Winter 12-2019

Data Collection Map: A Canvas for Shared Data Awareness in Data-Driven Innovation Projects

Liza Kayser

Roland M. Mueller

Tizian Kronsbein

Data Collection Map: A Canvas for Shared Data Awareness in Data-Driven Innovation Projects

Completed Research Paper

Liza Kayser
University of Twente
l.j.j.kayser@utwente.nl

Roland M. Mueller
Berlin School of Economics and Law
University of Twente
roland.mueller@hwr-berlin.de

Tizian Kronsbein
Dlighted
tizian@dlighted.de

Abstract

Digitization, Machine Learning, and Artificial Intelligence along with Big Data lead to more and more data-driven innovations. However, companies often lack the knowledge, skills, and experience about the potentials of data and its related technologies. Consequently, they might struggle to manage the related complexity when designing data-driven products, services, or processes. This paper presents the Data Collection Map—a visual collaboration canvas that can be used to help raising data awareness and identifying the key data stakeholders. Action design research was applied to develop and test the canvas through iterative cycles of building, intervention, and evaluation. Three workshops were conducted to collect and triangulate data from focus group discussions, surveys, and observations. The resulting canvas facilitates shared knowledge and understanding and hence, helps organizations to raise awareness of the (potentially) available data assets in innovation projects.

Keywords

Data awareness, data literacy, data-driven innovation, visual collaborative tool.

Introduction

Due to emerging technological trends of current digitization efforts, such as Machine Learning and Artificial Intelligence (AI), organizations are urged to identify and use the potentials of data and corresponding technologies. Therefore, more and more organizations try to identify relevant use cases, applications, or business models for taking advantage of key technologies around data and analytics (Manyika 2017; Newman 2018; Pettey 2019). Digitization also entails that more and more data is being generated (e.g. by products, services and processes), collected, and made accessible to organizations (Burn-Murdoch 2012; Cave 2017; Manyika 2017; Manyika et al. 2011; Marr 2018). Leveraging the capabilities and potentials of data can give companies a competitive edge by enabling, for example, faster decision making and generation of insights, but also to develop new process, product, and business innovations (Cronholm et al. 2017; Manyika 2017; Ransbotham and Kiron 2017; Vanauer et al. 2015). However, many companies are capturing only a fraction of the value resulting from the use of data (Manyika 2017; Nagle and Sammon 2017; OECD 2015).

Data and analytics are considered a powerful catalyst for innovation (Bauer et al. 2017; Manyika 2017; Parmar 2014; Ransbotham and Kiron 2017). Hence, data has to become part of the value creation process (Etventure 2018), which in turn requires a new way of thinking (also referred to as a mindset) and acting in a company (Harris 2012; Rogers et al. 2017; The unbelievable Machine Company 2016a, 2016b). That means, getting practical or rather working with data (Bange 2016; Bhargava and D'Ignazio 2015) and having data in mind while designing products, services, and processes that might eventually lead to data-driven innovations (Bange 2016; Kronsbein and Mueller 2019). In this context, Data Thinking has been introduced as an innovative, holistic approach that enables the development of value-creating use cases through the intelligent use of data at each stage of development (Etventure 2018; Kronsbein and Mueller 2019; The unbelievable Machine Company 2016c).

However, several obstacles for companies to become more data-driven have been identified, such as a lack of understanding how to use data and analytics, missing consistent methods and processes, a shortage of data science skills, as well as insufficient collaboration and lack of interdisciplinary communication between the IT department, data science, and business teams (Bauer et al. 2017; Cao 2018, p. 20; Cronholm et al. 2017; Kollwitz et al. 2018; La Valle et al. 2011; Panetta 2019; Rogers et al. 2017).

Data awareness—which includes identifying and understanding of data assets—is a crucial step within the exploration phase of data-driven innovations and can also inform the subsequent ideation and evaluation phases (Kayser et al. 2018; Kronsbein and Mueller 2019; Wirth and Wirth 2017). Limited understanding of possible data sources and knowledge limited to only a few data categories or sources are counted among one of many dimensions of poor data literacy (Sternkopf and Mueller 2018). Missing data awareness often goes hand in hand with missing data literacy.

Not knowing where to start (La Valle et al. 2011) and having no systematic way of learning what data assets are available results in the need of a visual collaborative tool to foster shared knowledge between different stakeholders involved in data-driven innovation projects. Such a boundary object (Carlile 2002; Star and Griesemer 1989; Thoring et al. 2019) could create awareness of data that exist within or outside of the organization, as well as a common understanding of its contents (Hornick 2018; Seidelin et al. 2018; Sternkopf and Mueller 2018). The goal of this paper is to develop a visual collaboration tool that helps to create a holistic overview of the data assets that are potentially available within and outside the organization. This tool should facilitate the collaborative discussion within teams of different data literacy levels. Hence, this paper aims to answer the following research question:

RQ: How could a visual collaborative tool help to raise the awareness of the potentially available data assets to facilitate data-driven innovations?

