

Situational Reference Model Mining

Jana-Rebecca Rehse

Institute for Information Systems (IWi) at the German Center for Artificial
Intelligence (DFKI GmbH) and Saarland University
Campus D3 2, Saarbruecken, Germany
`Jana-Rebecca.Rehse@iwi.dfki.de`

Abstract. Reference models can be considered as special conceptual models that serve to be reused for the design of other conceptual models. Due to an ongoing need for high-quality reference models, reference model mining, i.e. the (semi-)automatic derivation of reference models from a set of existing process models, has recently gained the attention of researchers. The presented dissertation project addresses the concept of Situational Reference Model Mining, i.e. the idea that mined reference models, although based on the same input data, are intended for different use cases and thus have to meet different requirements. These requirements determine the reference model character and thus the technique that is best suited for mining it. The dissertation's major objective is to design, elaborate, and validate a method for Situational Reference Model Mining, which provides reference modelers with a clear guideline on how to use automated reference model mining techniques to their advantage.

Keywords: Inductive reference modeling, Reference model mining, Reference model design principles, Context-aware process design

1 Motivation and Research Problem¹

Reference models can be considered as special conceptual models that serve to be reused for the design of other conceptual models [4]. They provide a template for process models in a certain industry and thus facilitate a resource-efficient implementation of the respective process and its adaption to the individual needs of an organization. This way, companies may benefit from best practices and industry-specific experience. The use of reference models is associated with a higher quality of processes and process models, as it simplifies internal communications by introducing a common terminology and considerably reduces the resources required for Business Process Management (BPM) [5].

Leveraging the multiple benefits of reference modeling in both research and practice depends on the widespread availability of high-quality reference models. Generally accepted and widely used reference models exist only for a few industries, such as the IT Infrastructure Library (ITIL) for IT service management or

¹This section is based on previous work presented in [1–3] and has been adapted.

the Supply Chain Operations Reference Model (SCOR) for supply chain management, eliciting an ongoing need for reference model construction methods.

Reference models may be constructed both deductively and inductively. Deductive methods or “top-down” approaches employ generally accepted theories and principles. Models are constructed on a theoretical base and gradually substantiated along the way. In contrast, inductive or “bottom-up” development of reference models makes use of real-world data such as existing process models or execution logs. It focuses on similarities and commonalities within the input data and abstracts from a model’s individual features. As is it possible to automate the inductive derivation of a potential reference model from a collection of existing process models (Reference Model Mining), inductive reference modeling may contribute to the trend of combining BPM with artificial intelligence (AI)².

Until a few years ago, although inductive development was generally acknowledged, the literature on methods for reference model development was dominated by deductive approaches. This is interesting, as many existing reference models were at least partially developed inductively [3]. Recently, however, there have been some research activities regarding inductive reference modeling. Methods have been proposed to (semi-)automatically derive reference models from both instance level data in the form of process logs [6] and type-level data, i.e. process models [7]. They make use of a multitude of existing computational techniques, such as subgraph isomorphism [3], genetic algorithms [8], or heuristic approximations of the graph edit distance [9].

When applying reference model mining (RMM), i.e. (semi-)automatic inductive reference model development, it becomes evident that the reference model content and character are significantly influenced by the choice of mining technique and its parametrization. Different mining techniques yield different models, even when applied to the same set of input models. Meanwhile, given a reuse-oriented understanding of reference models, the requirements to the reference model are determined by the situational context it will be used in, similar to the adaptation of software developments methods in Situational Method Engineering (SME) [10]. It will hence increase an inductively developed reference model’s usefulness and situational suitability, if the mining technique is determined by the circumstances of its intended reuse. We call this concept “Situational Reference Model Mining” (S-RMM), i.e. extending RMM towards consciously considering the situational context when designing and using a reference model (cf. Fig. 1).

Designing a method for S-RMM and understanding the influencing factors behind it is the main objective of the presented dissertation project titled “Situational Reference Model Mining — Conceptual Design and Selected Applications”, which is conducted at the Institute of Information Systems at Saarland University, under the supervision of Prof. Dr. Peter LOOS. It started in October 2015 and is planned to be completed in 2020. Some preliminary results and cited publications were already achieved during the author’s time as research assistant at the same institute (April 2011 until October 2015).

