Two Replication Studies for Evaluating Artefact Models in RE: Results and Lessons Learnt

Birgit Penzenstadler  
Institute for Software Research  
University of California, Irvine, USA  
bpensens@uci.edu

Jonas Eckhardt, Daniel Méndez Fernández  
Institut für Informatik  
Technische Universität München, Germany  
{eckharjo|mendezfe}@in.tum.de

Abstract—Artefact-based requirements engineering (RE) describes the idea of establishing a company-wide reference model by putting the focus on the RE artefacts and their dependencies rather than dictating a strict process with interconnected methods. Although we could make first empirical studies on the benefits and shortcomings of artefact-based RE, however, we still have little evidence for our first results. The reason is that the conducted case studies focus on the isolated application of artefact-based RE approaches in individual socio-economic contexts and, thus, the findings can hardly be generalised.

The contribution of this paper is to report on two conducted replication studies to strengthen our confidence on the benefits and shortcomings of applying artefact orientation in RE. To this end, we replicated an industrial case study with partners from two companies. Those replications form part of a research project where all partners are working with the same artefact-based requirements engineering approach and its tool-supported realisation.

Our results give deeper insights into artefact-based RE and contribute to a reliable database due to comparability among the studies. This allows for first conclusions on the actual impact artefact orientation has on requirements engineering.

Keywords—Replication Study; Case Study; Evaluation Research; Requirements Engineering; Artefact Orientation;

I. INTRODUCTION

One significant problem in requirements engineering (RE) is that we still lack empirical evidence that reliably and objectively indicates the effectiveness of a specific approach or technique when applied in industrial settings. Consequently, dissemination of research results in RE to practice remains a challenge because there is no sufficient knowledge on strengths and limitations of given approaches.

Condori-Fernandez et al. [4] conducted a systematic mapping study on empirical evaluations of specific software requirements specifications techniques where they identified only eight papers where the evaluations took place in an industrial setting, and further ten in a mixed industrial-academic setting. Moreover, each of those evaluations focussed on the isolated investigation of single techniques. This shortcoming makes it hard to generalise findings in an empirically sound manner and gain objectivity in the results, holding not only for software engineering:

“We do not take even our own observations quite seriously, or accept them as scientific observations, until we have repeated and tested them. Only by such repetitions can we convince ourselves that we are not dealing with a mere isolated coincidence, but with events which, on account of their regularity and reproducibility, are in principle intersubjectively testable.” [13]

The usage of isolated experiments and case studies without replication in different contexts is not sufficient for significant validation research in software engineering, as pointed out by Juristo and Vegas [6], Shull et al. [15], and Brooks et al. [2]. The need for replicating experiments is especially high for RE, because of the involved stakeholders and the sensitive context both have a strong influence on any experimental endeavour in validating an anyway hardly standardisable discipline.

For example, in [7], we report on a case study with a street traffic management business unit at Siemens and discuss the benefits and shortcomings of applying artefact orientation in RE. Although this study led to concrete results and we were able to induce particular improvement goals and new concepts, we also pointed out that the most critical threat to the validity was the envisioned socio-economic context hampering the generalisation of our results.

To tackle this more general shortcoming, we propose to create a reliable source for empirical studies on artefact-based RE by conducting a family of studies in [11]. To do so, we aim at performing independent study replications in various projects, which then—if the study designs are followed as closely as possible—will lead to a more reliable database of empirical studies.

In this paper, we perform one further step towards such a reliable database, by reporting on the first replication studies. We replicated an industrial case study with two industrial partners in the project ARAMiS, where 40 partners are working with the same artefact-based requirements engineering approach and its tool-supported realisation.

Those replications are based on the design introduced in [11]. The aim is to further investigate the benefits and shortcomings when applying an artefact-based requirements engineering approach in an industrial setting.
Outline: In Sect. [II] we discuss the fundamentals and the related work, including the domain-independent artefact model for RE, which we evaluate in our replication studies. In Sect. [III] we present the case study design, before reporting the results in Sect. [IV]. Finally, in Sect. [V] we provide some lessons learnt, before concluding our paper with a summary of conclusions in Sect. [VI].

II. FUNDAMENTALS AND RELATED WORK

In the following, we introduce the fundamentals and related work.

A. Related Work

The related work for the paper at hand includes artefact-based requirements engineering and empirical studies in requirements engineering.

