**Sonnenberg, C., & vom Brocke, J. (2012). Evaluations in the Science of the Artificial - Reconsidering the Build-Evaluate Pattern in Design Science Research. In K. Peffers, M. Rothenberger & B. Kuechler (Eds.), *Design Science Research in Information Systems. Advances in Theory and Practice. Proceedings of the 7th DESRIST Conference* (Vol. 7286, pp. 381-397). Las Vegas, NV, USA: Springer Berlin / Heidelberg.**

# Evaluations in the Science of the Artificial – Reconsidering the Build-Evaluate Pattern in Design Science Research

Christian Sonnenberg and Jan vom Brocke

University of Liechtenstein, Fuerst-Franz-Josef-Strasse 21,
9490 Vaduz, Principality of Liechtenstein
{christian.sonnenberg,jan.vom.brocke}@uni.li

**Abstract.** The central outcome of design science research (DSR) is prescriptive knowledge in the form of IT artifacts and recommendations. However, prescriptive knowledge is considered to have no truth value in itself. Given this assumption, the validity of DSR outcomes can only be assessed by means of descriptive knowledge to be obtained at the conclusion of a DSR process. This is reflected in the build-evaluate pattern of current DSR methodologies. Recognizing the emergent nature of IT artifacts this build-evaluate pattern, however, poses unfavorable implications regarding the achievement of rigor within a DSR project. While it is vital in DSR to prove the usefulness of an artifact a rigorous DSR process also requires justifying and validating the artifact design itself even before it has been put into use. This paper proposes three principles for evaluating DSR artifacts which not only address the evaluation of an artifact's usefulness but also the evaluation of design decisions made to build an artifact. In particular, it is argued that by following these principles the prescriptive knowledge produced in DSR can be considered to have a truth-like value.

**Keywords:** Design science research, evaluation, design theory, epistemology

## 1 Introduction

Design science research (DSR) in information systems comprises of two primary activities: build and evaluate [1]. Although the evaluation of DSR artifacts as well as of design processes is regarded as being “crucial” [2, p. 82] much of the contemporary information system DSR work focuses on the build activity and the creation of prescriptive knowledge in the form of IT artifacts [3]. This is consistent with the view that prescriptive knowledge is the basic outcome of DSR (cf. [4], [5]). However, the prescriptive knowledge created during the build activity is assumed to have no truth-like value [5] which basically questions if such knowledge is worth to be accumulated. Moreover, if prescriptive knowledge cannot be validated until it is applied in practice a design science researcher runs the risk of devoting a significant amount of time to building insignificant solutions to practical problems.

adfa, p. 1, 2011.

This paper suggests, however, that prescriptive knowledge can have a truth-like value if DSR is conducted according to three principles. These principles relate to the problem of evaluation of DSR artifacts and spur reconsideration of the build-evaluate pattern incorporated in many current DSR methodologies. These principles are derived from the work on modes of DSR inquiries [4], on design theories [6], and on evaluation patterns for DSR artifacts [7]. The paper aims at contributing to the body of knowledge on DSR methodologies in that it tries to clarify some epistemological implications of current DSR practices. Moreover, it links existing but still not integrated and isolated contributions regarding evaluation and theorizing in DSR with the purpose of providing guidance for design science researchers to rigorously produce valid DSR artifacts.

The paper proceeds as follows. After discussing knowledge types involved in DSR as well as current DSR practices the paper points to important epistemological implications of these practices. The paper then proposes and discusses three principles to circumvent the implications of current DSR practices. The paper concludes with a summary and an outlook on future research.

## 2 Knowledge Types in DSR and Their Truth Values

IIVARI [5] made the point that design science research in IS, just like research in economics, is basically conducted at three levels of research: (1) a conceptual level, (2) a descriptive level, and (3) a prescriptive level. Research on each level creates different types of knowledge having different truth values. Conceptual knowledge captures "what things are out there" [5] in terms of concepts, constructs, conceptual frameworks, classifications, taxonomies, or typologies. Conceptual knowledge forms the foundations upon which both descriptive as well as prescriptive research build. Descriptive research is concerned with describing, understanding, and explaining 'how things are out there' [5] and produces descriptive knowledge in the form of observations, empirical regularities, theories, and hypotheses [5]. Prescriptive research yields prescriptive knowledge in the form of IT artifacts (design product knowledge) and recommendations for practice (design process knowledge) [5]. Prescriptive research is interested in answering 'how one can effectively achieve specified ends' [5].

Among the three knowledge types DSR activities predominantly focus on the creation of prescriptive knowledge (cf. [2], [4], [5]). More particular, DSR essentially aims at building artifacts that have utility for practice [2]. Statements of truth in DSR therefore relate to the fact that an artifact is actually useful or not for solving a given class of practical problems. IIVARI [5] emphasizes that prescriptive knowledge has no truth or truth-like value. Ultimately, an artifact or recommendation as prescriptive knowledge has to prove its utility in practice. This evidence, however, materializes in descriptive knowledge about an artifact. According to IIVARI [5], only descriptive knowledge, i.e. observations, empirical regularities, and theories have a truth value. As a consequence evaluations in DSR are located at the descriptive research level and are considered to not differ much from evaluations conducted in other sciences like the natural or human sciences (cf. [2], [5], [8]). However, the science of the artificial is different to other sciences in that it deals with analyzing phenomena (artifacts) that usually have not been existent at the beginning of scientific inquiry [4]. Thus, it can be challenged if evaluations in DSR should be conducted in a similar way as in the natural or human sciences. The following sections briefly outline how evaluation is considered in current DSR practices and subsequently discusses the implication of these practices with regard to achieving 'true' knowledge in DSR.

