# Fail Early, Fail Often: Towards Coherent Feedback Loops in Design Science Research Evaluation[1]

*Research-in-Progress*

**Ralf Abraham**
Institute of Information Management
University of St.Gallen
Müller-Friedberg-Strasse 8
CH-9000 St.Gallen
ralf.abraham@unisg.ch

**Stephan Aier**
Institute of Information Management
University of St.Gallen
Müller-Friedberg-Strasse 8
CH-9000 St.Gallen
stephan.aier@unisg.ch

**Robert Winter**
Institute of Information Management
University of St.Gallen
Müller-Friedberg-Strasse 8
CH-9000 St.Gallen
robert.winter@unisg.ch

## Abstract

*We propose feedback loops that increase the coherence between evaluation activities in a design science research (DSR) process. While several scholars have proposed DSR cycles with frequent evaluation activities to provide timely feedback on design activities, the question of how to ensure coherence between these activities has remained largely unaddressed. Yet, coherence is essential to claim validity not only for the DSR artifact, but also for the DSR process.*

*Based on a review of existing DSR literature, we propose an approach that ensures coherence between initial problem definition and final evaluation activities by explicating the notion of relevance underlying the DSR project, and between design and construction activities by creating situational design specifications.*

*We exemplarily apply our approach to an ongoing DSR project. We conclude with a research agenda, where we build on the recent debate on generalizability in information systems to identify six fruitful avenues for further research.*

**Keywords:** Design Science, Design method, Evaluation, Evaluation methods and criteria

[1] This work has been supported by the Swiss National Science Foundation (SNSF).

## Introduction

As a science dealing with the artificial, information systems (IS) design science research (DSR) is different from natural, social, or human sciences. While the latter primarily produce descriptive knowledge, covering natural or social phenomena, DSR produces prescriptive knowledge: knowledge on means to achieve desired ends (Gregor 2009; Gregor and Hevner 2013; Hevner et al. 2004). While natural, social, and human sciences are primarily concerned with describing reality, eventually, their results may also bear traits of prescriptive knowledge (e.g., substances that lead to new drugs). By constructing artifacts, design science researchers create reality. In other words, "the phenomenon under study emerges as the research proceeds" (Sonnenberg and vom Brocke 2012b, p. 384). Hence, the DSR results need to be evaluated differently than contributions in natural or human sciences. In this paper, we follow the notion of evaluation by Gregor and Hevner (2013): showing that something is useful, based on criteria such as validity, utility, quality, or efficacy (Gregor and Hevner 2013, p. 350). We shall use evaluation as a generic term to refer to both the design product (evaluating the artifact) and the underlying DSR process (evaluating the process).

Theories in natural or human sciences are typically validated ex post, i.e., after they have been formulated. Validation yields either corroborating evidence in favor of the theory (e.g., by showing that the theory is accurate, simple, internally and externally consistent in explaining and/or predicting natural phenomena or human behavior), or it limits the scope of the theory (Kuhn 1977, p. 321). An ex post evaluation of an artifact, i.e., showing that an artifact is actually useful in the "three realities" of real tasks, real systems, and real users (Sun and Kantor 2006), is certainly desirable in DSR as well. However, focusing solely on ex post evaluations has serious drawbacks (Sonnenberg and vom Brocke 2012b).

(1) *Resource efficiency*: Both constructing an artifact, and testing it in the real world (if possible at all), consume enormous resources. There is considerable risk of wasting resources by missing design flaws until the artifact has been deployed. Such flaws could include addressing the wrong problems (not doing the right things) or functional defects that compromise the artifact's performance (not doing things right). (2) *Re-usability*: When only the design product is evaluated, rather than the design process itself, such a process cannot be re-used with confidence. A successful artifact could also be created ad-hoc, either by following intuition or by following a process that has never been evaluated. However, such an artifact could not be deemed constructed in a rigorous way. Only when the design process itself is evaluated, it can be expected to generate high-quality, predictable results.

(3) *Epistemology*: If only the design product can be evaluated, but not the design process itself, this would question the essential value of DSR. While DSR may also produce descriptive knowledge, (e.g., during problem identification activities), the ultimate goal of DSR is to produce prescriptive knowledge in the form of artifacts that eventually need to be evaluated (Österle et al. 2011). However, to claim scientific rigor for DSR results, the design process should be justified as well: following this process will, under specified circumstances, yield certain results (Karagiannis 2010). Design theories (Gregor and Jones 2007) aim at documenting knowledge obtained in design science in a rigorous way so that design decisions can be justified and design knowledge be communicated in the academic community.

Therefore, evaluation activities in DSR need to be conducted both while the artifact is being constructed (ex ante evaluation), and after the artifact has been constructed (ex post evaluation). Ex ante evaluations assess the design process, while ex post evaluations assess the design product. Combining ex ante and ex post evaluation leads to agile DSR processes: processes that can detect design flaws early, often, and therefore cost-efficiently (a paradigm dubbed "fail early, fail often" (Thomke and Reinertsen 2012)). The need for integration instead of separation between design and evaluation activities is also discussed in related fields like action design research (Sein et al. 2011). Still, many proposed DSR processes strongly focus on ex post evaluations (Hevner et al. 2004; Peffers et al. 2007). Only the more recent contributions by Sonnenberg and vom Brocke (2012b) and Venable et al. (2012) address both ex ante and ex post evaluations. In this research-in-progress paper, we extend the focus on ex ante evaluations. This extension implies that not just a single component of the overall DSR project is evaluated (the design product), but multiple components (the actual artifacts as well as the sequence of design activities that generate the artifact and form the underlying design process). However, these evaluation activities require different techniques, and are performed from different perspectives. While several works have addressed the needs for multiple, fast evaluation iterations (Nunamaker Jr et al. 1991; Sonnenberg and vom Brocke 2012b; Venable et al. 2012; Winter and Albani 2013) and some works have addressed the need for multiple evaluation perspectives (Frank 2007; Sonnenberg and vom Brocke 2012b; Venable et al. 2012), the question how to ensure coherence between different evaluation perspectives remains largely unaddressed.