1) Artefact-oriented Requirements Engineering: The basic idea of artefact orientation consists in defining a reference model of all relevant artefacts and their dependencies while leaving open the way of their creation. The focus thus lies on what to create rather than on how to create it. In RE, there exist several artefact models, such as the one of Berenbach et al. [1] chap. 2], who describe RE artefact modelling with the key components to be a measurable reference model and respective process guidelines.

In general, artefact models for RE range from a level of detail to provide tool support by capturing the basic concepts of the domain in detail data models to very generic checklists as given at the example of the Vol`ere Requirements Specification Template.

Hence, artefact orientation remains a paradigm with various interpretations and manifestations in practice. To tackle the problem of a blurry terminology and to foster the discussions about this paradigm, we introduced a meta model for artefact-based RE in [3]. This meta model defines the basic concepts of artefact-based RE, i.e. which elements are necessary to define an artefact (structure, content), or how an artefact relates to further software process concepts like “method” or “role”. This supports the systematic creation of artefact-based RE approaches covering all elements of software processes and, thus, the integration and customisation of an artefact-based RE as part of a software process.

2) Empirical Studies in Requirements Engineering: In [3], Cambell et. al describe that today’s experiments need cross-validation and replication at other times and under other conditions such that they can be objectively interpreted with confidence. This is confirmed for software engineering by Shull et al. [15] and Brooks et al. [2]. The paper at hand bases on the principle of replication to make the interpretation of empirical findings for artefact-based RE more confident.

Gómez et al. perform in [5] a study on different replication types and identify a classification for replications. In this paper, we refer with “replication” to the empirical generalisation, where “a previous study is repeated on different populations. The researcher runs an empirical generalisation to test the extent to which the study results can be generalised to other populations. It follows the original experimental procedures as closely as possible”.

As referenced in the introduction (Sec. [I]), Condori-Fernandez et al. perform a systematic mapping study on empirical evaluation of software requirements specifications techniques that is reported on in [4]. They could identify rather few studies compared to the number of publications in requirements engineering, and even these evaluations treated the isolated validation of individual techniques.

The research presented in this paper is, inter alia, motivated by this shortage of evaluation research. Our replication studies are based on the case study presented in [7]. In this initial case study, to which we will refer as the replication base, we report on a case study with a street traffic management business unit from Siemens on the application of an artefact-based requirements engineering approach. We discussed the different benefits and shortcomings of that approach, but remained aware that the empirical evidence is limited to the sensitive context of our study. Still, we are not aware of any other studies that evaluate and assess artefact orientation in requirements engineering, let alone in a family of studies, apart from the one proposed in [11].

B. The Domain-independent Artefact Model

This section describes the domain-independent artefact model under evaluation in the studies at hand. It is presented in further detail in [9].

1) Concept: Reference Model and Instances: The main principle of artefact-orientation is to define what has to be accomplished (the work product) instead of how it has to be accomplished (the steps that have to be taken). As introduced in [8], an artefact is thereby defined by its content, structure, the concepts that describe the content, and its notation (syntax and semantics).

Figure 1 illustrates the instantiation of the reference model in a specific project. The left hand side represents the structural view and the right hand side the content view defined along the structure. The reference model is based on a structure, the coarse-grained model of content item containers in the upper left corner, that is detailed in the upper right by the more specific concepts, depicted in a class diagram. The project-specific artefact instances in the lower left corner are, e.g., the requirements specification documents that are elaborated according to the structure given by the reference model, which are in detail using the concepts provided by the reference model as illustrated in the lower right corner.

2) Domain-independent Content Model: The content model was developed on the basis of experiences from previous research and industrial collaborations, e.g. [12], [8].
and [7], and in a series of workshops with the industrial partners in the project ARAMiS [9]. The model, depicted in Fig. 2, has four levels of abstraction: context level, system level, subsystem level, and technical architecture level. Requirements Engineering focuses on the upper two levels. The two lower abstraction levels are included to facilitate integration with design and to enable bottom-up propagation of technical constraints.

Each level features a number of content items, denoted in ovals in Fig. 2. The depicted relations denote only a first initial input for derived content items and are not exhaustive. We chose the reduced representation to indicate the primary information flow during elaboration but without the feedback cycles, because a complete net graph of the relations between content items would overload the figure and not improve the understanding of the model. A detailed description of the individual content items is provided in [9].

