level of abstraction of the incoming data does not fit the level of abstraction at the EA repository's side. Thus, we believe that computer aided processes with human task-lists should be used in order to include humans into the data collection processes where necessary. This leads to the third research question:

Q3 *How can the EA data collection be supported by semi-automated data collection processes from external data sources?*

In order to support the data collection processes and eventing mechanisms, an EA tool needs to be provided with context information. This context information can be responsibilities of stakeholders, available data sources, the data origin of elements in the model (manually entered or from a data source), or the date of the last change of a model element. This data has to be stored alongside with the actual EA model and has to be connected to the elements of the model. Hence, this information needs to be incorporated into the EA meta-model. This constitutes the fourth research question:

Q4 *How can an EA meta-model be devised that facilitates automated EA model updates with context information?*

As pointed out in the introduction of **Q3** the EA data collection from external data sources introduces several problems. The lack of human judgement, for example, can introduce inconsistencies such as duplicate model element entries or simply the representation of the EA at the an inappropriate (too detailed) level of abstraction. This leads to research question five:

Q5 *How can the data collection processes be supported by an EAM tool, e.g. identity reconciliation mechanisms to remove duplicates?*

After having introduced the research questions, we proceed with presenting the preliminary work that was already conducted in the course of this thesis and the research methodology that is applied.

2 Preliminary Work & Research Methodology

In the context of this thesis we have already conducted preliminary work that establishes the problem relevance and produced an initial prototype. The basic chronological steps of the research activities can be seen Table 1. Currently we are approximately at the end of phase 4.

#	Step	Activity
1	Evaluation of problem relevance	Practitioner interviews & survey & literature review
2	Elicitation of requirements for implementation and processes	Practitioner interviews & survey & literature review
3	Definition of integration processes	Practitioner interviews & survey & literature review
4	EA tool prototype and meta-model implementation	Iterative software development
5	Refinement of prototype with eventing and rules	Iterative software development & survey & interviews
6	Evaluation	Case study & expert interviews & inclusion of automation concepts in open-source EA tool <i>iteraplan</i>
7	Dissemination	Journal Publication

Table 1. Chronological steps of research including methodology.

As it can be seen in the table above, we started out with an evaluation of the problem domain. The first steps were three interviews with practitioners in the field of enterprise architecture management from the electric utility service and insurance fields. These unstructured interviews gave us a basic picture of the data collection problems in practice. A thorough literature review revealed that the related work in the area of automated data collection for EA model is scarce (see Section 6). In order to verify our assumption that the EA data collection is a major problem in EA practice, we conducted an online survey among EA practitioners. The results of the survey supported our assumption with 90% of the survey participants agreeing or strongly agreeing with the statement “The manual collection and maintenance of EA data in a sufficient quality, is one of the major challenges of EA practice.” [8]. As part of the design science process according to Hevner et al. [9] this establishes the relevance of the problem.

In the second step we elicited a set of requirements for an EA tool that supports automated maintenance of EA models from external data sources. These

requirements resulted from the literature review, the expert interviews, the survey and the experience of our industry partners of the company *iteratec* in Munich. The results of the survey are published in [8].

In the third step we defined data collection processes (independently from an implementation) that takes human interaction into account. The goal of these processes is to reduce the manual work of EA data collection and to include humans in the process via tasks only when necessary. The results of these process definitions were published in [10]. These processes are the first design science artifact produced in the course of this thesis.

Currently, we are in the phase of implementing the first prototype of an EA tool that supports the defined requirements and the proposed processes as the second design science artifact. Also, we are currently conducting another survey among EA practitioners, in order to elicit typical data sources in organizations that apply EAM. We expect that the results will allow us to make statements about what type of data can be typically gathered automatically and statements about the expected data quality. Also the results can help us to further refine our prototype in step 5 to better support realistic integration scenarios from data sources that are common in practice. The details of the prototype and our general approach will be explained in Section 3.

After the implementation phase, we plan to evaluate our work with further expert interviews and an extensive case study and the transfer of our concepts to the open-source EA tool *iteraplan* that is used in practice. We further detail our evaluation strategy in Section 5.

Finally, we aim at disseminating our findings in a journal publication.