The evaluation of a DSR process is only as strong as the weakest link in the chain. A desirable validation chain should therefore not only be agile, but also ensure coherence between the different evaluation perspectives, so that the entire DSR process can be evaluated. Thus, the drawbacks of ex post-only evaluations mentioned above can be addressed. We therefore ask the research question, *how can coherence between the evaluation perspectives in an agile DSR process be ensured*? To this end, we suggest coherent notions of relevance and situational design. We also identify fruitful areas for further research (research agenda). This paper is organized as follows. In section two, we discuss related work. We formulate an initial approach for ensuring coherence among evaluation activities in section three, and we provide an example of a coherent sequence of evaluation activities in section four. We conclude our paper with a discussion of current limitations and an outlook on our future research.

## Related work

The duality of design process and design product is highlighted by Nunamaker et al. (1991). They propose a systems development process as an information systems (IS) research method. Systems development is attributed a pivotal role: the result of the process (e.g., a database system as a design product) is seen as a "proof-of-concept" for the underlying "fundamental research". Overall, systems development can be "thought of as a 'proof-by-demonstration'" (Nunamaker Jr et al. 1991, p. 91), a research approach to choose when researchers aim at providing evidence for "the validity of the solution, based on the suggested new methods, techniques, or design" (Nunamaker Jr et al. 1991, p. 98).

Frank (2007) argues for multi-perspective evaluations. He proposes four perspectives for the evaluation of an artifact: an economic perspective, a deployment perspective, an engineering perspective, and an epistemological perspective. The economic perspective assesses whether using the artifact has a desirable cost-/benefit ratio. Challenges in this perspective include quantifying benefits (e.g., how to measure to what degree an artifact improves communication among stakeholders in an organization). The deployment perspective assesses how the users react towards an artifact. Relevant criteria in this perspective include understandability, users' attitudes towards the artifact, and adoption of the artifact. The engineering perspective assesses whether a solution implements the specified requirements. The epistemological perspective, finally, assesses whether an artifact fulfills the criteria of scientific rigor. The assumptions underlying the model design should not contradict existing theories; rather, existing (design) theories should be leveraged to inform the artifact design. This perspective also requires that the artifact and the design decisions are documented in a structured way.

In a frequently referenced DSR process model, Peffers et al. (2007) outline the activities of *problem identification and motivation*, *define the objectives for a solution*, *design and development*, *demonstration*, and *evaluation*. This process is then demonstrated in a number of case studies. The emphasis on ex post evaluations becomes clear from the naming of the activities in the DSR process model: the dedicated *evaluation* activity ("[o]bserve and measure how well the artifact supports a solution to the problem" (Peffers et al. 2007, p. 56)) is carried out after the *design and development* and the prototypical instantiation in the *demonstration* activity ("Demonstrate the use of the artifact to solve one or more instances of the problem" (Peffers et al. 2007, p. 55)).

In neighboring disciplines like software development, agile development processes (Dingsøyr et al. 2010; Dyba and Dingsøyr 2008) advocate to frequently revisit earlier development activities and to perform small, incremental rather than large, big-bang changes. Thus, rapid feedback on development activities can be provided, instead of delayed feedback that is received only after a major milestone has been reached (as would be the case with sequential development processes). Similarly, in business process management (Karagiannis 1995; Weske 2007), process life cycles are also concerned with constantly monitoring and iteratively improving the underlying artifact.

## Ensuring coherence between evaluation activities

Sonnenberg and vom Brocke (2012b) argue for reconsidering the traditional sequence of build first, evaluate later (denominated "build-evaluate pattern"). Instead, they propose a cyclic DSR process that alternates between four core activities ("Problem Identification-Design-Construct-Use"), and four evaluation activities. Hence, feedback loops from each evaluation activity (Eval 1 to Eval 4) allow for revisiting the preceding design activity. Together, these feedback loops form a feedback cycle, which runs counterclockwise to the clockwise DSR cycle (Sonnenberg and vom Brocke 2012a). While the feedback cycle will typically only revisit the preceding design activity in the event of a negative evaluation result, the design cycle is intended to be completed, alternating design and evaluation activities. The various design and evaluation activities, along with the produced outputs, a selection of applicable research methods, and the proposed linkages between evaluation activities to ensure coherence are illustrated in Figure 1. Note that we based our selection of applicable research methods on the peculiarities of the IS domain; for this reason, we do not discuss methods like formal proofs or simulations that are common in computer science, but may not be adequate for DSR projects which focus not only on technology, but also on social issues like use and adoption.