3) Tool Support: For the project-wide tool support in ARAMiS, we used the Enterprise Architect (EA) for model-based requirements engineering and implemented the artefact model as a Model Driven Generation (MDG) Technologies plugin for EA [1].

For turning the content model presented in Fig. 2 into a concrete artefact model that can actually be supported by a tool, it is necessary to define the structure and syntax of the content items. We decided to use a combination of SysML and UML to define the items of the content model syntactically, since UML and SysML are both frequently used by the industrial partners. Furthermore, we developed an exemplary guideline, which helps the requirements engineer at our industrial partner to use the tool.

III. CASE STUDY DESIGN

We organise the reporting on our replications according to Runeson and Höst [14]. Details on the case study design are provided in [11] where we define the overall study design for the family of studies. In order to form part of a family of studies, we reuse the three research questions, the case and subject selection as well as the data collection and analysis procedures throughout the replication study (see also [7]). We shortly describe how we select the case and the subjects and how we collect and analyse the data.

A. Research Questions

The overall goal is to investigate the advantages and limitations in the application of our RE artefact model in an industrial environment. The following research questions, replicated from [7], are detailed with the evaluation criteria shown in Table 1:

- http://www.sparxsystems.de/uml/neweditions/
- Available for download including slide tutorial at http://www4.in.tum.de/~ecxharjo/ARAMiS-MDG-Plugin.zip

Figure 1. Concept of Reference Model and Instances

Figure 2. Overview of contents of the ARAMiS model
RQ1 Does the artefact-based approach improve the usability of the requirements engineering reference process?

One aim of artefact orientation consists in the flexibility of the process, i.e., its usability in differing project environments while leading to reproducible results.

RQ2 Does the artefact-based approach improve the syntactic quality of the created artefacts?

Once we analysed the actual process for creating the artefacts with respect to individual project influences, we want to know whether the created artefacts are of higher syntactic quality due to the underlying model-based philosophy.

RQ3 Does the artefact-based approach improve the semantic quality of the created artefacts?

Finally, for implementing the requirements it is not only important that the syntactic quality is high, but also that the requirements are stated correctly and sufficiently detailed, i.e., the semantic quality is high.

B. Case and Subject Selection

The subjects are industry participants as focus group, researchers, and an external reviewer for comparing the specifications. As cases serve specific projects where two of our research partners from the ARAMiS project are actively involved in their respective companies: one from the automotive domain (BMW) and one from the avionics domain (Cassidian). We rely on interviews by the industry participants to analyse the usability of the reference model, and the created models, e.g., documents, UML models.

C. Data Collection Procedures & Assessment

We collect and analyse the data independently at the two aforementioned different partners’ sites in a series of requirements engineering workshops. The new artefact-based approach is applied in joint action research workshops and separate subsequent isolated modelling sessions. After finalising the requirements specifications, we perform an assessment of each research question with a set of closed (Likert-scale) and open (rationale for the rating) questions to comparatively rate both RE approaches (see Table I). Ease of use is assessed by the industry participants on basis of the experiences during the elaboration. The quality in the created artefacts is assessed on basis of the created models.

D. Analysis & Validity Procedures

The analysis is performed in a qualitative manner due to the sensitivity of the context and the low number of interview participants. Therefore, we interpret the answers to the open questions on a descriptive level of analysis. To ensure the validity of our replications, we apply a number of measures: method triangulation, focus groups, and researcher triangulation.

IV. Case Study Results

After the description of the cases and subjects, we describe the results and structure them according to the research questions. For reasons of confidentiality, we can not give detailed information on the system under consideration and related artefacts created during the data collection, but we will provide an illustrative overview.

A. Subject Description: The Roles

At the automotive partner’s site, the subjects are a developer and a project manager from the research department with a background in series development as well as mechanical and electrical engineering. At the avionics partner’s site, the subjects are also a project manager from the research department and a developer from series development.

The interviews are held by the authors of the paper at hand. The overall case study design [11] had foreseen an external review to be conducted to compare the ARAMiS specifications with specifications of the same systems elaborated by using the respective former reference model of the company. As this turned out to be impossible only during the interviews due to NDA issues, the reviewer’s participation was limited to a role as additional interviewer.