The remainder of this paper is structured as follows: the subsequent section discusses the related work about collaborative design and designing with data. In section "*Methodology*" the used research methodology is illustrated. The ontology of the Data Collection Map (DCM) as well as its visual representation is presented in section "*Artifact Description of the Data Collection Map*". In section "*Evaluation Workshops*", the paper discusses the key findings from evaluating the canvas and presents the implications as well as the limitations of this research. Finally, the paper concludes with an outlook to future work.

Related Work and Research Gap

When it comes to designing products, services or processes, co-design or participatory design is acknowledged as an important approach (Steen 2013). Benefits of co-design comprise (1) including stakeholders from different disciplines, (2) involving users and customers as participants in the design process, and (3) sharing and combining of knowledge about the design process and the design content in order to develop a shared understanding (Kleinsmann and Valkenburg 2008; Steen 2013). Mastering complexity in the form of ill-structured or wicked problems (Rittel and Webber 1973) initially requires a common understanding of the challenge or basis on which the different team members can build up on (de Graaf 2019). In this context, Design Thinking facilitates coping with ill-structured (wicked) problems (Mueller and Thoring 2012; Steen 2013).

As data-driven innovation arises, using data to support the design process (Kun et al. 2018; Speed and Oberlander 2016) and using data for innovation in order to develop new or significantly improve existing products, processes, methods or services (Bange 2016; Cronholm et al. 2017; Curley and Salmelin 2018; Hunke and Engel 2018; Meierhofer and Herrmann 2018; OECD 2015) is gaining more and more attention. Making data visible is considered as a facilitator for discussions with and through data in order to design data-related aspects of a future system, product, service etc. (Seidelin et al. 2018). Seidelin et al. (2018) highlight the complexity which exists when many different stakeholders across organizations interact or rather work with the same data, resulting in the need to make data more accountable for all actors of cross-organizational collaboration. Winter (2019) proposes "blackboxing data" that, in contrast to classical conceptual data models, is abstracting from operational details and instead focuses on the potential uses of data. However, only few approaches and tools exist that provide a shared language and visualization of the data aspects in innovation projects.

For instance, some approaches suggest a process or methodology (partly based on Design Thinking) to support the generation of data use cases and also highlight the ideation or creativity phase as one essential part of data-driven innovation (Bange 2016; Kayser et al. 2018; Kreutzer and Sirrenberg 2019; Paar and Blankenburg 2017; Stadelmann et al. 2019; Vanauer et al. 2015; Wirth and Wirth 2017). Furthermore, visual tools have been introduced that help getting started on identifying opportunities for AI or Machine Learning and could serve as a communication tool for domain experts, data scientists and business people (Agrawal et al. 2018; Dorard 2018). Others developed visual collaborative tools as boundary objects (Carlile 2002; Star and Griesemer 1989; Thoring et al. 2019) to tackle the complexity of working with data during the design process as well as having team members with different levels of data literacy (Kollwitz et al. 2018; Kronsbein and Mueller 2019). The Data Innovation Board (DIB) (Kronsbein and Mueller 2019) facilitates the collaborative ideation and development of data-driven products and services within the scope of Data Thinking workshops. A Data Vignette (Kollwitz et al. 2018) sums up the topics and structure of individual data sets to enable cross-disciplinary teams to design data-driven products and services by reducing knowledge boundaries. Nagle and Sammon (2017) introduced the Data Value Map, a visual strategy framework for data initiatives to tackle the problem of misalignment between data stakeholders while focusing more on a strategic level.

All these approaches and tools recognize identifying and collecting the available data assets, i.e. data awareness as a valuable design input (Agrawal et al. 2018; Bange 2016; Dorard 2016; Kollwitz et al. 2018; Kronsbein and Mueller 2019; Vanauer et al. 2015; Wirth and Wirth 2017). Knowing which data is available or learning more about potential data sources could inform data use case definitions right from the start (Bange 2016; Kayser et al. 2018; Kreutzer and Sirrenberg 2019; Vanauer et al. 2015; Wache et al. 2019; Wirth and Wirth 2017) or even entail that the initial use cases are revised (Wirth and Wirth 2017). Handling and getting practical with data requires a sound understanding of various types of available data sources, as well as experience in identifying and exploring data (Sternkopf and Mueller 2018). Stakeholders with varying data literacy levels need a shared collaborative tool for exchanging and combining their knowledge about data assets demands. This kind of a collaboration tool can function as a shared mental model as well as an extended memory (Thoring et al. 2019).

Hence, we seek to develop and evaluate a visual collaborative tool that raises data awareness among practitioners with varying data literacy levels during data-driven innovation projects.

Methodology

The research design of this paper follows Sein et al.'s (2011) action design research (ADR) methodology to design and evaluate the visual collaboration canvas in terms of its understandability, plausibility, usability, and perceived usefulness. ADR, which is part of design science research (DSR) (Hevner et al. 2004; Peffers et al. 2007), is particularly suitable for our research purpose as it allows the iterative and joint design of an artifact by both, the research team and organizational stakeholders such as practitioners and end users (Haj-Bolouri et al. 2018). The applied ADR methodology has been accompanied and enriched mainly by qualitative focus group discussions and a quantitative questionnaire.