## 3 The Build-Evaluate Pattern in DSR

Although suggesting that prescriptive knowledge as the central result of DSR has no truth value, IIVARI [5] also emphasizes that prescriptive knowledge "forms an area of its own and cannot be reduced to the descriptive knowledge of theories and empirical regularities" [5, p. 56]. According to his understanding, DSR is concerned with creating prescriptive knowledge that is assumed to have no truth-like value and with gathering evidence through descriptive research that an artifact proves to be useful. Current DSR methodologies reflect this sequencing of prescriptive and descriptive research. In DSR terms, design science researchers conduct two high level activities: build and evaluate [1], [3]. A prominent example of such a DSR process is provided by PEFFERS ET AL. [9]. Their DSR methodology has been synthesized from prior DSR process proposed in the literature and is depicted in Fig. 1.

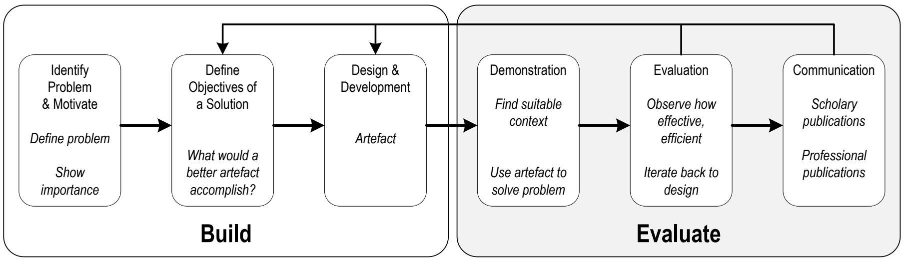


**Fig. 1.** Build-Evaluate in a representative DSR methodology (cf. [9])

What can be seen from Fig. 1 and what is also a typical assumption of other DSR processes is that evaluation activities and thus the articulation of truth statements about an artifact occur ex post, i.e. after an artifact has been constructed [3]. Truth about an artifact according to the build-evaluate pattern is known not until the evaluate phase which creates descriptive knowledge about an artifact. This applies also for DSR methodologies envisioning a concurrent or interweaved building and evaluation, like for example in Action Design Research (ADR) as proposed in [10]. Although ADR evaluation cycles appear to be much shorter when compared to a DSR process according to Fig. 1 evaluations still occur ex post, i.e. after an artifact has been constructed or revised. Thus, a validation of design decisions and the design principles incorporated by an artifact already in the design and construction phase is not a central theme in DSR evaluations. Evaluations rather focus on proving the usefulness of an artifact and less on the artifact design itself, i.e. on an artifact's rationale and specifications that are a constituent part of the prescriptive knowledge created in DSR.

In this regard it is interesting to note, however, that existing DSR methodologies emphasize the build activities, i.e. the actual artifact design, over evaluation activities [10]. This is consistent with what can also be observed in actual DSR projects. Much time is spent on designing and building an artifact, like for example when building new software systems or (re-) designing business process models. Given the significant amount of time on building an artifact and provided that the magnitude of a design decision's impact on the applicability and usefulness of an artifact is significantly higher at design-time than at run-time, i.e. when the artifact is actually constructed and instantiated (cf. [11]) it is less satisfying for a design science researcher to assume that the prescriptive knowledge holds no truth value.

It is the claim of this paper, however, that the evaluation of DSR artifacts should be approached differently compared to the study and evaluation of phenomena in the natural or human sciences. This difference emerges directly from the scope and interest of DSR which is not to explain or predict how the world is (through observations, theories, etc.) but to shape the world by means of artifacts [5]. Moreover, as GREGOR [4] points out, the truth value of DSR knowledge cannot be evaluated in terms of 'traditional' descriptive research since in DSR the researcher (or practitioner) would construct the object of study himself/herself, i.e. the phenomenon under study emerges as the research proceeds. Evaluations must account for this emergent nature and for the importance of design decisions made at the build-time of an artifact. Maintaining a 'build-evaluate'-like pattern embodied in current DSR methodologies would have significant epistemological implications on the validity of knowledge created while the artifact emerges. These implications are discussed within the next section.