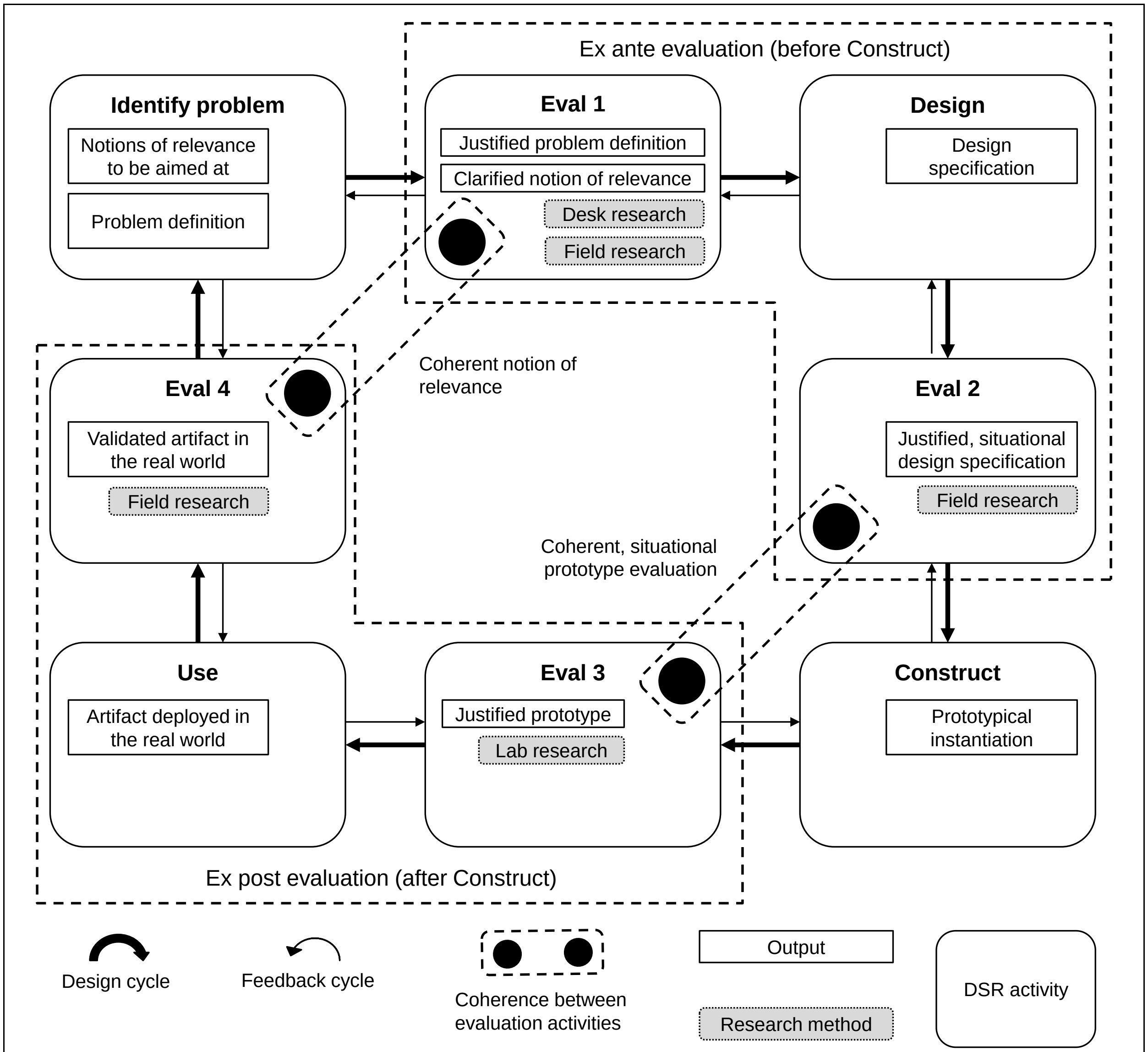


**Figure 1. Coherent design/feedback cycle based on Sonnenberg and vom Brocke (2012b)**

In the left part of Table 1, the evaluation activities (Eval 1–4) are matched to the preceding DSR process steps. For comparison, the corresponding evaluation perspectives from Peffers et al. (2007), Nunamaker et al (1991), and Frank (2007) are mapped against the Sonnenberg and vom Brocke (2012b) perspective in the right part of table 1, where applicable. While the ex post evaluation activities Eval 3 and Eval 4 (i.e., prototypical instantiations and actual use in the "three realities") are covered by all DSR process, there is a shortage on perspectives on the ex ante evaluation activities Eval 1 and especially Eval 2.

**Table 1. Comparison of evaluation perspectives**

| Sonnenberg and vom Brocke (2012b) perspective | | | Comparison perspectives: (Frank 2007; Nunamaker Jr et al. 1991; Peffers et al. 2007) | | |
|---|---|---|---|---|---|
| Evaluation activity | Ex ante/ Ex post | Evaluated DSR activity | Corresponding DSR evaluation activity based on (Peffers et al. 2007) | Systems development research process (Nunamaker Jr et al. 1991) | Evaluation perspective (Frank 2007) |
| Eval 1 | Ex ante | Problem Identification | No dedicated evaluation activity mentioned; Evaluation of the results from *problem identification and motivation* and *define the objectives for a solution* activities | n/a | Economic perspective |
| Eval 2 | Ex ante | Design | No dedicated evaluation activity mentioned; Evaluation of the results from the design part of the *design and development* activity | n/a | n/a |
| Eval 3 | Ex post | Construction | Demonstration | Proof-of-concept (prototyping) | Engineering perspective |
| Eval 4 | Ex post | Use | Evaluation | Proof-by-demonstration | Deployment perspective |

What do we mean by coherence between evaluation activities? We mean that the evaluated perspectives of the DSR process are logically connected and internally consistent. This allows designers to test certain attributes or components of the artifact where they are encountered first in the overall DSR process, instead of having to wait for the final evaluation activity (Eval 4). When designers can expect coherence between evaluation activities, they can also expect that evaluated intermediate design products lead to a useful final artifact. In cases of negative evaluation results, they only need to trace back to the preceding design activity, instead of the beginning of the overall design process. For example, when a prototype fails, designers only need to iterate back to the construction activity, but not to the design specification when they can assume coherence between the Eval 2 and Eval 3 activities.