B. Case Description: CPS, Automotive, and Avionics

The project Automotive, Railway and Avionics in Multicore Systems (ARAMiS) is a German publicly funded research project where 40 partners from academia and industry work on an integrated approach for developing and realising cyber physical systems (CPS) scenarios. These scenarios span over the mobility domains Automotive, Avionics, and Railway and involve cross-domain interaction that requires high computing power, which is why multi-core is expected to deliver a major input for technological realisation. All partners are using the same requirements engineering artefact model, which we introduced in [2] to illustrate its application in system-of-system contexts.

We elaborate two questionnaires, one on the ease of use of the model and the other one on syntactic and semantic quality of the resulting specifications. The questionnaire on ease of use contains the six criteria and semantic quality of the resulting specifications. The questionnaire on syntactic and semantic quality contains the eight criteria and statements listed under RQ 1 in Table I and the questionnaire on syntactic and semantic quality contains the eight criteria and statements listed under RQ 2 and 3, each of them assessing the ARAMiS model as well as the companiespecific reference model that was used before ARAMiS.

The Specified Systems: The least common denominator that is used as point of origin for all domain-specific activities in ARAMiS is a cyber-physical systems scenario that describes a smart travel service that accompanies a family on their journey through Europe [10].
<table>
<thead>
<tr>
<th>RQ</th>
<th>Criteria</th>
<th>Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ 1</td>
<td>Ease of use</td>
<td>The approach is clear and understandable.</td>
</tr>
<tr>
<td></td>
<td>Effectivity</td>
<td>The reference model allows modelling all contents important for a requirements spec.</td>
</tr>
<tr>
<td></td>
<td>Productivity</td>
<td>The perceived productivity was high.</td>
</tr>
<tr>
<td></td>
<td>Unambiguiousness</td>
<td>It was clear with which elements to model which kind of information.</td>
</tr>
<tr>
<td></td>
<td>Adequacy</td>
<td>The reference model provided all necessary elements to specify the requirements.</td>
</tr>
<tr>
<td></td>
<td>Reproducibility</td>
<td>The created artefacts are reproducible.</td>
</tr>
<tr>
<td>RQ 2</td>
<td>Syntactic Consistency</td>
<td>Elements in the specifications are used consistently.</td>
</tr>
<tr>
<td></td>
<td>Complexity</td>
<td>The complexity in the cross-references is low.</td>
</tr>
<tr>
<td></td>
<td>Syntactic Completeness</td>
<td>All syntactic elements needed to specify the requirements are given.</td>
</tr>
<tr>
<td></td>
<td>Syntactic Minimality</td>
<td>There are no unnecessary syntactic elements in the specifications.</td>
</tr>
<tr>
<td></td>
<td>Traceability</td>
<td>Each requirement has a rationale.</td>
</tr>
<tr>
<td></td>
<td>Ease of Perception</td>
<td>The specs are well-suited to be understood by people not involved in the process.</td>
</tr>
<tr>
<td>RQ 3</td>
<td>Unambiguity</td>
<td>The requirements are stated unambiguously.</td>
</tr>
<tr>
<td></td>
<td>Semantic Consistency</td>
<td>There are no contradictory statements in the specification.</td>
</tr>
</tbody>
</table>

On the basis of this CPS scenario, the mobility domains Automotive and Avionics selects some of the the domain-specific systems required to realise the scenario. Automotive works on intelligent infotainment and cooperative vehicles, while avionics work on cabin management systems and disaster control systems.

Figure 3 shows an excerpt of the goal model from the automotive domain, while Fig. 4 shows an excerpt of the usage model from the avionics domain.

On the basis of this CPS scenario, the mobility domains Automotive and Avionics selects some of the the domain-specific systems required to realise the scenario. Automotive works on intelligent infotainment and cooperative vehicles, while avionics work on cabin management systems and disaster control systems.

Figure 3 shows an excerpt of the goal model from the automotive domain, while Fig. 4 shows an excerpt of the usage model from the avionics domain.

Created Artefacts: Although model-based requirements engineering relies on a unified model database as a single information source, deliverables for specific milestones still have to be exported from the tool into a deliverable document. In ARAMiS, these were requirements specification documents exported from Enterprise Architect and, if necessary, polished manually. These deliverables are the basis on which the project sponsor evaluates the project’s progress.

At both companies, we refer to requirements specifications previously produced and used in same or similar contexts. However, it turned out that there was no requirements specification available of the exact same system under consideration as the ones produced with the ARAMiS artefact model.