## 4 Epistemological Implications of the Build-Evaluate Pattern

From a descriptive research point of view an artifact is considered to be true if some theory, observation, or empirical regularity exists that tells 'how an IT artifact actually behaves', 'why an IT artifact exists in the world', 'how an IT artifact actually relates to other things in the world' or 'if an artifact proved to be useful' (cf. [2], [5]). However, statements of truth in DSR do not primarily relate to 'what is' and 'how things are' but to 'what could and what should be' [5] and 'how useful things are expected to be'. This is consistent with the view of SIMON [8] who suggests that the sciences of the artificial "*are concerned not with the necessary but with the contingent – not with how things are but with how they might be – in short, with design*" [8, p. xii]. In this regard, GREGOR [4] argues that the study of IT artifacts by means of traditional descriptive research has to be reconsidered both in the building and the observation of IT artifacts in order to accommodate the particularities of the science of the artificial [5]. Notably, the sequencing of build and evaluate activities hardly accounts for the emergent nature of IT artifacts [10].

If DSR evaluations would be limited to descriptive knowledge it would only be possible to infer ex post if an artifact proved to be useful and why it did so. However, DSR requires IT artifacts to be built in a disciplined and "informed" way [2], [5] which necessitates making inferences on the truth contained in the prescriptive knowledge created throughout a DSR process. Therefore, it is important to infer on an artifact's expected impact on the world ex ante, i.e. before an artifact has been applied to some real world problem. A designer could refer to descriptive knowledge to justify and inform the design of a new artifact and thus ingrain descriptive truth into it. This would require the existence of kernel theories, a so called design theory, or meta-artifacts [5], [6], [12]. Nevertheless, an IT artifact emerges throughout a DSR process. The construction of an artifact precedes the knowledge of why it works [6] and thus design decisions also relate to conceptual and mainly prescriptive knowledge of an emergent design theory. These decisions have to be justified and validated by means of evaluations long before an IT artifact has been put into use.

Eventually, the assumption that the truth of an artifact cannot be inferred from prescriptive knowledge embodying an artifact's ideas, purpose, and structure ultimately affects the validity of early phases of a DSR process. If prescriptive research would result in knowledge that cannot be assumed to have truth value then no reasoning could be made about it. As a result, it can be questioned if prescriptive research could be characterized as research at all since no valid knowledge is created. Prescriptive knowledge as the major outcome of DSR would not be worth to be accumulated. Reusing parts of an artifact by other researchers of within other contexts might not be justifiable since these parts are also assumed to have no truth value. In this regard, a design science researcher would hardly be able to build an artifact in a rigorous and informed way as required by DSR guidelines [2] since design decisions could be validated not until an artifact has been constructed and applied to some reality. Some might argue that the science of the artificial would no longer be a science but rather a practice. In fact, PURAO [12] remarks that the scientific foundations underlying design research have remained largely undeveloped.

Is there a way to circumvent these epistemological implications? The key to a solution must be to acknowledge that the science of the artificial is different to the natural and human sciences and requires different modes of inquiry to reason about the truth of the knowledge created [4]. The most significant difference is that the phenomena under study cannot be assumed to be existent at the outset of a DSR endeavor but it emerges in the course of scientific inquiry. The next sections outline how an inquiry in DSR might be conducted in order to make truth-like statements about prescriptive knowledge while it emerges through design science research.

# 5 Progressing Towards a Truth – Principles for Evaluating DSR Artifacts

## 5.1 Three Principles for Evaluating DSR Artifacts

To demonstrate the validity of an artifact already in the design phase and to provide a rationale for the design decisions a design science researcher has to resort to a truth residing in conceptual and prescriptive knowledge, i.e. the ideas, metaphors, analogies, or other artifacts from which the artifact under study has been deduced. In order to make truth statements about an artifact corresponding prescriptive knowledge should be documented and accumulated in a way that allows for step-wise evaluations of an artifact as it emerges in the DSR process. In particular, such a documentation should not only allow for making inferences on the usefulness of an artifact but also on an artifact's expected suitability and importance as well as the validity and correctness of its design and construction. That means evaluations should also address the validation of incremental design decisions right from the start of a DSR process.

Prior work already pointed out that evaluation in DSR may address either the artifact design (i.e. the artifact characteristics) or the actual artifact as it is used by some relevant stakeholders. The former refers to *ex ante* evaluations occurring prior to the artifact "construction" whereas the latter refers to *ex post* evaluations after an artifact has been constructed [3]. However, ex ante evaluations in DSR are usually interpreted as a means to anticipate the effort required as well as the (economic) consequences implied by the envisioned artifact characteristics. Ex ante evaluations thus often employ complexity or profitability measures at the outset of a DSR project (cf. [3]). What has been neglected so far in ex ante evaluations is the emergent nature of IT artifacts. As has been outlined above, current DSR methodologies treat the inherent structure of an artifact, its principles of form and function, as a black box in both the build and evaluation phase. In particular, the evaluation of design decisions made by a researcher during the build phase is well out of scope of existing DSR methodologies.

It is the claim of this paper that the prescriptive knowledge that emerges throughout a DSR process has a truth-like value. This implies that incremental additions made to the prescriptive knowledge base throughout a DSR process, if evaluated and documented in a rigorous way, can be communicated early by design science researchers to interested peers or research communities. For example, a researcher could present intermediate products of a DSR process to the research community in order to build consensus on the relevance, novelty, and importance of a chosen problem domain, to discuss design objectives and features, to disseminate an initial blueprint of an IT artifact spurring joint or distinct developments of artifacts for a particular problem domain, or to demonstrate that an artifact can be put into practice by means of a prototype.

