12-12-2018

Cross-Disciplinary Collaboration for Designing Data-Driven Products and Services

Christoph Kollwitz
Chemnitz University of Technology, christoph.kollwitz@wirtschaft.tu-chemnitz.de

Mximilian Perez Mengual
Fraunhofer Institute for Integrated Circuits (IIS), maximilian.perezmengual@scs.fraunhofer.de

Barbara Dinter
Chair of Information Systems, Chemnitz University of Technology, Chemnitz, Germany., barbara.dinter@wirtschaft.tu-chemnitz.de

Cross-Disciplinary Collaboration for Designing Data-Driven Products and Services

Completed Research Paper

Christoph Kollwitz
Chemnitz University of Technology
christoph.kollwitz@wirtschaft.tu-chemnitz.de

Maximilian Perez Mengual
Fraunhofer Institute for Integrated Circuits (IIS)
maximilian.perezmengual@scs.fraunhofer.de

Barbara Dinter
Chemnitz University of Technology
barbara.dinter@wirtschaft.tu-chemnitz.de

Abstract

In today's digital world, data has become an essential resource for companies' business success. However, many organizations still struggle to use this resource effectively. We argue that one major reason for this problem is the lack of data literacy of individuals. In this paper we design a boundary object, called data vignette, which is proposed to enable cross-disciplinary teams to design data-driven products and / or services by reducing knowledge boundaries. Therefore, we apply design science research (DSR) and draw back on established standards like Dublin Core Metadata Element and feedback from experts from data science as well as service designers. Following the DSR approach we present a first artificial and naturalistic evaluation. Based on the naturalistic evaluation, which we conducted at a service design jam we derive potential improvements for the design and application of the data vignette. The paper concludes with an outlook on further application scenarios and development.

Keywords

Boundary Objects, Big Data, Business Analytics, Knowledge Boundaries, Cross-disciplinary Collaboration, Data Literacy, Data-driven Products, Data-driven Services, Design Science Research.

Introduction

In today's digital age, huge amounts of poly-structured data are produced by industry as well as by the public sector, non-governmental organizations (NGO) and individuals. Since big data and the related analytics technologies emerge and become more and more sophisticated, data is considered as a key resource for business success and value creation and thus, gains increasing attention by researchers (Cavanillas et al. 2016; Demirkan et al. 2015) as well as by practitioners (Manyika et al. 2011). Not only the digital industry develops novel data-driven business models (Hartmann et al. 2016) and enriches their offerings by data related services (Davenport 2013). Overall, organizations increasingly consider data as a valuable resource for innovation by the use of analytics technologies (Duan and Cao 2015; Kiron and Ferguson 2012). This results in a constantly increasing demand for well-educated analysts and so called data scientists from literally all industry areas (Davenport and Patil 2012).

However, highly specialized data scientists who possess all the necessary skills are often hard to find, to hire and to pay for. In addition, nowadays novel services and products are no longer developed in closed R&D departments by small groups of experts, but are often the result of collaboration between a variety of actors within and outside organizations (Chesbrough 2006; von Hippel 2005). In the figurative sense, this reminds of the movie Moneyball, in which Brad Pitt, as the coach of a baseball team, prefers to sign many specialists rather than a single superstar eventually leading the team to historic success. Co-design approaches are associated with a variety of positive effects, including but not limited to an improvement of ideation, service / product quality or market success (Steen et al. 2011; Zheng et al. 2018). In order to co-design in the world of data-driven products and services, it is particularly important that the different actors have a common basic understanding of the fundamental resources available to them in the design process, which is a particularly difficult endeavor with the "new resource" data. Thus, Bhargava and D'Ignazio (2015) argue that suitable approaches and methods for the work with data are required, especially for inexperienced participants. One area of competence that is of great importance in this context is represented by the concept of data literacy. Data literacy can be defined as an individual's "ability to understand, find, collect, interpret, visualize, and support arguments using quantitative and qualitative data" (Wolff et al. 2016, p. 11). In this paper, we would like to look at data literacy from a slightly different angle. Instead of accentuating data literacy as an individual´s inherent ability to handle data, we focus on it as a foundation for the collaborative design of data-driven products and services. In other words, we aim to enable individuals with different backgrounds, abilities and perspectives to develop a common understanding of working with data thereby facilitating collaboration.

A common approach to support collaboration and heterogeneous teams lies in the use of boundary objects. The basic idea of boundary objects is to use physical or digital objects (e.g. documents, diagrams or lists) to achieve a common understanding and thus literally overcome cognitive boundaries between individuals. This principle can be illustrated by the example of business models. An employee in the sales department of a company may have a completely different understanding when he or she hears the word business model than a researcher in the field of entrepreneurship at a university or a car mechanic. In order to achieve a shared understanding of business models, the popular Business Model Canvas (cf. Osterwalder and Pigneur 2010) can be used as a boundary object. All parties can now communicate about business models by referring to the boundary object and e.g. discussing the nine areas of the Business Model Canvas.

In this paper we aim on the development of a boundary object for the design of data-driven products and services. Therefore, we postulate the following research question: How can a boundary object be designed to enable cross-disciplinary teams to co-design data-driven products and / or services?

We call this boundary object Data Vignette (DV); a kind of index cards on which important information about the content and the structure of pre-selected data sets is illustrated. They are used to get an overview of different data sets, to understand what these data sets contain, and they can be combined in an innovative manner. In other words, DVs help to capture the meaning of data sets by simplifying their semantic and structural description. This is intended to facilitate the design of new data driven services and products in cross-disciplinary teams through combining, linking and / or serializing several DVs. The DVs are particularly intended to be used in workshop settings in which teams ideate and innovate collaboratively. The developed DVs should be used for (1) the visualization of available data sets and (2) the mapping of data sets that are interesting for combination or linkage, but are not available yet. Moreover, our approach is not limited to organizational data repositories but includes also freely available data sets, so called open data, as possible resources and encourages the workshop participants to combine internal and external data.

The paper at hand is structured as follows: The motivation and problem statement are discussed in the Introduction. Subsequently, the foundations in the fields of data literacy, collaboration and boundary objects are presented in the Foundations and Theoretical Background Section. Detailed information about our methodology and the underlying research method can be found in the Research Approach Section, while the subsequent section describes the Design of the DV. Afterwards, we discuss the application and improvement of the DV in the Section Evaluation of the Artifact before the paper ends with a Conclusion.