²Cf. the keynote speech “Intelligent Continuous Improvement, When BPM Meets AI” by Miguel VALDÉS at BPM 2017

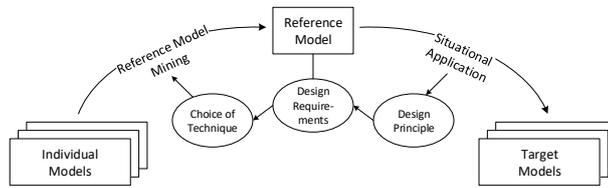


Fig. 1. Main idea behind Situational Reference Model Mining

2 Research Method and Research Questions

As the main objective of the project is the design of an artifact (a method for S-RMM), it follows the design science research paradigm (DSR), originally coined by HEVNER et al. [11]. More specifically, it applies the design science methodology by WIERINGA [12], which describes design science as the “design and investigation of artifacts in context”. According to its template for design questions [12, p.16], the overall goal of the dissertation is the following:

Improve availability and suitability of reference models
by introducing a method for reference model mining
that considers the situational context
in order to facilitate reference model development and usage.

This goal shall be reached by answering the following four research questions. The first two questions are knowledge questions, determined to gather knowledge on the reference modeling process. The second two questions are design questions, asking for new artifacts to improve the reference modeling process.

1. Which contextual factors influence reference model design?
2. Which techniques exist for reference model mining and which contextual factors influence their choice and applicability?
3. How can reference model mining techniques be matched with situational contexts to produce applicable reference models?
4. How can reference model mining be applied in a given application case, considering situational contextual factors?

The first question addresses the contextual factors that influence reference modeling. It is an open descriptive empirical research question, meant to observe reference modeling in real-world scenarios, gathering knowledge about the context of the artifact to be designed without the need for explanations. In contrast, the second question is an open explanatory empirical knowledge question with the goal to identify existing RMM techniques and to determine their potentials and limitations. We want to know under which situational circumstances a certain technique may and may not produce an applicable and well-suited reference model. Because these circumstances may apply to multiple techniques, they should be conclusively explained by the technique’s properties.

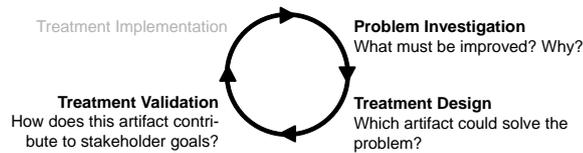


Fig. 2. The design cycle to be followed [12]

The third question joins the existing artifacts (i.e. the RMM techniques) with the context by designing a method to match an existing technique with a situational context, such that the produced reference models are applicable and useful for their intended purpose. Therefore, the results from the previous two questions need to be related to each other through some intermediary concept. This concept and the matching method are then used in the fourth question, which finally addresses the overall goal of the dissertation project, as sketched by the template above. It is set out to design a comprehensive method for S-RMM, ranging from the identification of a need for a new reference model to its application and validation in a concrete use case.

Knowledge questions and design questions are answered by means of different research methods, following WIERINGA’s design cycle [12, p.27ff.] in Fig. 2 and FRANK’s pluralistic conception of research methods in IS research [13]. Problem investigation concerns learning more about the problem context, including stakeholder goals, conceptual problem frameworks, and related phenomena. In our case, this means an in-depth analysis of the context factors that determine a reference model’s applicability and usefulness. Therefore, we will perform literature reviews and conduct interviews with expert modelers. Treatment design consists of defining requirements for the artifact and designing it accordingly. Therefore, we will translate the knowledge from the first stage into concrete requirements, before the main artifact (the method) is designed. In the treatment validation stage, the artifact is tested in its context. In our case, this validation will be conducted by applying the method for inductively developing a reference model for a to-be-defined use case, using observational case studies as well as single-case mechanism experiments along with qualitative research methods.

3 State of the Art and Related Solutions³

The concept of situational reference model construction based on design principles is not new, but has yet only been elaborated for deductive reference model development [4]. Inductive reference model development has only recently gained attention in research, so there is little methodological seminal work. We presented first ideas towards S-RMM in [2], where the choice of an appropriate mining technique is discussed. In [14], we illustrate concrete challenges of inductive refer-

³This section is based on previous work presented in [1] and has been extended.

ence modeling according to a seven-stage framework. A number of contributions describe concrete RMM techniques, but do not take on a methodological perspective, reflecting on the ways of model construction and the requirements of specific use cases. A first overview on existing techniques is given in [14]. Since then, we have identified a few additional the contributions, such as [15].