The reason was that in the early phases of ARAMiS, the level of abstraction on which a system was specified was significantly higher than what the daily business of both partners requires. Neither of them usually specifies systems as encompassing as the whole infotainment in a vehicle or aircraft in one business unit, but instead they specify subsystems of these such as, for example, a head unit or a specific controller.

Still, the study at hand is comparable to the replication base study [7] as the situation was exactly the same — in both studies we had a previously used reference model and the resulting specification documents differ not only in their structure but also in their contents. However, having specifications that documented requirements from the same application domain, we could still validate the reference models according to the chosen evaluation criteria.

Both of the partners had been using an established reference model for their requirements engineering: While the automotive partner had been using a mostly textual representation in specification documents, the avionics partner had been using a combination of textual and SysML-based representation.

C. Analysis Results

An overview of the rating of the closed questions from our automotive partner is depicted in Fig. 5 and 7 and from our avionics partner in Fig. 6 and 8. For reasons of illustration, the centre of the Kiviat diagram is labelled with the value of zero instead of one, since otherwise the data points would overlap in the center.

Both partners seem to perceive the ARAMiS model as improvement in comparison to their previous reference model. On first glance, the rating was more critical for both their own and the ARAMiS model by the automotive partner. When analysing the answers to the open questions (the rationale for their ratings), we found that some of the reasons given for a specific rating were quite similar, but
rated rather different. For example, for traceability, both stated that their own reference model provided the possibility to link requirements, but did not provide the possibility to document a rationale. At the same time, their rating of 1 (automotive) versus 5 (avionics) of the criterion "Traceability" shows a considerable difference in perceiving the importance of a reference model providing the possibility to document a rationale.

For each of the criteria, we summarize the statements and discussion with the interviewees that were received as answers to the open questions asking for the rationale for their ratings given in Fig. 5 and Fig. 6.

1) RQ1 — Usability: Both partners perceived the usability of the ARAMI S reference model as high.

   *Ease of Use:* The automotive partner perceived their previous reference model as easier to use as the ARAMI S model. They said this was due to the higher complexity of the ARAMI S model which required an initial learning effort. However, this effort was welcomed as applying the model was easy once that initial learning curve was taken, while the company-specific reference model was very simple template. The avionics partner rated both models as clear and understandable.

   *Effectivity:* The automotive partner rated the effectivity of their previous reference model low compared to the new reference model, because it is little structured and, although
basically any text or picture can be set as entry, it does not provide the possibility to link to any models elaborated in tools. At the avionics site, our partner rated the effectivity of both models rather high, although ARAMiS relies on a model-based specification while their own model relies on textual specifications.

**Productivity:** Our automotive partner said their elaborated specifications have a higher quality and the use cases make it much easier for automotive, but there is also higher effort to be spent on elaborating the specification using the ARAMiS model. By avionics, the productivity was rated equally high for both models.

**Unambiguousness:** Our automotive partner explained they had to think about what they exactly had to model when trying to specify a specific system characteristic according to the ARAMiS model but that made their specifications better. Their own reference model distinguished only requirement and information. The unambiguousness was rated high for both models by avionics, but their own reference model provides only textual requirements and so-called design notes for additional information.

**Adequacy:** At the automotive site, our partners we generally satisfied with the model elements provided in the ARAMiS model, while for their own reference model, they stated they were missing the abstraction levels above and below their specification level, and that use case and implementation links were both missing. The avionics reference model provides for all their expected requirements types and for ARAMiS, no missing element could be identified either.

**Reproducibility:** At the automotive site, they explained that the reproducibility of artefacts for their own reference model was strongly dependent on the author. Reproducibility in both models was rated the same in avionics, but for different reasons. In ARAMiS, the reference model provides all descriptions to reproduce the type of information required while the actual reproduction also requires other information (on the actual content) and for their own reference model, the artefacts are reproducible, but the content depends on the author.

2) **RQ2 — Syntactic Quality:** Our partner at the automotive site rated the syntactic quality of specifications according to the ARAMiS reference model significantly higher than the syntactic quality according to their old reference model, but also made clear that there was simply more freedom in their own reference model. Avionics rated the ARAMiS model slightly higher than their own reference model and also pointed out that the ARAMiS reference model was more detailed in providing guidance on a higher level of abstraction, while their company-specific reference detailed more on requirements documentation in terms of phrasing.