Foundations and Theoretical Background

Data Literacy as a Foundation for the Design of Data-Driven Product and Services

The ubiquitous digital transformation leads to a growing importance of data in all industrial areas and across all business functions. Although data is considered as one of the most important resources of the 21st century and organizations recognize (big) data as valuable for the development of data-driven services

(Demirkan et al. 2015) and as the foundation for data-driven business models (Hartmann et al. 2016), it does not create value by its own, but by handling, analyzing and visualizing it (Andrade et al. 2014). At this point the role of data literacy as an important skill for designing data-driven products and services becomes evident. Götzen et al. (2018, p. 9) highlight the important of data literacy to "create [...] a higher integration of data into the design of services", while Wolff et al (2016) particularly emphasis its significance for a data driven society, where in particular open data is used for the (co-)design of apps or other marketable products. In practice, organizations usually have access to the data needed to solve problems, but employees and managers don't know how to use this data to create solutions. For example, operational personnel might not be aware of the value of customer service data or managers do not recognize the potential value of external data sources, such as open data. Obstacles like a lack of information about data quality and licenses lead to a large gap between the potential benefits of data and its actual use (Zuiderwijk and Janssen 2014), which can be at least partially attributed to missing data literacy (Sternkopf and Mueller 2017).

In today's world, the ability to access data and to appreciate its value has become a tough job for a single individual. Collaborative, creative processes have proven themselves valuable when it comes to designing products and services, but are still in infancy when data-driven products and services are involved. Appropriate approaches that enable novices to "speak data" are needed that translate data for people with different backgrounds (Bhargava and D'Ignazio 2015, p. 2). In this sense, a certain level of data literacy is a prerequisite for individuals to be able to participate in the co-design of data-driven products and services at all (Bischof et al. 2017).

Collaborative Processes and Knowledge Transfer

As mentioned before, creating and sharing knowledge on how to utilize data in order to create value is still something organizations struggle with. Therefore, a look shall be taken at numerous studies examining the research field of knowledge creation. Knowledge has become one of the main assets of organizations and serves as an important source of innovation and resulting competitive advantage (Johannessen et al. 1999; Kandampully 2002; Numprasertchai and Igel 2005). Knowledge creation is therefore one of the key processes to enable the creation of innovation. In this respect, the management of knowledge in order to foster innovation is critical for the development of new products and services (Numprasertchai and Igel 2005). The core process of innovation is facilitated by the integration of the knowledge, skills and motivation of the employees. An organization can increase its innovation capacity by transforming tacit knowledge into external knowledge through externalization and sharing (Kogut and Zander 1992; Nonaka 2008). However, transforming tacit knowledge into explicit knowledge is a significant challenge. A great deal of attention has been paid to the issue of why knowledge is difficult to manage. Especially research on the tacit nature of knowledge (Von Krogh et al. 2000; Nonaka 1994; Polanyi 1966) and its stickiness (von Hippel 1994) are to be highlighted as it illustrates mentioned difficulties.

Knowledge Boundaries

When taking the position that knowledge in organizations is problematic, especially when it comes to innovating and developing new products and services, then knowledge can be seen as a source as well as a barrier to innovation (Carlile 2002). But although the characteristics of knowledge increase problem solving capabilities within a person / function, they hinder these capabilities across functions (Carlile 2002). According to Carlile (2002), exactly these knowledge boundaries between functions represent enormous challenges for organizations regarding knowledge management. He differentiates between three approaches towards knowledge boundaries: syntactic, semantic and pragmatic. A knowledge boundary can be seen as an angle separating two functions / actors. As information becomes more and more complex, the angles diverge further (Marheineke 2016).

Spanning a syntactic boundary is basically the mere passing of information. Developed in the mathematical theory of communication by Shannon and Weaver (1963), spanning a syntactic boundary means establishing a shared and stable syntax to ensure accurate communication between actors (Shannon and Weaver 1963). Semantic boundaries mean that although a common and stable syntax exists, communication is still difficult due to interpretation when knowledge from one domain is translated to another (Marheineke 2016). Different interpretation of information by diverse actors can thus pose a serious challenge to collaboration (Carlile 2002). The semantic boundary has received a considerable amount of attention in the past. Various researchers agree that individual contextual aspects must be taken into account in the creation and transmission of knowledge (von Hippel and Tyre 1996; Nonaka 2008). Creating a mutual understanding through communication and interaction is particularly important to tackle the semantic boundary (Nonaka 1994). Even if there is a common syntax and mutual understanding, this does not mean that actors automatically collaborate with each other. So called pragmatic boundaries exist when different interests are at odds with each other and hinder collaboration (Carlile 2002). To overcome these boundaries, political and practical effort is necessary in order to transform and utilize knowledge. For this reason, the concept of boundary objects was introduced. A boundary object is a sociological concept that describes the different use of information by different groups.

Boundary Objects

The term boundary object was first defined by Susan Leigh Star and James Griesemer in 1989. The notion is that scientific work is heterogeneous and requires collaboration between different actors to be successful (Star and Griesemer 19989). In the case of collaboration, however, each actor has different needs and concerns that arise from the social world he or she inhibits. The ability of these different actors to collaborate depends on two circumstances: the development of standards and the development of boundary objects. A boundary object is part of different social worlds and enables communication between them while maintaining a different meaning in each world. Therefore, a boundary object must be abstract and as concrete as possible to ensure that they can be recognized in each social world. Star and Griesemer (1989) state:

"Boundary objects are objects which are both plastic enough to adapt to local needs and the constraints of the several parties employing them, yet robust enough to maintain a common identity across sites" (Star and Griesemer, p. 393).

Although Star and Griesemer (1989) originally saw the creation of boundary objects as a core process of communication and information coherence between overlapping social worlds, Star (Star 2010) emphasizes in a reflection on the term the importance of context. While the term boundary objects has become widespread and sees frequent use when it comes to innovation processes and prototyping (Rhinow et al. 2012), little attention is paid to the contextualization of boundary objects. Star (2010) emphasizes that boundary objects are created when different groups are required to collaborate with each other although there are differing interests. A boundary object allows collaboration between these groups without them having to agree on their interests (Star 2010). Therefore, it is important to take a process-oriented approach to boundary objects to define common and shared goals in order to ensure purposeful collaboration.

Boundary Objects for Collaboration Processes and Knowledge Integration

Boundary objects fulfill different roles when it comes to cross-disciplinary collaboration. According to Nicolini et al. (2012) three different activities are served by boundary objects: (1) they motivate collaboration, (2) they allow actors to work across different types of boundaries and (3) they represent the fundamental infrastructure of collaborative activity. Boundary objects have an influence on the understanding and exchange of ideas and are therefore enormously important for work processes. Therefore, cross-disciplinary collaboration is first and foremost a social accomplishment (Orlikowski 2002) and although boundary objects cannot solely explain cross-disciplinary work in its entirety they play an active role (Carlile 2004; Orlikowski 2007).