Building on prior work on DSR evaluations this paper extends the notion of ex ante evaluations by emphasizing that in order to achieve rigor in DSR it is not sufficient to just letting the IT artifact emerge in the build phase and evaluate its use but to ensure that a design science researcher makes design decisions in a disciplined way order to consistently and rigorously converge to a feasible and useful artifact. To do so it is suggested that evaluations in DSR should be conducted according to three principles. These principles have been synthesized and combined from prior literature ([4], [6], [7]) and are summarized in Table 1. It is hold that by following these principles the unfavorable epistemological implications of the build-evaluate distinction of current DSR methodologies can be alleviated.

**Table 1.** DSR evaluation principles

| Principle | Description |
|---|---|
| **Distinction** between **interior and exterior modes** of DSR inquiry | This principle directs the foci of evaluations on two aspects: (1) the constituents of the artifact and the design decisions taken as well as on (2) the evaluation of the usefulness of the artifact. |
| **Documentation** of prescriptive knowledge as **design theories** | This principle necessitates the prescriptive knowledge to be documented in a structured way. This would facilitate the communication and dissemination of the prescriptive knowledge produced within a DSR process. Moreover, such documentation would already have a truth-like value that is worth to be accumulated in a DSR knowledge base. |
| **Continuous assessment** of the DSR progress achieved through ex ante and ex post evaluations | This principle prompts the design researcher to have multiple evaluation episodes throughout a *single iteration* of a DSR process. |

These principles are interrelated in that one principle supports the other principles. Their implications on DSR evaluations are explained in detail in the following sections.

### 5.2 Distinguishing Modes of DSR Inquiry

This principle directly points to the implications of the build-evaluate pattern. DSR should not only describe and predict '"what is"' and '"why it is"' (descriptive knowledge produced in the evaluation phase). DSR predominantly builds IT artifacts producing prescriptive knowledge. The question is how a design science researcher might infer on the truth residing in that prescriptive knowledge. GREGOR [4] proposed a framework which clarifies on a high level how knowledge creation, theory building and thus truth assessment can be achieved in DSR (cf. Fig. 2).

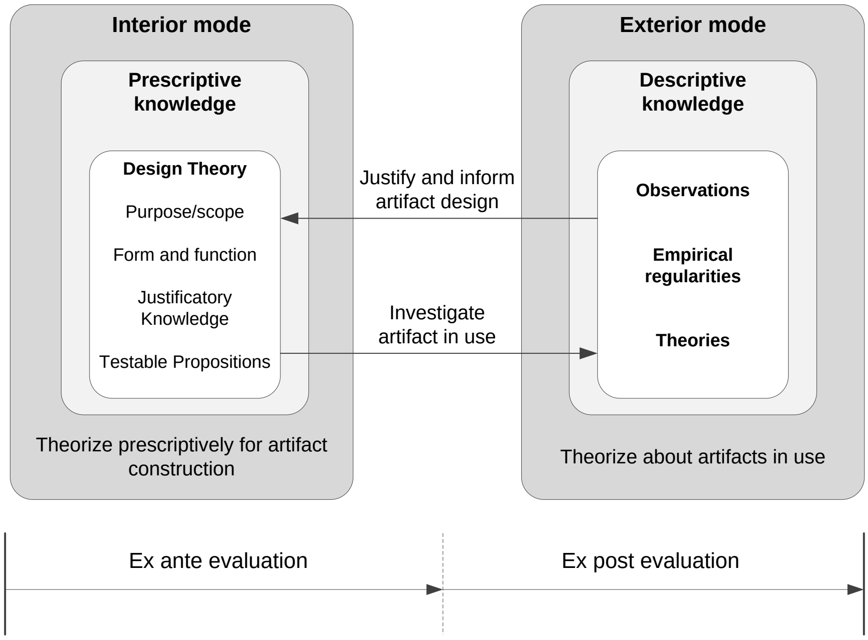


**Fig. 2.** Modes of DSR inquiry (based on [4, p. 8])

In their work [4] distinguishes two separate but linked modes of research activities that particularly affect the way artifacts should be evaluated: (1) an *interior mode* of DSR, and (2) *exterior mode* of DSR. The interior mode is concerned with producing "*prescriptive statements* about how artifacts can be designed, developed and brought into being" [4, p. 7, emphasis added]. The exterior mode aims "primarily at analyzing, *describing* and predicting what happens as artifacts exist and are used in their external environment" [4, p. 7, emphasis added]. Research in the interior mode would make use of inductive reasoning on prior descriptive or prescriptive knowledge when building an artifact. It is in this mode that prescriptive knowledge is produced. In the external mode descriptive knowledge about the artifact is produced treating the artifact more as a *black box* and only assessing significant design features with regard to achieving some utilitarian ends [4]. The relationships between interior and exterior research mode and the involved knowledge types are depicted in Fig. 2. The figure also illustrates how the application of each of the three evaluation principles stated above supports the creation of valid DSR knowledge.