3 Approach to Automating EA Model Maintenance

As outlined before, we have strong indication that full automation of data collection for EA models is not feasible. The reasons are that (i) EA data, especially in the business layer, is often modelled at a high-level of abstraction and these abstractions can only be made by the judgement of humans, (ii) not all data can be collected automatically from existing data sources in an organization, and (iii) in case data inconsistencies occur, such as entry duplicates, the resolution of the duplication can often only be decided by a human. Thus, the inclusion of humans in the data collection and quality assurance processes plays a major role in our approach. The goal of this thesis is to shift the work of humans in the EA process from the data collection to the quality assurance tasks. In addition to the integration of data sources that can automatically provide data on the EA, we consider events from the environment that can trigger manual or automated actions in the EA tool. In the following, we detail the prototypical architecture we are currently implementing, and explain the specifics of the meta-model we created to support the implementation with the required context information.

3.1 Tool Architecture

Figure 1 gives an overview of our implementation approach.

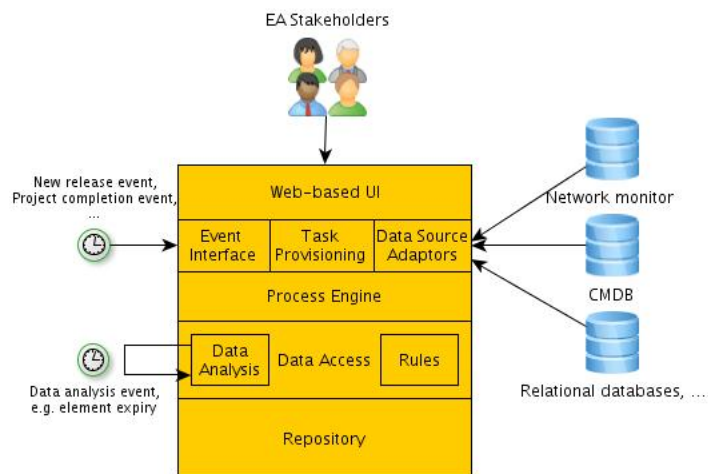


Fig. 1. Overview of our implementation approach for an EA tool that supports the recurring semi-automated collection of EA data from external sources.

The central part of the architecture is the **process engine**, based on Apache Activiti¹, which can be used to configure the semi-automated data collection with BPMN processes. The processes are driven by change events. The change events can come in the form of events from external information systems via the **event interface**, as data from external data sources and as human input from task lists created by the **task provisioning** component in the **web-based user interface**. In addition, the process can be driven by **data analysis** events, such as the expiry of a model element². The data analysis component periodically checks the EA repository for inconsistencies, duplicate entries, or expired model elements. Via the deployed processes, changes can be applied to the EA repository over the **data access layer**.

Currently, the **EA repository** is implemented as a OWL2 knowledge base using the Jena³ RDF repository. The reason for this technological choice are the ability to easily apply rules, the ability to adapt the meta-model at run-time, and being able to model on several meta-layers at runtime (refer to Section 3.2 for details).

External data sources provide EA relevant data via **data source adaptors** that have to be implemented for a each specific data source type. The adaptors take care of the mapping between external data sources and the internal EA repository data representation. For example, this could be a mapping, be-

¹ <http://www.activiti.org>

² In our approach a model element expires when it has not been changed or checked within a specified period of time

³ <http://incubator.apache.org/jena>

tween XML coming from a SOAP interface to the internal OWL representation. Note that implementing the adapters very likely has to occur individually for each company's data sources. For this prototype we plan to implement only few representative adapters, such as an adapter to a Configuration Management Database (CMDB)[11] and an adaptor to a network monitor. We acknowledge that in practice implementing the adapters can be a time consuming task as well. Hence, the trade off between the cost of implementing adapters and long term savings should be calculated as we describe in [10].

Information systems which cannot provide structured EA data, but are able to indicate EA relevant change events can fire events to the EA tool via the event interface. These events can hint responsible stakeholders, about changes such as finished projects or new information system releases. Event providers in these contexts are, for example, project portfolio management tools or software configuration management tools. This way manual data collection processes can be initiated right after the changes are applied in reality, even though no full automation is possible.

3.2 Meta-model

Another important part of our approach is the meta-model. Different EA frameworks such as TOGAF [1] and [12] prescribe the information model including the concepts as well as their attributes and relations that can be used to model the EA. Those EA information models often share some concepts like the concept *InformationSystem* but greatly differ in detail. In line with other researchers in the EA field [13–15] we believe that the EA information model should be organization specific to better resemble the specific EA of an organization. Hence, we provide a meta-model that is a generic foundation and can be used to create organization specific information models that precisely cover the desired concepts of an organization.