The Eval 1 activity (evaluation of the *problem definition*) is concerned with justifying the relevance of the addressed problem. In DSR, scholars strive to demonstrate that "the envisioned design science research project is important for practice" (Sonnenberg and vom Brocke 2012b, p. 394). However, the underlying notion of relevance or what is meant by adjectives like "relevant" or "important" in a DSR project is rarely explicated. In management science, three forms of relevance are distinguished by Nicolai and Seidl (2010): instrumental, conceptual, and legitimative relevance. Instrumental relevance means building artifacts that can be used by practitioners ("instrumented") to solve problems in organizations.

Conceptual relevance includes the uncovering of causal relationships or the definition of linguistic constructs: rather than prescribing specific courses of action, the practitioners' understanding of the decision situation is enhanced. Legitimative relevance provides rhetoric devices to help managers gain legitimacy and credibility when addressing their organizations. At this stage of our research, we cannot claim one notion of relevance to be preferable in a DSR project over another. We argue, however, that DSR researchers should clearly define which notion of relevance they are planning to achieve. Depending on the selected notion of relevance, different research methods are applicable in the Eval 1 and Eval 4 activities. In the Eval 1 activity, field research methods like expert interviews and focus groups may be more appropriate when aiming at instrumental relevance, whereas desk research methods like literature reviews or deductive reasoning may be more suitable when aiming at conceptual relevance. In the Eval 4 activity, researchers should assess whether an artifact solves a specific problem (instrumental relevance) or contributes to a better understanding of the decision situation (conceptual relevance). For example, expert interviews are a suitable evaluation method for both notions of relevance, but would need to involve different sets of questions. Summarizing, DSR researchers should stick to their initially selected notion of relevance when evaluating the final artifact in the Eval 4 activity.

In the Eval 2 activity, concerning the evaluation of the *design specification*, researchers should leverage the existing knowledge base and document their rationales for including particular kernel or design theories. This results in a design specification that builds on evaluated knowledge and may also lead to increased efficiency due to re-use (Gehlert et al. 2009). The challenge of evaluating the design specification (Eval 2) is exacerbated by the fact that many artifacts in the IS domain contain a significant social component. Long-term effects must be observed, where the ends are achieved long after the means have been applied. Thus, formal proofs or simulations may not be feasible for evaluating a design specification. Those methods only address the technological aspects of a solution. Instead, the duality of technological and social aspects should also be mirrored in the creation and evaluation of the design specification. Kernel theories that provide a link between technological artifacts and their adoption in social contexts, like the technology acceptance model (Davis 1989; Venkatesh and Davis 2000; Venkatesh et al. 2003), may be useful when creating a design specification. Such a design specification may be created using desk research methods like literature reviews to collect requirements and then be evaluated against field data (which may inform the researchers whether important requirements for the solution are missing). Moreover, researchers should also attempt to identify different situations. A design specification that takes different situations into account and formulates different sets of requirements towards the artifact depending on the situation is more likely to lead to a design that fits its specific use cases. Yet, defining situations is a challenging task in many domains (e.g., software engineering). Possible approaches to differentiate situations include national (Barrett and Oborn 2010; Jaakkola et al. 2010) or organizational culture (Aier 2014; Iivari and Huisman 2007).

After a prototypical instantiation of the artifact has been created in the *construct* activity, it should be demonstrated (Eval 3). Lab research methods like experiments are suitable for this evaluation activity, since researchers can assess their artifact in some of the three realities proposed by Sun and Kantor (2006), while surrogating the others by proxies. Thus, a resource-efficient evaluation can be done when a prototype is ready (e.g., many experiments involve students as proxies for real users when demonstrating an artifact), without having to wait for a full deployment into the real world. While this clearly represents a limitation, it nevertheless provides researchers an opportunity to test a "real system" with proxy users in a proxy setting (Mettler et al. 2014).

For the final evaluation activity (Eval 4), the artifact needs to be deployed in the real world, though, and be *used* by real users to conduct real tasks. The result of this evaluation activity shows the artifact's performance (i.e., the design product) in a given situation. Field research methods like case studies seem suitable for this activity, as they allow for the observation of complex effects. When designing a case study to evaluate an artifact, design science researchers should also revisit the notion of relevance they aim at (Nicolai and Seidl 2010). For example, conceptual relevance is attributed with better decision-making by providing linguistic constructs and uncovering hitherto unknown causal relationships. By interviewing decision makers, design science researchers may assess the conceptual relevance of their artifact (which would be given if the artifact contributed to better-informed decision making). Claiming instrumental relevance would potentially require researchers to observe the impact of their artifact on its environment over a sufficient time frame. Thereby, design science researchers can assess whether their artifact gives prospective users a means to reach their desired ends (this may be particularly challenging when effects do not occur immediately, but with some delay).