In order to theorize in the interior mode, i.e. to add truth to prescriptive knowledge, a design science researcher has to document the emerging IT artifact in a way that allows for reasoning about its purpose, its rationale, its inner structure, the conditions under which the artifact is expected to work, the steps required to actually use the artifact in practice, or testable propositions that can be evaluated in the exterior mode. Such prescriptive design knowledge can be documented by means of a *design theory* [6]. The next section briefly outlines the anatomy of a design theory according to GREGOR & JONES [6] and discusses how such an anatomy supports DSR evaluations.

The distinction between interior and exterior mode not only requires design knowledge to be documented as design theories. It also widens the perspective of how evaluations in DSR should be approached. Instead of only resorting to ex post evaluations in the exterior mode (i.e. analyzing and creating descriptive knowledge), evaluations should also be conducted ex ante during the build phase as part of the interior mode. Ex ante evaluations would then refer to design theories and the progress achieved in designing an IT artifact would be assessed by means of evaluation criteria pertinent to different aspects of a design theory. This will also be discussed further below.

### 5.3 Documentation of Cumulative Prescriptive Knowledge as Design Theories

Reasoning about IT artifacts in the interior mode, i.e. its build phase, requires the design researcher to document prescriptive knowledge in a particular way. GREGOR & JONES [6] refers to such a documentation as (information systems) *design theory* (ISDT) showing "*the principles inherent in the design of an IS artifact that accomplishes some end, based on knowledge of both IT and human behavior. The ISDT allows the prescription of guidelines for further artifacts of the same type. Design theories can be about artifacts that are either products (for example, a database) or methods (for example, a prototyping methodology or an IS management strategy)*" [6, p. 322].

According to [6] a design theory consists of eight components:

1. Purpose and scope (causa finalis)
2. Constructs (causa materialis)
3. Principle of form and function (causa formalis)
4. Artifact mutability
5. Testable propositions
6. Justificatory knowledge
7. Principles of implementation (causa efficiens)
8. Expository instantiation.

Some components could be specified and reasoned about right at the outset of a DSR project, while other components are specified and reasoned about as the IT artifact emerges throughout the build phase. What can be seen, however, is that documenting artifacts according to the eight components readily serves to evaluate an artifact in terms of 'what should be' and 'how it would be able to shape the world'. Reference to descriptive knowledge and thus to exterior modes of DSR is made through components (5), (6), and (8). Testable propositions can be investigated in ex post evaluations to create descriptive knowledge about the utility of the artifact. Justificatory knowledge serves to explain or anticipate why an artifact might work in a given context and ingrains truth of prior knowledge. Justificatory knowledge can be of a descriptive (theories, observations) or of a predictive type (other design theories that proved to be useful or principles of form and functions that are reused). Expository instantiations may help to reason about an artifact's feasibility and applicability at build-time (artificial evaluation in interior mode) or to reason about its usefulness when applied to some reality (naturalistic evaluation in exterior mode). The descriptive knowledge gained by evaluating instantiations in the interior mode can serve as additional justificatory knowledge for further developing the artifact in a subsequent build cycle (e.g. benchmark results).

Documenting IT artifacts as design theories is a prerequisite for enabling the interior mode of DSR and thus to create prescriptive knowledge that ingrains truth value. Moreover, it immediately affects the way evaluations can be conducted in DSR. The distinction of interior and exterior modes of DSR together with a dedicated means for documenting the IT artifact enables the reasoning about the validity of the artifact ex ante, i.e. before it has been put into use. The predominant build-evaluate pattern of DSR methodologies along with its unfavorable epistemological implications can be reconsidered in favor of a more fine-grained consideration of research rigor in the design process. Evaluations should not only be conducted at the conclusion of a DSR project but they should be conducted on a continuing basis to assess the progress achieved as the artifact emerges [3]. In this regard, principles (1) and (2) discussed above support principle (3) leading to an expansion of the common build-evaluate pattern into a *design-evaluate-construct-evaluate* pattern (e.g. as has also been put forward in [3].

### 5.4 Continuous Assessment of the Progress Achieved in a DSR Process

By following principles (1) and (2) prescriptive knowledge in the form of design theories can be regarded as having truth-like value. Thus, it is possible and also reasonable to consider the evaluation of design decisions ingrained in the artifact and not just its usefulness by means of continuous assessments of the progress achieved in the DSR process. Two aspects are central to enable such a continuous assessment. First, *evaluation criteria* have to be defined to be able to systematically demonstrate the progress achieved in DSR and to guide evaluation activities [14]. Second, it should be clarified how ex ante and ex post evaluations can be positioned in a DSR methodology leading to the definition of *evaluation patterns* in DSR (cf. [7]).