In the way they work, boundary objects are capable of overcoming all three kinds of knowledge boundaries. Syntactic boundary objects primarily serve the transfer of knowledge. They thus support the individual competence for knowledge-sharing between different actors and can, for example, take the form of repositories, status reporting tools or sticky notes (Marheineke 2016). Semantic boundary objects serve the exchange of meaning. Unlike pure knowledge transfer, this type of boundary object ensures that all actors have a common, shared understanding of knowledge, e.g. through the use of standardized forms (Carlile 2002). They serve the facilitation of negotiations, the discussion of information and the development of new ideas (Marheineke 2016). Pragmatic boundary objects serve the transformation of meaning. They serve to moderate constructive conflict by enabling interaction, sense-making and integration of all actors (Marheineke 2016). Examples of pragmatic boundary objects consist of malleable objects, models and maps (Carlile 2002).

The findings from the existing literature that (1) different types of knowledge boundaries exist, (2) that they can be overcome by boundary objects, and (3) that next to the boundary object itself context and process are relevant as well, will be applied in the following part of this research paper in order to develop a framework for data-driven collaboration. The boundary object to be designed should be able to address all three types of knowledge boundaries in order to (1) transfer the content of a data set and thus the knowledge (syntactic), (2) create a common understanding across all actors involved regarding the content of the data set (semantic) and (3) enable the possibility of collaboration through interaction and malleability (pragmatic).

Research Approach

While traditional research methods often focus on theory building and theory testing, thus trying to explain why a certain phenomenon is occurring, this research follows a problem-solution finding approach. An established qualitative research method in the field of information systems research is the constructivist design science research (DSR) approach. DSR strives to develop artifacts that on the one hand solve real-world problems from practice and on the other hand contribute to the body of knowledge, i.e. to theory building (Hevner 2007). In general, DSR artifacts are broadly defined as constructs, models, methods or instantiations (March and Smith 1995). Peffers et al (2007) extend this understanding by arguing that "Conceptually, a design research artifact can be any designed object in which a research contribution is embedded in the design" (p. 55). These artifacts, embedded in a natural environment, are designed to help people with problems by providing decision support or preventing problem-inducing behavior (Fuller 1992).

Since our research aims at the development of a boundary object, which provides support for people in cross-disciplinary collaboration, a design-oriented approach, such as DSR, is particularly suitable. Our approach is based on the process model introduced by Peffers et al. (2007), which consists of the six main steps (1) Problem identification & motivation, (2) Objectives of the solution, (3) Design & Development, (4) Demonstration, (5) Evaluation, and (6) Communication.

As already mentioned in the Introduction, our (1) Problem concerns the challenges of digitization and the increasing importance of data for product and service design. In particular, we focus on conveying data literacy to individuals regardless of their prior knowledge and competencies in order to enable data-driven products and service design. The (2) Objectives of our solution are mentioned in the Theoretical Background section. We aim at the development of a conceptual framework, which describes the content and structure of the DVs. We use boundary objects that represent individual data sets from organizations but also freely available open data in order to foster cross-disciplinary collaboration in product and service design. In order to be able to use the DVs in workshops, they must first be instantiated, i.e. the information describing the data sets must be depicted on the DVs. Which information is required for this and how it is displayed is the subject of the (3) Design & development. The boundary objects (DVs) basically contain two types of information: (a) information about the subject and semantic meaning of the data and, (b) information describing the structure, quality and nature of the data. Therefore, we distinguish two views on the DV:

- The (a) thematic view defines which information is relevant for describing the data sets and thus should be represented on the DV. For the design of this view we are guided by established standards for the description of information objects in the field of software development and electronic library catalogues.
- The (b) structural view represents the user's perspective on the DV. In particular the focus lies on the presentation of semantic information in a way that can be captured and processed by the users. In other words, that users have not only be provided with relevant information about the data sets, they must also be able to understand them. We use design principles and guidelines from the field of software ergonomics, visual design and usability engineering and draw on findings from research on boundary objects (cf. Section Theoretical Background).

The segmentation in views is a common procedure used in the conceptual design of information systems and software development to reduce complexity and differentiate between structural and procedural issues. Besides the two views, we aim to use and test the boundary object in various application scenarios (workshops, design jams, etc.). In order to use the DVs in such settings, they must be instantiated, i.e.

metadata have to be filled in. We use this instance as an initial (4) Demonstration of the artifact, which is shown in the Structural View Section (cf. Figure 2). During the application, we aim at collecting feedback from the participants, which is used for an initial naturalistic (5) Evaluation and improvement of the artifact.

In summary, the conceptual framework for cross-disciplinary collaboration for designing data-driven products and services, illustrated in Figure 1, results from the considerations described above. The thematic and structural view as well as the application are described more detailed in the next section.

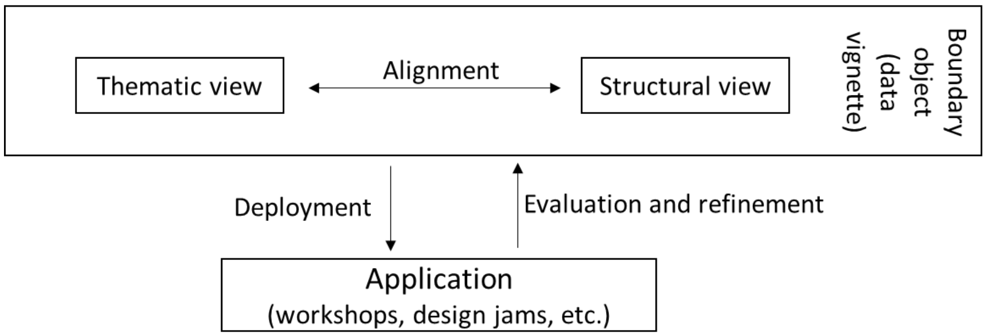

Figure 1. Conceptual Framework for Cross-Disciplinary Collaboration for Designing Data-Driven Products and Services

Design of the Data Vignette

Thematic View

To describe the meaning and structure of information objects in software development or for cataloguing, metadata is used. In general, the term metadata includes "all information which makes data useful (Bretherton and Singley 1994, p. 166)". The importance of metadata has increased enormously due to the spread of the Internet, especially Web 2.0, as metadata is used to structure the virtually unlimited amount of documents, websites and articles on the world wide web (Lee-Smeltzer 2000). There are different types of metadata. A rough distinction can be made between functional metadata addressing human users and technical metadata used directly by information systems (Bretherton and Singley 1994). Functional metadata in turn describes different aspects of the data, e.g. definitions, data quality, relationships and lineage (Foshay et al. 2007).

In the course of this development, various metadata standards from different application areas have been established. For example, ebXML represents an established standard for the process design and the exchange relationships in electronic business (ISO 15000-5 2014) or ADEPT addresses the development of distributed georeferenced information libraries (Janée and Frew 2002). These standards define, among other things, which elements are required for the description of respective information objects. Here, the elements of the Dublin Core Metadata Initiative (DCMI) have been established as a de facto standard. The so called Dublin Core elements (DCE) consist of 15 components, which are closely aligned to the standard for Information Technology -- Metadata Registries (ISO/IEC 11179-3 2013). The elements are introduced briefly in Table 1.