The important characteristic of the meta-model is that it contains automation-related concepts that are independent from the organization specific realization of the information model. However, we argue that the concept provided in our meta-model could also be applied to existing information models such as TOGAF if the organization-specificity is not considered as important by an organization. The meta-model, among others, contains concepts that provide means for:

- assigning responsible persons or roles to model element types, to specific model elements and to events,
- modeling data sources and the model element types and attributes they provide,
- tracing the origin of model elements and their attributes (i.e. whether they were entered manually or came from a data source),
- updating single model element attributes from a data source,
- definition of identifying properties of model elements that are used to eliminate duplicates and
- generating change events via the definition of expiry durations.

The result is that the data collection processes and event mechanisms have the necessary context information to operate independently of the organization specific meta-model. Details of this meta-model have been submitted for publication.

4 Contributions/Artifacts

The main contributions and produced artifacts of this thesis will be the following:

1. A thorough literature analysis on the current state of the art in enterprise architecture model maintenance.
2. The collection and analysis of potential events and data sources that can be used to automate the EA data collection.
3. Specified data collection and quality assurance processes that include the data collection via events that trigger manual input and collection from external data sources.
4. A meta-model that supports the data collection process with context information and the ability to create organization specific information models.
5. A prototypical EA tool implementation that provides eventing, process execution, and modeling functionality and supports the recurring import of EA data from external data sources.

These artifacts in combination will provide a holistic view on automation possibilities in organizations that has, so far, not been covered by research literature. They will be of practical relevance for organizations that want to optimize their EA data collection, and for EA tool vendors and users to optimize their data collection tooling.

5 Evaluation Strategy

We are aware of the fact that applying our prototype in practice will be very difficult to achieve. Thus, we apply alternative means to evaluate the artifacts produced in the course of this thesis.

First, we will create case studies that show the applicability of our approach, taking into account the findings of our survey on the EA-relevant environment in organizations, such as their processes, events, and available data sources. This way, we can give an estimate on the amount of automation that can be achieved by our approach. As part of these case studies, we will simulate semi-automated data collection from typical data sources such as CMDBs or network monitors.

Second, we will present these case studies as part of further interviews to EA experts, in order to gather input on how relevant the produced concepts are in practice.

Third, we are already in the process of transferring some automation concepts of our prototypical implementation to the open-source EA tool *iteraplan* of our industry partner *iteratec GmbH*. This way, parts of our approach can be evaluated in practice.

6 Related Work

As stated in the introduction, the related work in the EA literature is very limited. Several publications acknowledge the problem of EA model maintenance, however only one publication could be identified that presents an implemented solution.

An example of a paper that acknowledges the problem is the recent vision paper by Brückmann et al. [5] on real-time EA monitoring. The authors have the goal of providing real-time EA models. Since it is a vision paper no concrete solution approaches are presented.

Another publication that mentions EA automation is the work by Moser et al. [16]. The authors describe a process that includes automated data collection. However, the authors do not describe how the process is supplied with relevant context information about data sources and do not provide implementation details. Our meta-model may provide a basis for realizing these processes.

The work of Fischer et al. [6] discusses the federated nature of data collection in EAM, that is also relevant for the automation concepts. However, the publication also does not present any details about the meta-model requirements for automation or other implementation details.

The only publication that actually presents a concrete implementation approach for automated EA model maintenance is the work by Buschle et al. [7]. They present a tool for the instantiation of an EA model via the use of a security network scanner. The tool can be used for the initial import of model element instances describing the IT infrastructure. We argue that this approach can only be used for initial import of EA data and may introduce inconsistencies because no explicit (manual) checks are included in the approach. Our foundational meta-model could help to develop and enhance the presented tool prototype. In particular, recurring updates from arbitrary data source including network scanners could be enabled based on our meta-model.

On the technical side, several approaches originating from different IT-fields address the topic of federating information from different data sources in the business context and are hence relevant as foundations for our work. In particular, Extract-Transform-Load (ETL), data warehouses [17], Master Data Management [18] (MDM) as well as Configuration Management Databases (CMDBs) [11] target these topics. These disciplines provide a large body of foundational research that relates to the issues of automated EA model maintenance, but differ in detail. We will build on the concepts of these related disciplines where applicable.

On the practical side, the current generation of commercial EA tools mostly support the batch import of EA data from external sources. I.e. the import of data from files in common formats such as XML, CSV or Excel. In addition, some tools support the data import from relational databases and selected information systems such as CMDBs. To the best of our knowledge no current EA tool supports the traceability of sources of model element instances and recurring updates that map elements to their original sources, as well as the process and eventing features sketched in this paper.