**Evaluation Criteria**

Table 2 below lists DSR evaluation criteria proposed by [1]. These criteria could be applied in both ex ante and/or ex post evaluations. While this criteria set is considered being comprehensive [14], however, the proposed evaluation criteria are not independent of the artifact type under consideration. AIER & FISCHER [14] suggest criteria that are independent of an artifact type and particularly apply for evaluating design theories. These criteria are: utility, internal consistency, external consistency, broad purpose and scope, simplicity, fruitfulness of further research. Another set of evaluation criteria is proposed by ROSEMANN & VESSEY [15]. Their criteria set aims at particularly ensuring the relevance of a DSR artifact, i.e. if an artifact is expected to be applicable in practice. The suggested criteria are: importance, suitability, and accessibility of an artifact [15]. Applicability checks in that sense are considered particularly suitable for ex ante evaluations.

**Table 2.** Evaluation criteria for DSR artifacts (cf. [1])

| | Construct | Model | Method | Instantiation |
|---|---|---|---|---|
| Completeness | X | X | | |
| Ease of use | X | | X | |
| Effectiveness | | | | X |
| Efficiency | | | X | X |
| Elegance | X | | | |
| Fidelity with real world phenomena | | X | | |
| Generality | | | X | |
| Impact on the environment and on the artifact's users | | | | X |
| Internal consistency | | X | | |
| Level of detail | | X | | |
| Operationality | | | X | |
| Robustness | | X | | |
| Simplicity | X | | | |
| Understandability | X | | | |

Depending on the type of object to be evaluated and on the point in time an evaluation should be conducted some criteria might better reflect the progress achieved in designing an artifact then others. To structure evaluation activities and corresponding evaluation criteria the concept of evaluation patterns for DSR artifacts has been proposed in [7]. The core ideas behind these patterns as well as their specifications are presented in the next section.

**Evaluation Patterns**

Patterns are useful to describe a good solution to a recurring problem (cf. [16], cited in [17]). Patterns can be useful for both researchers and practitioners in that they incorporate "high-level solutions to classes of problems that can be converted into specific best practices" [17, p. 9]. For researchers patterns may serve to "synthesize and capture knowledge in a given domain as well as highlight areas for future research" [17, p. 9]. SONNENBERG & VOM BROCKE [7] introduced the concept of evaluation patterns for DSR artifacts. Such patterns should provide design science researchers with an orientation when configuring particular evaluation strategies. Essentially, these patterns can be positioned within a global design-evaluate-construct-evaluate pattern.

Fig. 3 below sketches a cyclic high level DSR process incorporating a design-evaluate-construct-evaluate pattern. The DSR process includes the DSR activities *problem identification*, *design*, *construction*, and *use* followed by corresponding evaluation activities. As can be seen, the process suggests that evaluations in DSR should be conducted throughout the whole process. In such a process, ex ante evaluations validate the *design of an artifact* and ex post evaluations validate *artifact instances* and *artifacts in use*. In particular, ex ante evaluations are conducted before the construction, ex post evaluations are conducted after the construction of any artifact [3].

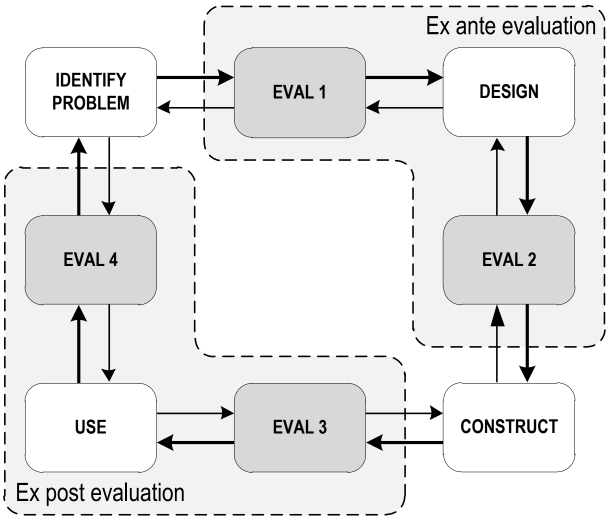


**Fig. 3.** Evaluation activities within a DSR process

The evaluation activities in Fig. 3 have been given generic names. Depending on the context and the purpose of an evaluation within the DSR process different evaluation methods and evaluation criteria could be applied for an evaluation activity [18]. Such a combination resembles ‘best practices’ in the form of evaluation patterns.

Design science researchers could benefit from such evaluation patterns as they would be able to disseminate their (validated) research findings also in early stages of their research. Ultimately, a design science researcher has to proof the utility of an artifact. However, even design objectives or principles of form and function, if related to a generic problem and evaluated rigorously might already inform other researchers and thus present a useful contribution to a DSR knowledge base.

In order to formulate such evaluation patterns it is required to broadly understand the purpose and scope of individual evaluation activities of the DSR sketched in Fig. 3. The nature of these activities as well as possible evaluation criteria and methods are summarized in Table 3 and are further discussed below. Moreover, their purpose and scope as well as their significance for supporting the accumulation of (incremental) prescriptive knowledge by means of design theories is discussed below.