Name	Short description
TITLE	A name given to the resource
CREATOR	An entity primarily responsible for making the resource
SUBJECT	The topic of the resource
DESCRIPTION	An account of the resource (e.g., abstract, table od content, graphical representation)
PUBLISHER	An entity responsible for making the resource available
CONTRIBUTOR	An entity responsible for making contributions to the resource
DATE	A point or period of time associated with an event in the lifecycle of the resource
TYPE	The nature or genre of the resource
FORMAT	The file format, physical medium or dimensions of the resource
IDENTIFIER	An unambiguous reference to the resource within a given context
SOURCE	A related resource from which the described resource is derived
LANGUAGE	A language of the resource
RELATION	A related resource
COVERAGE	The spatial or temporal topic of the resource, the spatial applicability of the resource or the jurisdiction under which the resource is relevant
RIGHTS	Information about rights held in and over the resource

Table 1. Brief Description of the Dublin Core Elements (DCMI 2012)

For the design of the DV, we take the DCE as starting point, since they are widespread and generally accepted. However, we have to modify the elements for the DV for two main reasons: (1) Dublin Core was created for the description of resources in the world wide web, but we need a description of the data in the form of a boundary object. The DCE are too complex for this purpose. (2) Not all core elements are actually relevant for developing data-driven services and products. We need a reduction to the essential aspects of the data sets. Therefore, we have modified the elements in an iterative process following four steps:

- First, we have prioritized all 15 elements using the MoSCoW method, which means that we have built categories of elements we consider as “must have”, “should have”, “could have”, and “won't have” for the design of data-driven products and services.
- Second, we have conducted academic focus groups (AFGs) where we discussed the element list with scholars from the field of information systems research and service design.
- Third, with respect to our prioritization and discussion with the scholars within the AFGs, we consolidated the results.
- Fourth, we conducted a first pilot test in a service design jam (SDJ) consisting of participants who were no experts in data science or service design. We used their feedback for a first evaluation of our approach.

The AFGs took place in August and September 2018 and were conducted at the Chemnitz University of Technology and the Fraunhofer Institute for Integrated Circuits (IIS). Participants were two to five researchers with particular interests in innovation management, service design, big data analytic and computer science. The AFGs were moderated by one of the authors. Although our DVs should also be used by non-experts, we have decided to initially only include experts in the prioritization process. There are two main reasons for this: First, a basic understanding of metadata and its structure as well as service design was required for the DCE prioritization. The AFG participants possess this knowledge and have relevant experience in conducting collaborative design workshops. Second, we believe that non-experts can be better involved in the development by testing an initial draft of the DV and incorporating the resulting feedback into a next iteration. Non-experts can thus evaluate the DV in its operational use instead of carrying out abstract prioritizations.

Table 2 shows the prioritization of the single DCE by the authors, the three AFGs and the consolidated rating.

Core element	Authors priority	AFG1	AFG2	AFG3	Consolidated priority
Title	4	4	4	4	★★★★
Description	4	4	4	4	★★★★
Coverage	3	3,5	4	4	★★★★
Type	3	3	4	2	★★★
Format	2	2	4	2	★★½
Identifier	4	1	1	4	★★½
Rights	2	1	4	3	★★½
Subject	4	2	1	3	★★½
Creator	3	2	-	1	★★
Publisher	3	2	-	1	★★
Date	1	2	3	1	★★
Language	2	1	1	1,5	★½
Source	1	2	-	1	★½
Contributor	1	2	-	1	★½
Relation	1,5	1	1,5	1	★

1= Won`t ★

2=Could ★ ★

3=Should ★ ★ ★

4= Must ★ ★ ★ ★

Table 2. Prioritization of Dublin Core Elements

When consolidating the elements for the creation of the DV, we noticed that for the most elements there was consensus between the authors and the AFGs. All groups concurred that the elements Title and Description must be available. Coverage was also indicated as a must-have element. In the discussion of the AFG, Type and Format were mentioned as very similar elements. All groups agreed that Type is more important, or at least equally important to Format, since it is more general and no technical background knowledge is necessary to understand it. A more diverse picture has emerged for the elements Identifier, Rights and Subject. Although we and AFG3 considered the Identifier as must-have, AFG1 and AFG2 considered that it is redundant to Title and, therefore, unnecessary. While we and AFG1 agreed that Rights have little relevance for the design of services and products, the other two AFGs argued that Rights are a particularly important and sensitive issue (especially in Europe) and, thus, should be considered on the DVs. The element Subject was seen as rather important by us and AFG3, because the element offers the possibility to get an overview by means of upper categories. AFG1 and AFG2 argue that this could inhibit participants creativity, as upper categories prefigure assignments between data sets. Most discussions occurred regarding the elements Creator, Publisher, Contributor and Source. The main issue focussed on the question whether these elements are useful at the abstraction level of a DV or whether they can be differentiated from a user's perspective. Inspired by the discussion in the AFGs, we decided as follows: As Creator we define a contact person (if available), while as Publisher we name an organization or institution. The contact person should encourage the users to ask upcoming questions regarding the data sets, the Publisher should enable the users to better assess the reliability of a source. We dispense with the elements Contributor and Source, because the former is not considered to have any added value and the latter is deemed as redundant to Publisher. For the remaining elements Date, Language and Relation the picture is quite homogeneous again. Date is considered as already included in Coverage and otherwise as not relevant. Language is mentioned as negligible and Relation should be avoided as it is potentially confusing and restricts creativity of the users.

Structural View

While the MoSCoW prioritization has defined on the semantic level of the boundary object which information of a given data set is to be shared, the following subsection shall address the syntactic level, thus, how the DCE are displayed on the DVs in order to facilitate the transfer of meaning. For the design of the DV and the arrangement of the DCE, findings from software ergonomics, in particular eye movement analysis and visual design / usability engineering are applied. In this paper, and in the early design stage of the DV, we use existing data sets from open data repositories. The aggregation level of the data sets depicted on the DV as well as information on granularity were issues discussed in the AFGs and will be addressed over the course of the naturalistic evaluation, but are not part of this paper.

To illustrate the process of transforming the metadata into the DV an exemplary data set (shown in Table 3) is used. The exemplary data set "NCDC Storm Events Database" was accessed through the open data repository data.gov and is provided by the National Climatic Data Center (NCDC). Originally the metadata for this data set is structured in the ISO-19139 format, a metadata type specifically designed for geographic information. The metadata have therefore been slightly altered to fit the elements of the DCMI.