## Prototypical application of the feedback loops

We now provide an example of a coherent DSR feedback cycle using an actual DSR project we are involved in. This project has recently undergone the Eval 3 activity, and we are now preparing the Eval 4 activity. The problem we identified is the generation of shared understanding in enterprise transformation (ET), i.e., fundamental, non-continuous change in organizations. Shared understanding among diverse communities of practice (e.g., line managers, project managers, or enterprise architects) on the goals and plans of an ET is considered a key success factor (Bisel and Barge 2010; Elving 1996; Ford and Ford 1995; Stensaker et al. 2008). Enterprise architecture (EA) is a promising approach to foster shared understanding in ET (Asfaw et al. 2009; Labusch and Winter 2013; Simon et al. 2013). EA provides models on current (as-is) and future (to-be) enterprise states on a high level of abstraction by modeling the interplay of business processes and supporting software systems (Aier and Saat 2011). While EA models have the potential to be used across different communities of practice, the adoption of EA models for communication in ET is still low (Asfaw et al. 2009). One way to address this challenge is to consider and design EA models as boundary objects (Valorinta 2011). Boundary objects (Star and Griesemer 1989) are abstract or physical artifacts that help to form a shared understanding among diverse communities of practice. They provide a platform for cooperative action, but without requiring the involved communities of practice to "abandon the distinctive perspectives, positions, and practices of their 'base' social world" (Winter and Butler 2011, p. 103).

Hence, the envisioned artifact is a set of design principles for constructing EA models as boundary objects. In this DSR project, we strive at both conceptual and instrumental relevance. We identify a lack of shared understanding between diverse communities of practice in enterprise transformation as a critical business problem. We offer the notion of boundary objects to describe artifacts—conceptual models from the EA domain—that help diverse communities of practice to achieve a shared understanding. By introducing the boundary object notion to the EA domain, we aim at conceptual relevance by providing a linguistic construct. Since our envisioned artifacts are design principles for the construction of boundary objects, we also aim at instrumental relevance by giving practitioners a means to shape organizational reality towards a desired end (turning an EA model into a boundary object and thereby foster shared understanding). To assess the conceptual relevance in the Eval 1 activity, we analyze the literature on both organizational change and EA, and arrive at the notion of boundary objects as a bridge between these two worlds. To assess the instrumental relevance, we conduct a series of expert interviews (Ghauri and Grønhaug 2010) to gather information on an ongoing transformation project in the air traffic management industry.

To obtain a design specification, we perform a structured literature review (Webster and Watson 2002) to identify a set of boundary object properties. We additionally conduct a focus group with enterprise architects to collect additional properties (Tremblay et al. 2010), since EA models are a prominent example of boundary objects (Pareto et al. 2010; Valorinta 2011). In the Eval 2 activity, we evaluate this design specification empirically by conducting a series of twelve expert interviews and a survey. We formulate some hypotheses describing different situations: we use the construct of knowledge boundaries to distinguish three situations of increasing distance among communities of practice (distance in terms of knowledge, values, and goals). Based on these situations, we analyze which properties from the initial set are required in which situation. Thus, we are able to perform an ex ante demonstration (Sonnenberg and vom Brocke 2012b) of the artifact: we expect a boundary object with a specific set of features to work in a specific situation. The problem we identified for our DSR project also serves as an example how deeply connected technological and social issues are in many IS research projects: in order to become a boundary object and obtain instrumental relevance for generating shared understanding, an object must not only possess certain constructional properties, but it must also be accepted and adopted by the involved communities of practice (Levina and Vaast 2005).

In the construction activity, we formulate some principles to implement a specific property ("visualization"). We reference the existing literature in the conceptual modelling field and the DSR knowledge base (specifically, the physics of notations design theory (Moody 2009)) when formulating our principles. In the demonstration activity (Eval 3), we test a subset of our artifact (principles on visualization) in an experimental setting: we describe an illustrative transformation setting (a merger of two telecommunication providers) and provide a fictitious capability model. We teach our visualization principles to doctoral candidates in the EA field. Finally, our subjects employ these principles to the fictitious model and comment on the feasibility of the principles for turning the capability map into a boundary object. In this activity, we have demonstrated a chunk of the artifact (i.e., the principles on visualization) to proxy users (i.e., students acting as proxies for enterprise architects) in a proxy setting (i.e., a fictitious model in a fictitious transformation setting). Thus, we test a part of our artifact in one of the three realities ("real systems") and approximate the remaining two ("real tasks, real users") (Sun and Kantor 2006).

The next step in our DSR project is the Eval 4 activity. We would need to deploy a complete instance of our artifacts (i.e., all our principles) in a real world setting: an organization conducting a real enterprise transformation would need to apply our design principles to one of their EA models. This model would then have to be used by members from diverse communities of practice and help them to form a shared understanding on the enterprise transformation goals and plans. A possible research method would be a single case study (Yin 2009): the unit of analysis would be a single boundary object that would have to be constructed and managed based on our design principles. By obtaining interview partners from different communities of practice (purposeful sampling strategy), the effect of the boundary object on achieving shared understanding could be observed. We can thus assess both the conceptual and the instrumental relevance of our artifact. Another reason for doing a case study in the Eval 4 activity (versus an experiment in the Eval 3 activity) is the time factor: when all three realities have to be observed to evaluate our artifact, this can only be done over an extended time period (Levina and Vaast 2005).