**Table 3.** DSR evaluation activities and evaluation criteria

| Activity | Input | Output (mandatory) | Eval. Criteria (exemplary) | Eval. Methods (exemplary) |
|---|---|---|---|---|
| **Eval 1** | Problem statement/ Observation of a problem<br><br>Research need<br><br>Design objectives<br><br>Design theory<br><br>Existing solution to a practical problem | Justified problem statement<br><br>Justified research gap<br><br>Justified design objectives | Applicability, suitability, importance, novelty, (economic) feasibility | Literature review, review of practitioner initiatives, expert inter-view, focus groups, survey |
| **Eval 2** | Design specification<br><br>Design objectives<br><br>Stakeholders of the design specification<br><br>Design tool/ design methodology | Validated design specification<br><br>Justified design tool/ methodology | Feasibility, accessibility, understandability, clarity, simplicity, elegance, completeness, level of detail, internal consistency, ap-plicability, operationality, | Mathematical proof, logical reasoning, demonstration, simulation, benchmarking, survey, expert interview, focus group |
| **Eval 3** | Instance of an artifact (prototype) | Validated artifact instance in an *artificial setting*<br><br>(proof of applicability) | Feasibility, ease of use, effective-ness, efficiency, fidelity with real world phenome-non, operationali-ty, robustness, suitability | Demonstration with prototype, experiment with prototype, experiment with system, benchmarking, survey, expert inter-view, focus group |

| | | | | |
|---|---|---|---|---|
| **Eval 4** | Instance of an artifact | Validated artifact instance in a naturalistic setting<br><br>(proof of usefulness) | Applicability, effectiveness, efficiency, fidelity with real world phenomenon, generality, impact on artifact environment and user, internal consistency, external consistency | Case study, field experiment, survey, expert interview, focus group |

*Eval1 Activity*

The evaluation of the problem identification activity serves the purpose of ensuring that a meaningful DSR problem is selected and formulated. It should be demonstrated whether the envisioned design science research project is important for practice, is novel and thus adds to the existing knowledge base. The Eval1 activity might have different inputs depending on what actually triggers the interest in the DSR project (cf. [9]). A DSR process might start with a problem observed in practice, with a research need observed in the literature, with an existing artifact (design theory) which needs refinement in a given context, or with an existing practical solution that has not been rigorously documented or developed. Mandatory outputs of this activity are a justified problem statement, a justified research gap, and justified design objectives which serve as input for subsequent activities. Thus, the evaluation criteria and methods all serve to justify the engagement in a DSR project. Therefore, an evaluation pattern pertinent to the Eval1 activity could be termed “Justification” describing how a design researcher can justify the value of a solution and the prospective artifact. Criteria to be used here may predominantly refer to applicability checks regarding the suitability of a design idea and the perceived importance of the problem. With regard to developing an artifact, i.e. to specify a design theory, the Eval1 activity is concerned with validating the purpose and scope as well as the constructs to be used. The appropriateness of constructs might be justified by referring to constructs that have been used for solving similar problems (justificatory prescriptive knowledge). An artifact’s idea could be further validated by means of descriptive justificatory knowledge in the form of results from surveys or interviews. Moreover, a design science researcher may already derive testable propositions at this point.

*Eval2 Activity*

The evaluation of the design activity result serves the purpose of showing that an artifact design progresses to a solution of the stated problem. Since the artifact has not yet been constructed (instantiated) and thus not been applied to some reality this evaluation is artificial [19]. Possible inputs to this activity are a design specification ('blueprint', initial principles of form and function), the design objectives, information on the stakeholders of a design specification, as well as the tools and methodologies used for creating a design specification. The design specification is evaluated against its correctness and completeness to assess whether the design flawed. In particular, it should be evaluated whether the constructs used in the design specification as well as their relationships correspond to the stated design objectives. Moreover, it should be assessed whether the design specification is understandable and meaningful to all of its stakeholders (e.g. managers, IT staff) Thus, the use of particular design tools and methodologies has to be justified. Possible evaluation patterns pertinent to the validation of the design specification could be termed "demonstration" (show analytically that an artifact behaves as intended for a single test case), "simulation", or "formal proof". With regard to the justification of the design tool or methodology a pattern could be termed "tool evaluation". With regard to a design theory, the Eval2 activity validates the principles of form and function which have been specified during the design activity. Moreover, a design science researcher might want to formulate principles of implementation. Demonstrations and simulations may result in descriptive justificatory knowledge in the form of observations and empirical regularities. A formal proof may yield prescriptive justificatory knowledge in the sense that a formal proof confirms the consistency of assumptions about "what should be".

*Eval3 Activity*

This evaluation activity serves to initially demonstrate if and how well the artifact performs while interacting with organizational elements. In this activity, some inferences on the utility of an artifact could already be made. Since this activity links ex ante as well as ex post evaluations it is central for reflecting an artifact's design and stimulate subsequent iterations of the design activity if necessary (see feedback loop). The "realities" considered here may comprise of subsets of "real tasks", "real system", and "real users" (these "realities" have been suggested in [20]). Inputs to this activity are instantiations of artifacts ("constructed" artifacts) which should be evaluated regarding their applicability. At this point, the application context of the artifact instance tends to be artificial (in the sense of [19]) and might only prove that an instance is applicable to a task, within a system, or by a real user. The interplay of all three realities together with the artifact instance would be the focus of the Eval4 activity. Prototypes are frequently used at this stage. Besides demonstrating the applicability of an artifact instance, this evaluation activity should also proof that the artifact instance is consistent with its specification, i.e. that it ingrains the principles of form and function validated in the preceding evaluation activity Eval2. Possible evaluation patterns pertinent to the Eval3 activity could be termed "prototyping" and "experimentation". With regard to developing a design theory this activity is concerned with validating the component "expository instantiation" as well as artifact mutability. Moreover, evidence is gathered with regard to the ability of the artifact to behave according to its purpose and scope.