Core element	**Exemplary metadata**
TITLE	NCDC Storm Events Database
DESCRIPTION	Storm Data containing statistics on personal injuries and damage estimates. The data contains a chronological listing, by state, of hurricanes, tornadoes, thunderstorms, hail, floods, drought conditions, lightning, high winds, snow, temperature extremes and other weather phenomena.
COVERAGE	Storm Data covers the United States of America, recorded by county. The data began as early as 1950 through to the present, updated monthly with up to a 120 day delay possible.
TYPE	Relational data set
FORMAT	Table / CSV
IDENTIFIER	NCEI DSI 3910_03
RIGHTS	Cite data set when used as a source
SUBJECT	Atmosphere; Atmospheric phenomena; Climatic; United States of America
CREATOR	National Weather Service (NWS)
PUBLISHER	Data.gov
DATE	February 8, 2018

Table 3. Dublin Core Metadata of Exemplary Data Set

In order to determine a basic structure of the DV, it was considered how the most important information could be displayed in a way that it is caught at first sight, as not all information displayed is perceived simultaneously. Instead only small information units are perceived at a time. Through suitable graphic grouping and formatting of the information, logical information contexts can be conveyed and stressful search movements of the eyes can be reduced (Rudlof 2006). Research on attention has shown that there is a different distribution of attention when looking at screens. If a screen is divided into four squares, the quadrant on the top left will get 40% of attention on average, bottom left 25%, top right 20% and bottom right 15% of attention (Rudlof 2006). Applied to the DV, this means that we align the prioritized DCE according to the average attention distribution, placing general data before special data, and the most important data at the top left.

For the exact arrangement of the DCE on the DV, we take rules and laws from the research fields of psychology, communication design and visual design. The Gestalt psychologists, especially Max Wertheimer (1923) developed a number of laws that predict how people perceive complex stimuli and how they shape their perception. The Gestalt laws do not refer to specific content, but to the shape of the content, abstract patterns, connections and characteristics (Rudlof 2006). In particular, two laws are applied to highlight and group the individual DCE (Wertheimer 1923): (1) the law of proximity and (2) the law of similarity.

Applied to the DV and combined with the MoSCoW prioritization of the DCMI, the following design decisions result from the Gestalt laws. The law of proximity states that logically related information should be grouped together locally. Differences in prioritization are realized by spatial separation. DESCRIPTION and COVERAGE are grouped together as the most important elements that describe the actual content of a data set. TYPE, FORMAT and RIGHTS are grouped as the feature key information about how the data set can be put to immediate use. CREATOR, PUBLISHER and DATE are grouped as they offer information about the origin of the data set.

The law of similarity states that the human eye creates relationships between elements with similar design. Similarity can be developed by using elements such as common shapes and colors. With regard to the DV this phenomenon is used to develop a certain kind of category system. Thus TYPE, FORMAT and RIGHTS are represented visually stylized. On the one hand, this should help to introduce a degree of optical variance and thus increase the attractiveness and, on the other hand, simplify the perception of information. In addition, the stylized information shall facilitate the ability to make fast connections with other DVs, matching TYPE and FORMAT. Together, the Gestalt laws and findings from software ergonomics form the basis for the DV as shown in Figure 2.