The term “context” can be defined as “any information that can be used to characterize the situation of an entity” [16]. The idea of adapting an artifact, such as a method or a model, to fit the specific context in which it is used, originated in the software engineering domain, where a rigid one-size-fits-all methodology for software development is not only unattainable but also inefficient [10, p.5]. Their explanations and formalizations of the terms “situation” and “context” can be adapted to the BPM domain [17]. Regarding business processes, the context describes the “environment in which a business process artefact is used” [18], formalized in terms of a “minimum set of variables containing all relevant information that impact the design and execution of a business process” [19]. Similar to reference modeling, context is also considered in process mining research to achieve better and more specific results (e.g. [20]).

S-RMM is innovative in the way that it combines the domain of reference modeling with the idea to adapt an artifact to the context in which it is used, as coined by SME, and techniques to derive a process model from collected real-life data, as in process mining. Giving a manual procedure model for inductive reference modeling based on RMM techniques, S-RMM goes beyond SME, because it does not directly influence the artifact itself (the model), but instead adapts the method used for deriving it. It differs from process mining, as both process models and execution logs are considered potential data sources. To our knowledge, no comparable approaches or solutions exist in state-of-the-art literature.

4 Contributions⁴

Whereas most of the author’s contributions have so far addressed research question 2 (e.g [2, 21, 14]), the main contribution to S-RMM is made in [1], where a first answer is given to questions 3 and 4. With a reuse-oriented conceptualization of reference models, their main purpose is to serve as an orientation in the design of new business process models. In this context, we decipher two general design processes [4]. Deriving an individual model from a reference model is known as “Design With Reuse” (DWR), i.e. an existing model is used as a blueprint offering guidance to the process model designer by giving suggestions for both content and design of the individual model. On the other hand, “Design For Reuse” (DFR) describes the process of constructing a (reference) model for the purpose of being reused, i.e. composing model parts and domain knowledge, such that they achieve a certain degree of universality.

Considering a model construction process, there exist several different techniques for deriving a conceptual model from another one. These so-called design

⁴This section is based on previous work presented in [1] and has been adapted.

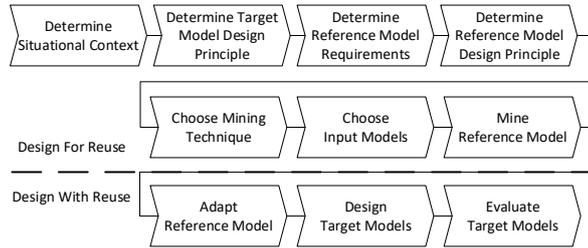


Fig. 3. Procedure model for Situational Reference Model Mining [1]

principles describe how the content of the original model is adopted, adapted, and extended in order to create a new model. Five design principles are described in the literature [4]. Each configuration, instantiation, specialization, aggregation, and analogy may be used in the context of reference modeling and applied to both DFR and DWR. However, not every design principle may be applied to every reference model, nor may every intended target model be derived by any design principle. Instead, the choice of model design principle depends on the situational circumstances, i.e. the requirements posed to the target model and the construction process itself. These factors also determine the character and the choice of an appropriate reference model for a certain application context.

The main idea behind our method for S-RMM is to match RMM techniques with applicable contexts by means of selecting appropriate design principles for the construction of both target and reference models. Depending on the situational circumstances and the target model requirements, a certain design principle is applied to derive the target model from the reference model. This design principle poses certain restrictions and requirements to the reference model design, which is mainly influenced by the choice of technique that was used to mine the reference model. Hence, the choice of mining technique is ultimately determined by the situational context of the reference model application.

Research question 4 is addressed by the first conceptual design for a procedure model for S-RMM, shown in Fig. 3. It constitutes the main artifact to be delivered in the dissertation. It is designed around the conceptualization of S-RMM in Fig. 1 and describes a generic execution process of an S-RMM application. Of the ten steps, seven belong to DFR and focus on the reference model construction (i.e. the actual mining), whereas three belong to DWR and are concerned with the target model construction (i.e. the reference model application).