# Discussion

In this paper, we propose a DSR feedback cycle that does not only evaluate a DSR project from multiple perspectives, but also ensures coherence between these perspectives. We argue for ensuring coherence between the different evaluation activities by defining the notion of relevance in the problem definition activity. Based on this definition, we may then choose appropriate methods for the Eval 1 activity. The notion of relevance must also be revisited in the Eval 4 activity, when the final design product is evaluated in the real world. In other words, the Eval 1 and Eval 4 activities become coherent when the same concept of relevance is evaluated. We also argue for ensuring coherence between the design specification and prototypical instantiations. Via the definition of situations in the design specification, and prototypical instantiations of the artifact in these situations, coherence between the Eval 2 and Eval 3 activities can be achieved. By ensuring coherence between ex ante and ex post evaluation activities, DSR projects can overcome the initially stated drawbacks associated with performing an ex post evaluation only.

Revisiting the notion of "evaluation", this term might be reserved for the Eval 4 and Eval 3 activities: there, the artifact is evaluated in the real world, or demonstrated in a subset of the three realities. The ex ante activities Eval 1 and Eval 2 might possibly be called Valid 1 and Valid 2: validating a problem definition against a certain notion of relevance, and validating a design specification against certain situations (ex ante validation). Note that these activities do not involve any of the three realities. Evaluation would then arguably describe the evaluation of the design product (the artifact) in some of the three realities (Eval 3), and possibly the evaluation of the artifact in all three realities (Eval 4). A process that is reliable, reproducible, and efficient in leading to an evaluated artifact could then be called a validated design process.

# Research agenda and conclusion

Our work is *research-in-progress*. It contains several limitations, from which a research agenda can be derived. A major limitation in the cycle presented in this paper is that it does not explicate the often differing levels of abstraction of a problem description (Identify problem/Eval 1), respective solution components (Design/Eval 2), and the artefact and its instance itself (Construct/Eval 3 and Use/Eval 4). Generalizability often bears a statistically-oriented connotation (i.e., generalizing from empirical observations to underlying theory). However, in IS, other forms of generalizing have been suggested (Lee and Baskerville 2003; Lee and Baskerville 2012; Tsang and Williams 2012). This wider conceptualization of generalizability is explicitly considered to extend beyond "statistical and positivist research" and hence to be compatible with qualitative, interpretive, and design science research (Lee and Baskerville 2012, p. 759). In the following, we build on this wider conceptualization of generalizability to structure a research agenda on coherent evaluation activities in DSR. This research agenda consists of the following areas we consider to be fruitful for further research (Kuhn 1977, p. 321).

(1) Investigate additional evaluation activities before Eval 1 and after 4. An Eval 0 activity should evaluate the specialization of a generic problem class into the specific problem to be solved by the proposed artifact. An Eval 5 activity should, vice-versa, evaluate the generalization of a solution from a specific problem to a generic problem class. Depending on the artifact type, different notions of generalizability (Lee and Baskerville 2003, p. 233) are applicable. For example, when empirically derived principles are applied in another field, this implies a generalization from data to description. When a design theory is applied in another field (i.e., when its scope is extended), this implies a generalization from theory to description. Both examples represent an *exaptation* type of DSR knowledge contribution ("extend known solutions to new problems" (Gregor and Hevner 2013, p. 345)). Of particular interest are polar sampling strategies (Eisenhardt 1989; Eisenhardt and Graebner 2007), i.e., striving to increase the heterogeneity of application fields. Thus, the artifact's range of applicability, its usefulness, robustness, and eventually the users' trust in the artifact's utility can be increased.

(2) Check the transferability of the conceptual, instrumental, and legitimative notions of relevance from management science to IS DSR. This can be approached by researching which notions are applicable to which artifact types. For example, instrumental relevance can be associated with artifact types like methods, principles, or design theories, whereas conceptual relevance can be associated with descriptive artifact types like taxonomies or classification schemes. Nicolai and Seidl (2010) also suggest a trade-off relationship between instrumental and conceptual relevance. Investigating whether this is also the case in IS DSR and whether it is only associated with a certain notion of generalizability or certain artifact types, is another potentially fruitful research avenue.

(3) Ensure coherence between the Eval 2 and Eval 3 activities via situational design specifications. A generalized design specification can be transformed into a situational design specification and then instantiated in a prototype. To identify how coherence between Eval 2 and Eval 3 activities has been achieved in other studies, a literature review on DSR projects in IS and related fields like computer science or management can provide important insights. To examine different situations in artifact design, high-level concepts like organizational culture are promising. For example, Aier (2014) proposes culture-specific approaches for designing and managing EA principles using the competing values model (Denison and Spreitzer 1991). Also, the number of intermediate levels required between a generalized design specification and a specific artifact (Winter 2012) instance should be an area of future research.

(4) Identify types of failure and revisit design activities in past DSR projects. For example, which kind of failures led to revisiting which previous design steps? Did failure in the construct activity lead to revisiting only the design activity, or has the root cause been an insufficient problem description? However, data collection may be difficult, since negative results often do not get published due to publication bias (Hunter and Schmidt 1990).

(5) Incorporate affordances into a generalized design specification. Originating from perceptual psychology (Norman 1999), affordances describe properties of things that are obvious, readily available and intuitive, so that they do not require further explanation or training (e.g., the handle of a tea cup). Solution properties with affordance character would not need to be evaluated on instance level, but could (much more efficiently) be validated as part of a generalized design specification. Research on national culture can provide important insights, since affordances are likely to vary between national cultures. Affordances are a potential avenue for researching the distinction between validation and evaluation in the DSR (Gregor and Hevner 2013).

(6) Integrate with DSR process models. To increase the penetration in research and practice, the discussed evaluation perspectives should be incorporated into popular DSR process models (e.g., the DSR methodology process model (Peffers et al. 2007)).