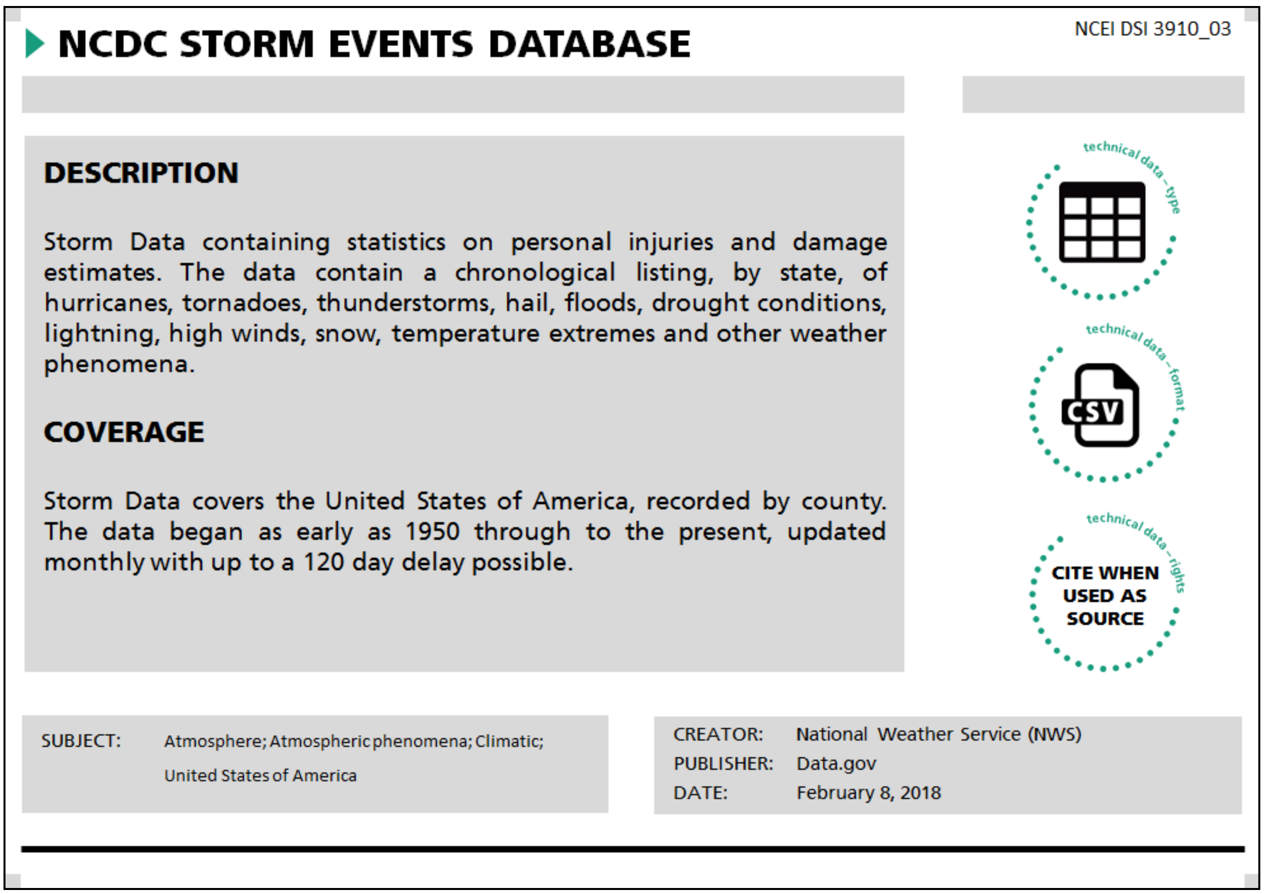

Figure 2. Data Vignette Example

It should be noted that the categorization of TYPE, FORMAT and RIGHTS is still work in progress. The challenge here is to keep the categorization as simple as possible in order to prevent overwhelming of actors with little data literacy. Further coding possibilities for the DV, e.g. through symbols and color schemes are conceivable. It needs to be avoided though that an explanatory guide of the codes has to be created in addition to DV. The symbols should therefore be as self-explanatory as possible, e.g. the TYPE can be represented by a camera symbol for photo files, a video symbol for moving images, a table symbol for relational data sets. The FORMAT, on the other hand, specifies the file format (e.g. JPEG, MPEG-4, CSV). Although the TYPE can often be derived from the FORMAT, both are given in order to avoid ambiguities. It may be the case, especially with unorganized data, that relational data is available but has been stored as a screenshot or image, thus having the TYPE "relational data" but the FORMAT "JPEG". In order to use these data sets, they would have to be processed first, thereby increasing the necessary effort. Therefore, in early data assessment phase, for which the DV is intended, both information is needed. Color schemes pose a similar challenge as most colors already have an associated characteristic, e.g. red is perceived as" alarming", green as "functioning" (Rudlof 2006). Coding through colors therefore needs careful consideration in order to prevent unwanted effects. For the mentioned reasons the assessment and development of coding possibilities will become part of the evaluation described in the next section.

Evaluation of the Artifact

Artificial Evaluation Using the Guidelines of Modelling

According to the DSR methodology, a designed artifact must be evaluated in order to show its usefulness (Peffers et al. 2007). As Venable et al. (2016) show in their "Framework for Evaluation in DSR" there are different purposes and approaches for evaluating. Thus, a DSR evaluation can either have the purpose of improving the artifact or to show its usefulness. While artificial evaluation is performed under laboratory conditions or using guidelines as well as theoretical argumentations, naturalistic evaluations are performed in the real world. For the evaluation of the DV, we first perform an argumentative and artificial evaluation to ensure that the DV has been properly designed.

Therefore, we use the "principles of orderly modeling" introduced by Beckeret al. (2000), which were originally seen as guidelines for the development of information models to counteract the problems of large projects, such as unstructuredness and complexity. While a single DV typically represents a data set that is often manageable in terms of complexity, the true value comes from combining it with other DV, i.e. data sets, requiring a structured approach. To ensure clarity, we verified the design of the DV according to six principles: (1) Correctness, (2) Relevance, (3) Economic efficiency, (4) Clarity, (5) Comparability and (6) Systematic construction.

In order to ensure the principles of (1) Correctness and (2) Relevance, we conducted three AFGs in addition to our own assessment, in which the individual DCE were discussed and prioritized. The intention of the principle of (3) Economic efficiency is to ensure that the modeling activities lie within a reasonable cost-benefit ratio. With respect to this, our main focus was on reducing the effort for instantiation and usage as much as possible. The use of the DCE facilitates this, since the required information can be derived from open data portals or metadata repositories of organizations, where DCE are frequently used for data descriptions as well. Regarding the principle of (4) Clarity, we drew on findings from the field of Gestalt psychology as illustrated in the Structural view section. With regard to the principles (5) and (6), we can identify the clearest potential for improvement by means of our boundary object. The (5) Comparability is even favored by the use of de facto standards (see DCMI) and the (6) Systematic construction is reflected in the distinction between the Thematic and the Structural view. Nevertheless, we can still improve (5) Comparability by creating a meta model based on established modeling languages and (6) Systematic construction by establishing additional views on the boundary object (e.g. an organizational view). We will realize these improvements after we have finished with our first round of evaluation.

Application of the DV - A First Pilot

The naturalistic evaluation is a still ongoing process and will be performed in several different workshop formats. A first round of evaluation was conducted as part of the Digital Festival Nuremberg in the Open Service Innovation Lab JOSEPHS® in October 2018. The overall topic of the SDJ was mobility. In particular, the needs and problems people face when planning their mobility in urban and rural areas were addressed. Eleven participants of different ages and professions took part in the two-day SDJ.

The participants were initially introduced to three mobility challenges (regarding urban mobility, rural mobility and multimodal mobility) from which they could choose one. Afterwards, they elaborated a persona in groups of up to five participants and mapped which data this persona produces on a normal day, thereby creating a "customer data journey map". For collecting these data sets, the participants used blank DVs that were supposed to be filled. Based on the persona and the self-created DVs, the participants should then design a mobility service that addresses one of the introduced challenges. As additional resources a data pool of approximately 50 DVs was given to the groups. These prefabricated DVs were based on data sets available via open data portals (e.g. data.gov.uk) and contained information on data sets from the fields of mobility & transport (e.g. accessibility of long-distance railway stations), energy & environment (e.g. CO_2 emissions from rail transport), society & social affairs (e.g. land use and population density), and technology (e.g. mobile network coverage). During the SDJ, the groups repeatedly presented the preliminary results to the other groups in order to receive feedback and to develop their service idea in an iterative way. As final result, the respective groups presented their service idea in form of an advertising prototype, a poster that illustrates approach, process and value proposition.

Although the naturalistic evaluation is still in progress and further workshops are planned, we would like to give a short overview of our first experience in the SDJ and the learnings for the refinement of the DVs. We captured the feedback of the participants in a group discussion at the end of the SDJ and through short standardized surveys. In general, the feedback of the participants on the SDJ was quite positive. The purpose of the DVs as well as the information displayed on them was easy to understand for almost all participants. However, the participants rated the usefulness of the DVs for the service design challenge as average. In the feedback discussion, a central point of criticism was that the DVs were too detailed, especially in the early phase of data collection and ideation, although it was mentioned that information on RIGHTS, TYPE and FORMAT could be quite useful in a later phase, when ideas need detailed assessment. This also coincides with our observations during the SDJ: Although the participants were zealous in writing down the data that the persona produces during the day, they concentrated mainly on the fields TITLE, DESCRIPTION and COVERAGE. Further insights refer to the use of the prefabricated DVs. On the first day, we made this data pool available from the beginning of the SDJ, so that the participants could already access the prefabricated DVs during the definition of the persona and the initial brainstorming. This was partially perceived as distracting and as inhibiting to the creativity of the participants. On the second day we made the prefabricated data pool available after the personas were determined, the “customer data journey map” was created and first ideas were discussed. An entirely different picture emerged here. The participants now started to link the additional DVs with the self-created ones and to further develop their service ideas using this additional information. The prepared DVs were modified and adapted by the participants, for example, new SUBJECTS were added to label related data.

In summary, we can derive two essential learnings that are well reflected in a quote from a participant from the feedback discussion: "sometimes less can be more". First, we learned, that especially in early design phases a simplified version of the DVs should be used, which can be extended in the later phases. In addition, the DV should be combined or embedded with other process-oriented methods, in this case the “customer data journey map” serving as an auxiliary structure for ideation. Second, it means that the provision of additional resources (such as the prefabricated DVs) for the service design process needs careful consideration and timing in order to avoid confusing the participants with information right from the beginning. With reference to data literacy however, it can be stated that the metadata-approach of the DVs has supported the workshop participants to collect, understand and consciously utilize data in the context of ideation.