The procedure model starts with defining the situational context, which determines the target model design principle in the second step. As stated below, this step is not yet conclusively elaborated and requires a lot of additional research. In the third step, the reference model requirements are defined, which are partly determined by the used principle and partly by the context itself. The reference model design principle is derived from these requirements, with the help of additional methods and artifacts, as explained below. For certain combi-

nations of target model and reference model principle, several mining techniques are available; their choice may depend on additional factors. The input models can only be finally selected after the mining technique, as some techniques pose additional requirements to their input data. Afterwards, the model can be mined, concluding the DFR process. In the DWR process, the model is manually or automatically adapted, before the target models can be derived and evaluated regarding their purpose.

To provide a guideline for applying the procedure model and to answer research question 3, we analyzed existing mining techniques regarding their underlying principles and requirements (cf. Table 1). For each target model design principle, we suggest corresponding reference model design principles and, for each pair, a example of a suitable mining technique. A more detailed version of this table, classifying many state-of-the-art techniques can be found in [1].

Table 1. Analysis of matching principles and according mining techniques

Target Model Design Principle	Reference Model Design Principle	RMM Technique (Example)
Configuration	Aggregation	[2]
	Analogy	[7]
Instantiation	Aggregation	[21]
Specialization	Aggregation	[2]
	Analogy	[7]
Aggregation	–	–
Analogy	Configuration	[6]
	Aggregation	[2]
	Analogy	[8]

5 Intended Future Work

Although the to-be-designed artifacts are roughly sketched in the previous sections, there are numerous steps to be conducted in order to complete the research activities that are required to satisfactorily answer the posed research questions. Readers will notice that the above procedure model was designed without performing an in-depth requirements analysis first. This analysis will be built on an identification of relevant context factors, which will be identified by means of a literature analysis of inductive reference modeling case studies and seminal works on reference modeling. The goal is an integrated and conclusive model of influential context factors for reference model development. This model will then be validated and enhanced with expert interviews and provide a comprehensive answer to research question 1. From that, we can derive requirements, which can be used to adapt and refine the procedure model.

To conclusively answer research question 2, it is necessary to update the literature analysis from [14], in order to capture the newest developments in the

field of RMM techniques. In addition, there needs to be an extensive literature review for the identification of techniques that are not explicitly set out for RMM, but could be used for that purpose, e.g. from the field of process mining.

Our analysis of matching existing RMM techniques with accepted design principles for reference models in Tab. 1 is a preliminary sketch of a conclusive answer to research question 3. It needs to be complemented by additional principles such as modification, elimination, or union, potentially by means of a structuring framework. The RMM techniques identified in the previous research question need to be matched the the individual principles. Most importantly, there needs to be some matching procedure between the context factors from the first question to the design principles, and then, the RMM techniques.

The procedure model in Fig. 3 is the framework for an answer to research question 4. Not only does it require an in-depth analysis of posed requirements and an according adaptation and refinement, it also has to be substantiated by concrete operationalizations of the individual stages. At this point, it is unclear how the stages should be executed. To complete the design cycle, the procedure model is supposed to be evaluated in practical case studies, where it is used to inductively develop a reference model for a certain context. These case studies have yet to be determined, planned, and executed.

References

1. Rehse, J.R., Fettke, P.: Towards Situational Reference Model Mining - Main Idea, Procedure Model & Case Study. In: Leimeister, J.M., Brenner, W. (eds.) Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik. Internationale Tagung Wirtschaftsinformatik (WI-2017), February 12-15, St. Gallen, Switzerland. AIS (2017)
2. Rehse, J.R., Fettke, P., Loos, P.: An execution-semantic approach to inductive reference models development. In: 24th European Conference for Information Systems (ECIS). European Conference on Information Systems (ECIS-16), June 12-15, Istanbul, Turkey. Association for Information Systems (AIS) (2016)
3. Rehse, J.R., Fettke, P., Loos, P.: A graph-theoretic method for the inductive development of reference process models. *Software & Systems Modeling* 16(3), 833–873 (2017)
4. vom Brocke, J.: Design principles for reference modeling: Reusing information models by means of aggregation, specialisation, instantiation, and analogy. In: Fettke, P., Loos, P. (eds.) Reference Modeling for Business Systems Analysis, pp. 47–75. Idea Group Publishing, Hershey (2007)
5. Fettke, P., Loos, P.: Perspectives on reference modeling. In: Fettke, P., Loos, P. (eds.) Reference modeling for business systems analysis, pp. 1–20. Idea Group Publishing, Hershey, PA (2007)
6. Gottschalk, F., van der Aalst, W.M., Jansen-Vullers, M.H.: Mining reference process models and their configurations. In: Meersman, R., Tari, Z., Herrero, P. (eds.) On the Move to Meaningful Internet Systems: OTM 2008 Workshops, OTM Confederated International Workshops and Posters, Monterrey, Mexico, November 9-14, 2008. Lecture Notes in Computer Science, vol. 5333, pp. 263–272. Springer Berlin Heidelberg (2008)