In this research-in-progress paper, we argue for coherence between different DSR evaluation perspectives by explicating the underlying notion of relevance and by using situational design specifications. We identify fruitful areas for further research (research agenda). In particular, we call for an investigation into the different forms of generalizability in DSR design and evaluation activities.

## References

Aier, S. 2014. "The Role of Organizational Culture for Grounding, Management, Guidance and Effectiveness of Enterprise Architecture Principles," *Information Systems and e-Business Management* (12:1), pp. 43-70.

Aier, S., and Saat, J. 2011. "Understanding Processes for Model-based Enterprise Transformation Planning," *International Journal of Internet and Enterprise Management* (7:1), pp. 84–103.

Asfaw, T., Bada, A., and Allario, F. 2009. "Enablers and Challenges in Using Enterprise Architecture Concepts to Drive Transformation: Perspectives from Private Organizations and Federal Government Agencies," *Journal of Enterprise Architecture* (5:3), pp. 18-28.

Barrett, M., and Oborn, E. 2010. "Boundary object use in cross-cultural software development teams," *Human Relations* (63:8), pp. 1199-1221.

Bisel, R. S., and Barge, J. K. 2010. "Discursive positioning and planned change in organizations," *Human Relations* (64:2), pp. 257-283.

Davis, F. D. 1989. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly* (13:3), pp. 318-340.

Denison, D. R., and Spreitzer, G. M. 1991. "Organizational Culture and Organizational Development: A Competing Values Approach," *Research in Organizational Change and Development* (5), pp. 1-21.

Dingsøyr, T., Dyba, T., and Brede Moe, N. (eds.). 2010. *Agile Software Development: Current Research and Future Directions*. Berlin, Heidelberg, Germany: Springer.

Dyba, T., and Dingsøyr, T. 2008. "Empirical studies of agile software development: A systematic review," *Information and Software Technology* (50:9-10), pp. 833-859.

Eisenhardt, K. M. 1989. "Building Theories from Case Study Research," *The Academy of Management Review* (14:4), pp. 532-550.

Eisenhardt, K. M., and Graebner, M. E. 2007. "Theory Building from Cases: Opportunities and Challenges," *The Academy of Management Journal* (50:1), pp. 25-32.

Elving, W. J. L. 1996. "The role of communication in organisational change," *Corporate Communications: An International Journal* (10:2), pp. 129-138.

Ford, J. D., and Ford, L. W. 1995. "The role of conversations in producing intentional change in organizations," *The Academy of Management Review* (20:3), pp. 541-570.

Frank, U. 2007. "Evaluation of Reference Models," in *Reference Modeling for Business Systems Analysis,* P. Fettke and P. Loos (eds.). Hershey, PA, USA: Idea Group, pp. 118-140.

Gehlert, A., Schermann, M., Pohl, K., and Krcmar, H. 2009. "Towards a research method for theory driven design research," in *Wirtschaftsinformatik Proceedings 2009*, Vienna, Austria, Paper 42.

Ghauri, P., and Grønhaug, K. 2010. *Research Methods In Business Studies: A Practical Guide*. Upper Saddle River,NJ, USA: Pearson Education.

Gregor, S. 2009. "Building Theory in the Sciences of the Artificial," *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology*, Philadelphia, PA, USA: ACM, pp. 4:1-4:10.

Gregor, S., and Hevner, A. R. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," *MIS Quarterly* (37:2), pp. 337-355.

Gregor, S., and Jones, D. 2007. "The Anatomy of a Design Theory," *Journal of the Association for Information Systems* (8:5), pp. 312-335.

Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), pp. 75-105.

Hunter, J. E., and Schmidt, F. L. 1990. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Newbury Park, CA, USA: Sage Publications.

Iivari, J., and Huisman, M. 2007. "The Relationship between Organizational Culture and the Deployment of Systems Development Methodologies," *MIS Quarterly* (31:1), pp. 35-58.

Jaakkola, H., Heimbürger, A., and Linna, P. 2010. "Knowledge-oriented software engineering process in a multi-cultural context," *Software Quality Journal* (18:2), pp. 299-319.

Karagiannis, D. 1995. "BPMS: Business Process Management Systems," *ACM SIGOIS Bulletin* (16:1), pp. 10-13.

Karagiannis, D. 2010. "The role of IT in design oriented information systems research," *Memorandum on design-oriented information systems research*, August 27, 2014 (http://memo.iwi.unisg.ch/index.php?id=1&L=1#c57)

Kuhn, T. S. 1977. *The Essential Tension. Selected Studies in Scientific Tradition and Change*. Chicago, IL, USA: University of Chicago Press.

Labusch, N., and Winter, R. 2013. "Towards a Conceptualization of Architectural Support for Enterprise Transformation," in *ECIS 2013 Completed Research*, Utrecht, The Netherlands, Paper 116.

Lee, A. S., and Baskerville, R. L. 2003. "Generalizing generalizability in information systems research," *Information Systems Research* (14:3), pp. 221-243.

Lee, A. S., and Baskerville, R. L. 2012. "Conceptualizing Generalizability: New Contributions and a Reply," *MIS Quarterly* (23:3), pp. 749-761.

Levina, N., and Vaast, E. 2005. "The Emergence of Boundary Spanning Competence in Practice: Implications for Implementation and Use of Information Systems," *MIS Quarterly* (29:2), pp. 335-363.