**Syntactic Consistency:** While the automotive reference model does not indicate a specific syntax, this is prescribed by ARAMiS. Both models were perceived as leading to
consistent specifications by avionics.

Complexity: The automotive partner rated the complexity of the cross references as low for both models as long as they are added in a sensible way. At the avionics site, they rated the complexity in the cross references of specifications according to their own model as higher than in ARAMiS specifications, and stated that it was not specified how to evolve requirements and their references.

Syntactic Completeness: Generally, for automotive, all syntactic elements needed to specify the requirements are given, while the old reference model does not define them explicitly. Syntactic completeness received full scores by avionics for ARAMiS and one point less for their own model.

Syntactic Minimality: Although our partners in automotive rated the syntactic minimality high for both models, they stated that developers working on a rather technical level might point to their experience and consider some of the ARAMiS elements superfluous. In avionics, this is the only criterion where ARAMiS has scored less than the company-specific reference model, because it “has not proven so far” that there are no needless requirements in the specification.

Traceability: While ARAMiS provides traceability for the automotive partner, their own reference model lacks it. Traceability is perceived as present but not necessarily carried out in the best way in either model by avionics.

Ease of Perception: According to automotive, the ARAMiS specifications are suited to be understood by people not involved into the process as it is clear where the requirements come from. For avionics, ARAMiS specifications were well understood in external reviews.

3) RQ3 — Semantic Quality:

Unambiguity: The automotive partner rated unambiguity higher in ARAMiS than for their own model, because it requires to provide additional information with a requirement, while their own reference model requires only a single sentence. Hardly any ambiguity could be detected in specifications according to either model by avionics.

Semantic Consistency: For automotive, it is clear, which content has to be documented where, so this leads to semantic consistency if the author adheres to the guidelines. Semantic consistency was rated high for both models by avionics.

D. Comparison to the Replication Base

In comparison to our first case study conducted in an environment hosted by Siemens (our replication base), we could make the following observations:

• Both replications again showed very good ratings for the artefact-based approach in syntactic consistency and completeness and also high ratings for traceability and ease of perception. Still, the quality was rated slightly higher in the Siemens study and the ease of use slightly higher in the ARAMiS study. We attribute this fact to a simplified artefact model that was used in ARAMiS compared to a more complex and fine-grained artefact model at Siemens. Our hypothesis is that a simpler model improves ease of use while a more detailed model improves the quality of the specifications.

• The major difference we found was that in both replications the company’s former standard was not rated as low as the former standard in the Siemens study. However, we do not draw conclusions on the actual difference in quality of the respective former company standards.

E. Evaluation of Validity

For the evaluation of the validity of the study, we report on construct, internal, and external validity.

Construct Validity: Regarding construct validity, we see the threat that the used questionnaire might not completely represent the research questions. Although the questionnaire was developed in joint discussions with the participants, it cannot be ensured that there are no topics missing. In addition, the answers of the participants are inherently subjective. However, the intensive discussions of the answers and different view points of the participants increased their objectivity.

Internal Validity: The internal validity could be threatened by a bias towards either model. The “not invented here” syndrome could lead to favouring the former reference model. On the other hand, there is the threat of a bias towards the ARAMiS model as both the automotive and the avionics partner were involved in developing the ARAMiS reference model. Furthermore, one of the academic developers of the model was involved in the interviews, which might also lead to a bias towards the ARAMiS model. However, by having a co-interviewer who was not involved in the development of the model, we minimised this threat.

Finally, the assessment of the quality of the specifications was partially speculative on the basis of the ARAMiS specifications and other specifications that had been developed using the previous company-specific reference models. This is due to the fact that the requirements specifications had not been reproduced from existing ones but had been newly elaborated for different system scopes than the ones usually produced at the partners’ sites. We rely, however, on the same family of systems in the same domain whereby we do not consider this threat to be a major one.

External Validity: Regarding external validity, a major concern is, in general, the generalisability of the results. As we conduct the replications as part of a family of studies,
we come close to overcome this threat. However, we are still working towards a family of studies with a reliable database by replicating the study.

V. LESSONS LEARNT

Our lessons learnt refer mainly to the diversity of information and the usability assessments.

Diversity of Information: Different domains result in a wide range of knowledge of the individuals with different information needs. There are many specific expert areas involved with automotive and avionics, from mechanics to electronics to economics, all of which need to be accommodated for by the artefact model.