Conclusion

The paper at hand serves as a starting point for a novel perspective on data-driven product and service design. In order to foster individual data literacy and collaborative work, we suggest the application of DVs, which sum up the topics and structure of individual data sets. We have defined the thematic and structural dimension of the DVs and conducted a first artificial evaluation. In addition, we piloted the DVs in the context of a SDJ. The results of this first naturalistic evaluation point to the great potential of the DVs, but at the same time show that further testing and iterative improvement is necessary. From a theoretical perspective, the paper contributes by providing first insights about the role of metadata for the development of data-driven products and services. In general, the exploitation of metadata as a central tool in collaborative design workshops represents a completely new approach that is promising for future research. From a practitioner’s point of view, our approach contributes by dismantling obstacles of service and product design in cross-disciplinary teams and by fostering data literacy of individuals. Thus, we expect the DVs to be useful tools that facilitate participation and collaboration in a digital work environment.

However, the paper also has limitations in some areas. Our artifact is still in a very early stage, so we need further evidence on which kind of information is really an indispensable factor for service design on a conceptual level and which is just a nice-to-have information. Although our learnings from the first evaluation round are in line with the expectations we derive from the prioritization of the DCE (see Table 2), they show that the DVs could be too complex, at least in the early phase of data collection and ideation. Another issue relates to the instantiation of the DVs. Currently, we do not have a suitable method to evaluate the potential benefit of data sets for service and product design, i.e. in order to decide which data sets are "worth" to be represented on a DV, a lot of manual work is required. Furthermore, we need a better understanding how we can fuse the conceptual level of the DVs with the operational creation (e.g. manufacturing or programming) of services and products.

As mentioned in the Evaluation Section, the next step for further development and refinement of our approach is to test the DVs in different empirical settings to get feedback from the users. We expect this to lead to improvements at all levels, including the content, the structure and the application of the DVs. Moreover, future research can be engaged in automating our approach, e.g., metadata can be automatically converted into DVs and curated DVs can be used as a basis for software development the other way around. Another possibility of further development is the transformation of our concept into a web application, which would allow the use of DVs over the Internet and thus, without geographical limitations.

References

Andrade, P. L., Hemerly, J., Recalde, G., and Ryan, P. S. 2014. "From Big Data to Big Social and Economic Opportunities: Which Policies Will Lead to Leveraging Data-Driven Innovation's Potential?" Global Information Technology Report 2014.

Becker, J., Rosemann, M., and von Uthmann, C. 2000. "Guidelines of Business Process Modeling," Business Process Management (1806), pp. 241–262. (https://doi.org/10.1007/3-540-45594-9_3).

Bhargava, R., and D'Ignazio, C. D. 2015. "Designing Tools and Activities for Data Literacy Learners," in Workshop on Data Literacy, pp. 1–5.

Bischof, A., Kurze, A., Storz, M., Totzauer, S., Lefeuvre, K., Jakob, S., and Berger, A. 2017. "On the Fingerprinting of Electronic Control Units Using Physical Characteristics in Controller Area Networks," in INFORMATIK 2017, M. Eibl and M. Gaedke (eds.), Chemnitz, pp. 1237–1242. (https://doi.org/10.18420/in2017).

Bretherton, F. P., and Singley, P. T. 1994. "Metadata: A User's View," in Proceedings of the Sventh International Working Conference on Scientific and Statistical Database Management, pp. 166–174.

Carlile, P. R. 2002. "A Pragmatic View of Knowledge and Boundaries: Boundary Objects in New Product Development," Organization Science (13:4), pp. 442–455.

Carlile, P. R. 2004. "Transferring, Translating, and Transforming: An Integrative Framework for Managing Knowledge Across Boundaries," Organization Science (15:5), pp. 555–568. (https://doi.org/10.1287/orsc.1040.0094).

Cavanillas, J. M., Curry, E., and Wahlster, W. 2016. "The Big Data Value Opportunity," in New Horizons for a Data-Driven Economy - A Roadmap for Usage and Exploitation of Big Data in Europe, J. M. Cavanillas, E. Curry, and W. Wahlster (eds.), Cham: Springer, pp. 3–11.

Chesbrough, H. W. 2006. "Open Innovation: A New Paradigm for Understanding Industrial Innovation," in Open Innovation: Researching a New Paradigm, H. W. Chesbrough, W. Vanhaverbeke, and J. West (eds.), Oxford, USA: Oxford University Press, pp. 1–12.

Davenport, T. H., and Patil, D. J. 2012. "Data Scientist: The Sexiest Job of the 21st Century," Harvard Business Reviewess Review, pp. 70–76.

DCMI. 2012. "Dublin Core Metadata Element Set Version 1.1." (http://dublincore.org/documents/2012/06/14/dces/, accessed November 10th, 2018).

Demirkan, H., Bess, C., Spohrer, J., Rayes, A., Allen, D., and Moghaddam, Y. 2015. "Innovations with Smart Service Systems: Analytics, Big Data, Cognitive Assistance, and the Internet of Everything," Communications of the Association for Information Systems (CAIS) (37:1), pp. 733 – 752.

Duan, Y., and Cao, G. 2015. "Understanding the Impact of Business Analytics on Innovation," in Proceedings of the 23th European Conference on Information Systems (ECIS), Münster.

Foshay, N., Mukherjee, A., and Taylor, A. 2007. "Does Data Warehouse End-User Metadata Add Value?" Communications of the ACM (50:11), pp. 70–77. (https://doi.org/10.1145/1297797.1297800).

Fuller, R. B. 1992. Cosmography: A Posthumous Scenario for the Future of Humanity, New York City: Macmillan Publishers.

Götzen, A. De, Kun, P., Simeone, L., and Morelli, N. 2018. "Making Sense of Data in a Service Design Education," in Linköping Electronic Conference Proceedings, Milan, pp. 1–10.

Hartmann, P. M., Zaki, M., Feldmann, N., and Neely, A. 2016. "Capturing Value from Big Data – a Taxonomy of Data-Driven Business Models Used by Start-up Firms," International Journal of Operations & Production Management (36:10), pp. 1382–1406.

Hevner, A. R. 2007. "A Three Cycle View of Design Science Research A Three Cycle View of Design Science Research," Scandinavian Journal of Information Systems (19:2), pp. 87–92.

von Hippel, E. 1994. "'Sticky Information'" and the Locus of Problem Solving: Implications for Innovation," Management Science (40:4), pp. 429–439.

von Hippel, E. 2005. "Democratizing Innovation: The Evolving Phenomenon of User Innovation," Journal Fur Betriebswirtschaft (55:1), pp. 63–78. (https://doi.org/10.1007/s11301-004-0002-8).

von Hippel, E., and Tyre, M. 1996. "The Mechanics of Learning by Doing: Problem Discovery during Process Machine Use," Technology and Culture (37:2), pp. 312–329.