Mettler, T., Eurich, M., and Winter, R. 2014. "On the Use of Experiments in Design Science Research: A Proposition of an Evaluation Framework," *Communications of the Association for Information Systems* (34:1), pp. 223-240.

Moody, D. L. 2009. "The “Physics” of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering," *IEEE Transactions on Software Engineering* (35:5), pp. 756-778.

Nicolai, A., and Seidl, D. 2010. "That’s Relevant! Different Forms of Practical Relevance in Management Science," *Organization Studies* (31:9-10), pp. 1257-1285.

Norman, D. A. 1999. "Affordance, Conventions, and Design," *ACM Interactions* (6:3), pp. 38-43.

Nunamaker Jr, J. F., Chen, M., and Purdin, T. D. M. 1991. "Systems Development in Information Systems Research," *Journal of Management Information Systems* (7:3), pp. 89-106.

Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Krcmar, H., Loos, P., Mertens, P., Oberweis, A., and Sinz, E. 2011. "Memorandum on design-oriented information systems research," *European Journal of Information Systems* (20:1), pp. 7-10.

Pareto, L., Eriksson, P., and Ehnebom, S. 2010. "Architectural Descriptions as Boundary Objects in System and Design Work," in *Model Driven Engineering Languages and Systems,* D.C. Petriu, N. Rouquette and Ø. Haugen (eds.). Berlin, Heidelberg, Germany: Springer, pp. 406-419.

Peffers, K., Tuunanen, T., Rothenberger, M., and Chatterjee, S. 2007. "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems* (24:3), pp. 45-77.

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., and Lindgren, R. 2011. "Action Design Research," *MIS Quarterly* (35:1), pp. 37-56.

Simon, D., Fischbach, K., and Schoder, D. 2013. "An Exploration of Enterprise Architecture Research," *Communications of the Association for Information Systems* (32:1), pp. 1-72.

Sonnenberg, C., and vom Brocke, J. 2012a. "Evaluation Patterns for Design Science Research Artefacts," in *Practical Aspects of Design Science,* M. Helfert and B. Donnellan (eds.). Berlin, Heidelberg, Germany: Springer, pp. 71-83.

Sonnenberg, C., and vom Brocke, J. 2012b. "Evaluations in the Science of the Artificial – Reconsidering the Build-Evaluate Pattern in Design Science Research," *Design Science Research in Information Systems. Advances in Theory and Practice,* K. Peffers, M. Rothenberger and B. Kuechler (eds.), Las Vegas, NV, USA: Springer, pp. 381-397.

Star, S. L., and Griesemer, J. R. 1989. "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39," *Social Studies of Science* (19:3), pp. 387-420.

Stensaker, I., Falkenberg, J., and Grønhaug, K. 2008. "Implementation Activities and Organizational Sensemaking," *The Journal of Applied Behavioral Science* (44:2), pp. 162-185.

Sun, Y., and Kantor, P. B. 2006. "Cross-Evaluation: A New Model for Information System Evaluation," *Journal of the American Society for Information Science and Technology* (57:5), pp. 614-628.

Thomke, S., and Reinertsen, D. 2012. "Six Myths Of Product Development," *Harvard Business Review* (90:5), pp. 84-94.

Tremblay, M. C., Hevner, A. R., and Berndt, D. J. 2010. "Focus Groups for Artifact Refinement and Evaluation in Design Research," *Communications of the Association for Information Systems* (26:27), pp. 599-618.

Tsang, E. W. K., and Williams, J. N. 2012. "Generalization and Induction: Misconceptions, Clarifications, and a Classification of Induction," *MIS Quarterly* (36:3), pp. 729-748.

Valorinta, M. 2011. "IT alignment and the boundaries of the IT function," *Journal of Information Technology* (26:1), pp. 46-59.
Venable, J. R., Pries-Heje, J., and Baskerville, R. L. 2012. "A Comprehensive Framework for Evaluation in Design Science Research," *Design Science Research in Information Systems. Advances in Theory and Practice,* K. Peffers, M. Rothenberger and B. Kuechler (eds.), Las Vegas, NV, USA: Springer, pp. 423-438.
Venkatesh, V., and Davis, F. D. 2000. "A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies," *Management Science* (46:2), pp. 186-204.
Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. 2003. "User Acceptance of Information Technology: Toward A Unified View," *MIS Quarterly* (27:3), pp. 425-478.
Webster, J., and Watson, R. T. 2002. "Analyzing the Past to prepare for the Future: Writing a Literature Review," *MIS Quarterly* (26:2), pp. 13-23.
Weske, M. 2007. *Business Process Management: Concepts, Languages, Architectures*. Berlin, Heidelberg, Germany: Springer.
Winter, R. 2012. "Construction of Situational Information Systems Management Methods," *International Journal of Information System Modeling and Design* (3:4), pp. 67-85.
Winter, R., and Albani, A. 2013. "Restructuring the Design Science Research Knowledge Base - A One-Cycle View of Design Science Research and its Consequences for Understanding Organizational Design Problems," in *Designing Organizational Systems: An Interdisciplinary Discourse,* R. Baskerville, M. de Marco and P. Spagnoletti (eds.). Berlin, Heidelberg, Germany: Springer, pp. 63-81.
Winter, S. J., and Butler, B. S. 2011. "Creating bigger problems: grand challenges as boundary objects and the legitimacy of the information systems field," *Journal of Information Technology* (26:2), pp. 99-108.
Yin, R. K. 2009. *Case Study Research: Design and Methods*. Los Angeles, CA, USA: Sage Publications.