*Eval4 Activity*

This evaluation activity serves to ultimately show that an artifact is both applicable and useful in practice. Evaluations reflect the organizational context by means of all “three realities” (real tasks, real systems, and real users). Inputs to this activity are artifact instances that are fully embedded within the organizational context. Possible patterns pertinent to the Eval4 activity could be termed “case study”, “field experiment”, “survey”, or “applicability check”. With regard to design theories the main focus of the Eval4 activity would be to finally validate the artifact based on the testable propositions specified in the design theory.

# 6 Conclusions

This paper suggests reconsidering the build-evaluating pattern of current DSR methodologies in favor of a more fine grained evaluation pattern that accommodates the emerging nature of IT artifacts. Therefore, three principles for DSR evaluations have been proposed that particularly support a design science researcher to make inferences on the truth contained in the prescriptive knowledge produced by individual DSR activities.

These principles have not been invented from scratch but have been synthesized from prior literature in the field and combined to fit the purpose of this paper. However, some aspects need to be explored in more detail. In particular, the definition of a comprehensive set of evaluations patterns related to the outlined evaluation activities is expected to be particularly beneficial to better guide design science researchers and to foster the rigor and discipline of the artifact development throughout the whole DSR process. Future DSR methodologies could build on the principles put forward in this paper and verify, whether they prove to be effective.

# References


1. March, S.T., Smith, G.: Design and Natural Science Research on Information Technology. Decision Support Systems , 15 (4), 251--266 (1995)
2. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems. MIS Quarterly, 28 (1), 75--105 (2004)
3. Pries-Heje, J., Baskerville, R., Venable, J.: Strategies for Design Research Evaluation. In: 16th European Conference on Information Systems (ECIS 2008), Galway, Ireland (2008)
4. Gregor, S.: Building Theory in the Sciences of the Artificial. In Proceedings of the International Conference on Design Science Research in Information Systems and Technologies (DESRIST) 2009, Malvern, PA, (2009)
5. Iivari, J.: A Paradigmatic Analysis of Information Systems As a Design Science. Scandinavian J. Inf. Systems 19(2), (2007)
6. Gregor, S., Jones, D.: The anatomy of a design theory. Journal of the Association of Information Systems, 8 (5), Article 2, 312--335, (2007)
7. Sonnenberg, C., vom Brocke, J.: Evaluation Patterns for Design Science Research Artefacts. In: Proceedings of the European Design Science Symposium 2011. CCIS, vol. 286. Springer, Dublin, Ireland (2012)

8. Simon, H.: The sciences of the artificial, 3rd ed. MIT Press, (1996)
9. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems, 24 (3), 45--77 (2007)
10. Sein, M.K., Henfridsson, O., Purao, S., Rossi, M., Lindgren, R.: Action Design Research. MIS Quarterly, 35 (1), 37--56 (2011)
11. vom Brocke, J., Recker, J., Mendling, J.: Value-oriented process modeling: integrating financial perspectives into business process re-design. Business Process Management Journal, 16 (2), (2010)
12. Walls, J., Widmeyer, G.R., El Sawy, O.A.: Building an information system design theory for vigilant EIS, Information Systems Research, 3(1), pp. 36-59, (1992)
13. Purao, S.: Design research in technology and information systems: truth or dare," unpublished paper, School of Information Sciences and Technology, The Pennsylvania State University, University Park, State College, PA, 2002.
14. Aier, S., Fischer, C.: Criteria for Progress for Information Systems Design Theories. Information Systems and E-Business Management, 9 (1), 133--172 (2011)
15. Rosemann, M. Vessey, I.: Toward Improving the Relevance of Information Systems Research to Practice: The Role of Applicability Checks. MIS Quarterly, 32 (1), 1--22 (2008)
16. Alexander, C., Ishikawa, S., Silverstein, M.: A Pattern Language. New York, Oxford University Press, (1977).
17. Petter, S., Khazanchi, D., Murphy, J. D.: A Design Science Based Evaluation Framework for Patterns. The DATA BASE for Advances in Information Systems, 41 (3), 9--26, (2010)
18. Vaishnavi, V.K., Kuechler, W.: Improving and Innovating Information & Communication Technology: Design Science Research Methods and Patterns, Taylor Francis (2008)
19. Venable, J.: A Framework for Design Science Research Activities. In Proceedings of the 2006 Information Resource Management Association Conference. Washington, DC, USA (2006)
20. Sun, Y., Kantor, P.B.: Cross-Evaluation: A new model for information system evaluation. Journal of the American Society for Information Science and Technology, 57 (5), 614--628 (2006)