7. Li, C., Reichert, M., Wombacher, A.: Mining business process variants: Challenges, scenarios, algorithms. *Data & Knowledge Engineering* 70(5), 409–434 (2011)
8. Yahya, B., Bae, H., Bae, J., Kim, D.: Generating valid reference business process model using genetic algorithm. *International Journal of Innovative Computing, Information and Control* 8(2), 1463–1477 (2012)
9. Ardalani, P., Houy, C., Fettke, P., Loos, P.: Towards a minimal cost of change approach for inductive reference model development. In: *Proceedings of the 21st European Conference on Information Systems. European Conference on Information Systems (ECIS-13)*, 21st, June 5-8, Utrecht, Netherlands. vol. Paper 127. AIS (2013)
10. Henderson-Sellers, B., Ralyté, J., Ågerfalk, P.J., Rossi, M.: *Situational method engineering*. Springer (2014)
11. Hevner, A., March, S., Park, J., Ram, S.: Design science in information systems research. *MIS quarterly* 28(1), 75–105 (2004)
12. Wieringa, R.J.: *Design Science Methodology for Information Systems and Software Engineering*. Springer Berlin Heidelberg (2014)
13. Frank, U.: Towards a pluralistic conception of research methods in information systems research. ICB-Research Report No. 7, Institut für Informatik und Wirtschaftsinformatik (ICB) der Universität Duisburg-Essen (2006)
14. Rehse, J.R., Hake, P., Fettke, P., Loos, P.: Inductive reference model development: Recent results and current challenges. In: Mayr, H.C., Pinzger, M. (eds.) *INFORMATIK 2016 (LNI Volume P-259)*. pp. 739–752. *Lecture Notes in Informatics, Gesellschaft für Informatik (GI)*, Bonn (2016)
15. Leng, J., Jiang, P.: Granular computing–based development of service process reference models in social manufacturing contexts. *Concurrent Engineering* 25(2), 95–107 (2017)
16. Dey, A.: Understanding and using context. *Personal and ubiquitous computing* 5(1), 4–7 (2001)
17. Kornysheva, E., Deneckère, R., Claudepierre, B.: Contextualization of method components. In: *Fourth International Conference on Research Challenges in Information Science (RCIS)*. pp. 235–246. IEEE (2010)
18. Born, M., Kirchner, J., Müller, J.P.: Context-driven business process modeling. In: *The 1st International Workshop on Managing Data with Mobile Devices (MDMD 2009)*, Milan, Italy. pp. 6–10 (2009)
19. Rosemann, M., Recker, J.: Context-aware process design: Exploring the extrinsic drivers for process flexibility. In: *The 18th International Conference on Advanced Information Systems Engineering. Proceedings of Workshops and Doctoral Consortium*. pp. 149–158. Namur University Press (2006)
20. Li, J., Bose, R.P.J.C., van der Aalst, W.M.P.: Mining context-dependent and interactive business process maps using execution patterns. In: zur Muehlen, M., Su, J. (eds.) *Business Process Management Workshops: BPM 2010 International Workshops and Education Track*, Hoboken, NJ, USA, September 13-15, 2010, Revised Selected Papers. pp. 109–121. Springer Berlin Heidelberg (2011)
21. Rehse, J.R., Fettke, P., Loos, P.: Eine Untersuchung der Potentiale automatisierter Abstraktionsansätze für Geschäftsprozessmodelle im Hinblick auf die induktive Entwicklung von Referenzprozessmodellen. In: Alt, R., Franczyk, B. (eds.) *Proceedings of the 11th International Conference on Wirtschaftsinformatik. Internationale Tagung Wirtschaftsinformatik. Internationale Tagung Wirtschaftsinformatik (WI-2013)*, February 27 - March 1, Leipzig, Germany (2013), (In German)