Different domains result in different challenges and priorities for creating the artefact model. For example, different roles have to find appropriate ways to exchange information across disciplines, e.g. from economics to mechanics by means of constraints.

Usability Assessment: The usability of an artefact model, especially if it is rather complex in terms of modelling concepts and their relationships, depends on the individual acceptance of the requirements engineer. Furthermore, it is necessary to conduct workshops to transfer knowledge to the engineer.

The usability as well as syntactic and semantic consistency depend on a proper realisation of the artefact model in a tool. Consistency depends on the one side on a profound understanding of the modelling concepts and on the other side on a proper realisation in the tool in terms of the possibility to interrelate the relevant concepts with each other and display these relationships.

VI. CONCLUSION

This paper presented a replication on the usability and quality of an artefact-based RE approach conducted in two industrial settings with one partner from the Automotive domain and one partner from the Avionics domain. We evaluated a domain-independent requirements engineering artefact model with regard to its usability as well as the quality of the resulting specifications in comparison to the respective former reference model used by the company.

Our investigation forms part of family of replications of a study on artefact-based requirements engineering artefact models. Traceability was also rated very high in both studies.

A. Summary of Conclusions

For most of the characteristics evaluated within this study, the results indicate that both companies have benefitted from the ARAMIIS reference model and perceive it as better compared to their previously used reference models.

In particular, regarding the usability, they rated the effectiveness to have increased same as the adequacy in the process while still supporting for precise artefacts. With regard to the quality of the created artefacts, syntactic consistency and traceability were the best rated criteria in direct comparison to the previously created specifications. The results also confirm our first expectations that the simpler artefact model improved the ease of use but constrained the improvement of the syntactic quality of the artefacts.

Our results thus strengthen our confidence in the general benefits of artefact-orientation in terms of supporting consistency and traceability while enabling a flexible process, as indicated by the results of the replication base study [7].

B. Relation to Existing Evidence

Our evaluation confirms the results reported in [7] and our expectations formulated in [11]: artefact-based requirements engineering brings benefits with regard to the usability of the requirements engineering reference process and the syntactic and semantic quality of the created artefacts.

The content model used in the study at hand was further developed after the previous study by abstracting from domain-specific details in the modelling concepts. This simplification increases its applicability across multiple domains as present in the ARAMIIS project. In consequence, and as expected, the productivity has increased while we could still preserve a high syntactic quality. Syntactic completeness was rated very high in both studies while syntactic minimality differs considerably in comparison with the respective artefact models. Traceability was also rated very high in both studies.

C. Limitations

Although this study is a replication study and forms part of a family of studies that is not yet complete, our results provide first confirmations on artefact-based RE expectations. However, the database is still limited and needs further extension for a higher generalisability of the results. Only this way we can examine whether the application of the artefact model by others not involved in the development leads to similar results.

D. Impact/Implications

From a practitioner’s point of view, the elaborated artefact models can be applied to further projects and their usage can be spread throughout the companies. From a researcher’s point of view, the artefact model can be evaluated in a different context to extend the scope and generalizability of the family of studies. Furthermore, the artefact model provides a solid basis for future improvement as indicated below.

E. Future Work

Within ARAMIIS, we are planning a further replication of this study with more and other industrial partners, probably from the micro-controller domain. Furthermore, the automotive partner has indicated that other business units
from series production might be interested in applying and evaluating the approach in their context and specific situation. The avionics partner offered his help in replicating the study in a further edition by re-elaborating a specification from series production according to the ARAMiS model to allow for comparison by an external reviewer as planned in the original study design.

Independent from ARAMiS, we will further improve and detail the model with the objective to find the ‘sweet spot’ between enough detail to enable high quality requirements specifications and not too much detail to delay requirements engineers by making them over-specify. Furthermore, we will develop enhanced tool support and evaluate it in different application domains in order to provide an artefact model for domain-independent requirements engineering. Finally, we would like to encourage the research community to participate in our empirical studies contribute to the further evaluation of artefact-based requirements engineering.

ACKNOWLEDGMENT

This work was funded within the project ARAMiS by the German Federal Ministry for Education and Research with the funding IDs 01IS11035. The responsibility for the content remains with the authors.

We would like to thank our industry partners from the ARAMiS project, in particular from BMW and Cassidian, who supported us in performing the study at hand. Furthermore, we thank Sebastian Eder for helpful comments on an earlier version of this paper.

REFERENCES