ISO/IEC 11179-3. 2013. Information Technology--Metadata Registries (MDR)--Part 3: Registry Metamodel and Basic Attributes.

ISO 15000-5. 2014. Electronic Business Extensible Markup Language (EbXML) -- Part 5: Core Components Specification (CCS).

Janée, G., and Frew, J. 2002. "The ADEPT Digital Library Architecture," in Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL), Portland, pp. 342–350. (https://doi.org/10.1145/544220.544306).

Johannessen, J.-A., Olsen, B., and Olaisen, J. 1999. "Aspects of Innovation Theory Based on Knowledge-Management," International Journal of Information Management (19:1), pp. 121–139. (https://doi.org/10.1007/s10561-017-9623-8).

Kandampully, J. 2002. "Innovation as the Core Competency of a Service Organisation: The Role of Technology, Knowledge and Networks," European Journal of Innovation Management (5:1), pp. 18–26. (https://doi.org/10.1108/14601060210415144).

Kiron, D., and Ferguson, R. B. 2012. "Innovating with Analytics," MIT Sloan Management Review (54:1), pp. 1–8.

Kogut, B., and Zander, U. 1992. "Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology," Organization Science (3:3), pp. 383–397.

Von Krogh, G., Ichijo, K., and Nonaka, I. 2000. Enabling Knowledge Creation: How to Unlock the Mystery of Tacit Knowledge and Release the Power of Innovation, New York City: Oxford University Press.

Lee-Smeltzer, K.-H. 2000. "Finding the Needle: Controlled Vocabularies, Resource Disvoery, and Dublin Core," Library Collections, Acquisitions, & Technical Services (24:2), pp. 205–215. (https://doi.org/10.1016/S1464-9055(00)00131-7).

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A. H. 2011. "Big Data: The next Frontier for Innovation, Competition, and Productivity," McKinsey Global Institute.

March, S. T., and Smith, G. F. 1995. "Design and Natural Science Research on Information Technology," Decision Support Systems (15:4), pp. 251–266. (https://doi.org/10.1016/0167-9236(94)00041-2).

Marheineke, M. 2016. Designing Boundary Objects for Virtual Collaboration, Wiesbaden: Springer.

Nicolini, D., Mengis, J., and Swan, J. 2012. "Understanding the Role of Objects in Cross-Disciplinary Collaboration," Organization Science (23:3), pp. 612–629. (https://doi.org/10.1287/orsc.1110.0664).

Nonaka, I. 1994. "A Dynamic Theory of Organizational Knowledge Creation," Organization Science (5:1), pp. 14–37. (https://doi.org/10.1287/orsc.5.1.14).

Nonaka, I. 2008. The Knowledge-Creating Company, Boston: Harvard Business Review Press.

Numprasertchai, S., and Igel, B. 2005. "Managing Knowledge through Collaboration: Multiple Case Studies of Managing Research in University Laboratories in Thailand," Technovation (25:10), pp. 1173–1182. (https://doi.org/10.1016/j.technovation.2004.03.001).

Orlikowski, W. J. 2002. "Knowing in Practice: Enacting a Collective Capability in Distributed Organizing," Organization Science (13:3), pp. 249–273. (https://doi.org/10.1287/orsc.13.3.249.2776).

Orlikowski, W. J. 2007. "Sociomaterial Practices: Exploring Technology at Work," Organization Studies (28:9), pp. 1435–1448. (https://doi.org/10.1177/0170840607081138).

Osterwalder, A., and Pigneur, Y. 2010. Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers, Hoboken: John Wiley & Sons.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. 2007. "A Design Science Research Methodology for Information Systems Research," Journal of Management Information Systems (24:3), pp. 45–77. (https://doi.org/10.2753/MIS0742-1222240302).

Polanyi, M. 1966. "The Logic of Tacit Inference," Philosophy (41:155), pp. 1–18.

Rhinow, H., Köppen, E., and Meinel, C. 2012. "Design Prototypes as Boundary Objects in Innovation Processes," in Proceedings of the Internatonal Conference on Design Research Society, Bangkok, pp. 1–10.

Rudlof, C. 2006. "Handbuch Software-Ergonomie," Usability Engineering.

Shannon, C. E., and Weaver, W. 1963. The Mathematical Theory of Communication, Urbana: University of Illinois Press.

Star, S. L. 2010. "This Is Not a Boundary Object: Reflections on the Origin of a Concept," Science Technology and Human Values (35:5), pp. 601–617. (https://doi.org/10.1177/0162243910377624).

Star, S. L., and Griesemer, J. R. 19989. "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology," Social Studies of Science (19:3), pp. 387–420.

Steen, M., Manschot, M., and Koning, N. De. 2011. "Benefits of Co-Design in Service Design Projects Marc," International Journal of Design (5:2), pp. 53–60.

Sternkopf, H., and Mueller, R. M. 2017. "Doing Good with Data: Development of a Maturity Model for Data Literacy in Non-Governmental Organizations Abstract," in Proceedings of the 51st Hawaii International Conference on System Sciences (Vol. 9), Waikoloa, pp. 5045–5054. (https://doi.org/10.24251/HICSS.2018.630).

Venable, J., Pries-Heje, J., and Baskerville, R. 2016. "FEDS: A Framework for Evaluation in Design Science Research," European Journal of Information Systems (25:1), pp. 77–89.

Wertheimer, M. 1923. "Untersuchungen Zur Lehre von Der Gestalt. II," Psychologische Forschung (4:1), pp. 301–350. (https://doi.org/10.1007/BF00410640).

Wolff, A., Gooch, D., Montaner, C., Rashid, J. J., Kortuem, U., Wolff, A., Gooch, D., Montaner, J. J. C., Rashid, U., and Kortuem, G. 2016. "Creating an Understanding of Data Literacy for a Data-Driven Society," The Journal of Community Informatics (12:3), pp. 9–26.

Zheng, P., Xu, X., and Chen, C. H. 2018. "A Data-Driven Cyber-Physical Approach for Personalised Smart, Connected Product Co-Development in a Cloud-Based Environment," Journal of Intelligent Manufacturing, pp. 1–16. (https://doi.org/10.1007/s10845-018-1430-y).

Zuiderwijk, A., and Janssen, M. 2014. "Barriers and Development Directions for the Publication and Usage of Open Data: A Socio-Technical View," in Open Government: Opportunities and Challenges for Public Governance, M. Gascó-Hernández (ed.), New York: Springer, pp. 115–135.