7 Conclusion & Next Steps

In this paper we have shown that the maintenance of EA models is a relevant problem in practice. This has been acknowledged by both researchers and practitioners. Nevertheless, related approaches are scarce as we outlined. We presented the key research questions that we aim to answer in this thesis in order to provide holistic approach to solving the socio-technical problem of enterprise architecture model maintenance. We then discussed the implementation approach that is based on processes for data collection from external data sources and for quality assurance. The underlying meta-model provides means to create organization specific information models and holds relevant data to drive the data collection processes. For example, the meta-model enables maintaining the connection of a model element to its data source, and thus enables recurring updates from its source. Finally, we highlighted the main expected contributions of this thesis and pointed out related EA research literature and neighboring research fields from which foundations can be drawn.

The next steps in the course of this thesis will be the evaluation of our survey on EA data sources and the further implementation of our prototype with the input from the survey. We will specifically focus on the concepts of model element identity reconciliation.

References

1. The Open Group: TOGAF “Enterprise Edition” Version 9. <http://www.togaf.org> (cited 2011-06-08) (2009)
2. Zachman, J.A.: A framework for information systems architecture. *IBM Systems Journal* **26**(3) (1987) 276–292
3. Winter, K., Buckl, S., Matthes, F., Schweda, C.: Investigating the state-of-the-art in enterprise architecture management methods in literature and practice. In: *MCIS 2010 Proceedings*. (2010)
4. Buckl, S., Matthes, F., Schweda, C.M.: Future research topics in enterprise architecture management—a knowledge management perspective. In: *Service-Oriented Computing. ICSOC/ServiceWave 2009 Workshops*, Springer (2010) 1–11
5. Brückmann, T., Gruhn, V., Pfeiffer, M.: Towards real-time monitoring and controlling of enterprise architectures using business software control centers. In Crnkovic, I., Gruhn, V., Book, M., eds.: *Software Architecture*. Volume 6903 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg (2011) 287–294
6. Fischer, R., Aier, S., Winter, R.: A federated approach to enterprise architecture model maintenance. In Reichert, M., Strecker, S., Turowski, K., eds.: *EMISA*. Volume P-119 of *LNI*, GI (2007) 9–22
7. Buschle, M., Holm, H., Sommestad, T., Ekstedt, M. In: *A Tool for Automatic Enterprise Architecture Modeling*. (2011) 25–32
8. Farwick, M., Agreiter, B., Ryll, S., Voges, K., Hanschke, I., Brey, R.: Requirements for automated enterprise architecture model maintenance. In: *13th International Conference on Enterprise Information Systems (ICEIS)*, Beijing. (2011)
9. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. *MIS Q.* **28**(1) (March 2004) 75–105

10. Farwick, M., Agreiter, B., Breu, R., Ryll, S., Voges, K., Hanschke, I.: Automation processes for enterprise architecture management. In: 2011 IEEE 15th International Enterprise Distributed Object Computing Conference Workshops, IEEE (Aug 2011) 340–349
11. OGC: ITIL Lifecycle Publication Suite Books, 2nd impression. TSO (2007)
12. Lankhorst, M.: Enterprise Architecture at Work: Modelling, Communication and Analysis. Springer, Berlin (2005)
13. Buckl, S., Ernst, A.M., Lankes, J., Schneider, K., Schweda, C.M.: A pattern based approach for constructing enterprise architecture management information models. In Oberweis, A., Weinhardt, C., Gimpel, H., Koschmider, A., Pankratius, V., Schnizler, eds.: Wirtschaftsinformatik 2007, Karlsruhe, Germany, Universitätsverlag Karlsruhe (2007) 145–162
14. Aier, S., Kurpjuweit, S., Riege, C., Saat, J.: Stakeholderorientierte dokumentation und analyse der unternehmensarchitektur. In: GI Jahrestagung (2). (2008) 559–565
15. Lagerström, R., Saat, J., Franke, U., Aier, S., Ekstedt, M.: Enterprise meta modeling methods – combining a stakeholder-oriented and a causality-based approach. In Halpin, T.A., Krogstie, J., Nurcan, S., Proper, E., Schmidt, R., Soffer, P., Ukor, R., eds.: Enterprise, Business-Process and Information Systems Modeling, EMMSAD 2009, Springer (2009) 381–393
16. Moser, C., Junginger, S., Brückmann, M., Schöne, K.: Some process patterns for enterprise architecture management. In: Proceedings, Workshop on Patterns in Enterprise Architecture Management (PEAM2009), Bonn. (2009) 19–30
17. Vassiliadis, P.: Survey of Extract-Transform-Load Technology. International Journal of Data Warehousing and Mining 5(3) (2009) 1–27
18. White, A., Newman, D., Logan, D., Radcliffe, J.: Mastering master data management (2006) Gartner, Stamford.