The Data Collection Map was designed following the tree design principles suggested by Avdiji et al. (2018): (1) framing the ill-structured problem by developing an ontology, (2) representing the ontology into a shared visualization, and (3) instantiating the visualization in a way that supports shared prototyping of the solution. More details on the creation process of the tool can be found in the subsequent section "*Artifact Description of the Data Collection Map*".

The evaluation of the canvas was done using primarily qualitative methods by conducting three design workshops. We applied a mixed-method research design called convergent design by gathering qualitative insights mainly via semi-structured group discussions and observations complemented by quantitative insights from surveys (Creswell and Creswell 2017; Creswell and Plano Clark 2017, p. 68). Throughout the process of data collection and analysis, triangulation was used as a validity strategy. Hence, different data sources, such as focus group discussions, questionnaires, observations, and photographs of filled canvases were used to determine the accuracy of the findings and to add to the validity of the study. This data triangulation is considered favorable when ADR is conducted on dynamic research topics like workshop formats (Flick et al. 2004; Hussein 2009).

We decided to study the use of the canvas in three workshops that were developed and analyzed going through three iterations. Following the principles of ADR, the feedback and learnings as a result of the first two iterations were incorporated into the artifact to further improve the canvas. One iteration of the evaluation consists of three different steps: observations, focus group discussions, and questionnaires. After completing the task of working hands-on with the DCM and presenting the completed canvases, participants were asked to take part in a focus group. The focus group discussions were audio recorded and transcribed in order to analyze and cluster the statements by the participants into thematic categories. The data were complemented by a questionnaire to underpin the results from the discussions and to provide additional information to the primary qualitative data set.

The first two parts of the focus group discussions dealt with questions that address the general understanding of the tool in order to evaluate whether the objective of using the canvas and the canvas itself were clear. One exemplary question during the focus group discussion was: *"Was the structure of the Data Collection Map with the different blocks coherent?"*. The two following parts of the discussion involved questions covering the aspect of the usability and perceived usefulness of the canvas. One sample question was: *"What did you like most about using the Data Collection Map?"*.

The questionnaire follows a similar structure as the focus group discussion. For the questionnaire we used single-choice questions and Likert-scales partly leaning on existing measures, like perceived usefulness (Pigneur and Fritscher 2014; Torkzadeh and Doll 1999). We used a simplified questionnaire based on (Davis 1989). In addition, we extended the questionnaire with questions that covered the concepts of practicality and usability. Our focus group discussion guideline and the survey questions can be provided on request.

WS ID	**Work-shop partici-pants**	**# of groups**	**Partici-pation in focus group**	**Partici-pation in survey**	**Data literacy level**	**Industry**	**Domain knowledge**	**Data challenge**
1	8	1	4	6	Low-medium	Automotive	High	Real life: "How might we make our engineering change processes faster?"
2	20	4	19	20	High	Academia	Low	Fictional: "How might we use consumer data to build new digital products and services for clinics, hospitals or doctors?"
3	7	1	5	6	Low-medium	Automotive	High	Real life: "How can new technologies (like AI) help simplify the scheduling in Engineering Change Management? How could the data be used to support the core roles in the scheduling?"

Table 1. Research Design of the Workshops

In total, three workshops in different settings were conducted. Similar to Design Thinking workshops, participants were faced with a data challenge. The first workshop was a pilot project with a car manufacturer involving practitioners with mixed business and IT backgrounds and high domain knowledge about the business challenge. The second workshop was held in an academic setting with participants who had high data literacy but low domain knowledge about the raised challenge. The third workshop was with a car manufacturer involving participants with business background and high domain knowledge, the only workshop where no participant with IT background was involved. Table 1 outlines the setup of the three workshops with the different foci.

Artifact Description of the Data Collection Map

The following subsections describe the origin and objectives, the conceptual model (ontology), as well as the reasoning behind the visual structure of the Data Collection Map (DCM), which intends to facilitate the creation of data awareness.

Origin and Objectives of Designing a Visual Ideation Tool

The Data Innovation Board (Kronsbein and Mueller 2019) included three consecutive sections: explore, ideate, and evaluate. It functioned as a summary of new data-driven, user-centric ideas. The "explore" section includes existing data that should contain all the data that is currently stored or collected by the organization. While running the workshops, it became clear that most of the participants did not know what data was available. They knew the IT systems in place (e.g. Google Analytics or a specific CRM) but not the data that is stored within these systems.

Hence, there was the initial need to improve the awareness of workshop participants when it comes to data, and to simultaneously frame and further focus this awareness to only the most frequently used data types. Based on these requirements, we developed a list of objectives that should be covered by the developed artifact.

- R1: Create a better understanding about available data.
- R2: Focus the current understanding of participants towards data instead of IT systems.
- R3: Show the potential for data that can be stored or collected further (optional).

Ontology

This subsection describes the ontology, that is, the building blocks and their relations in the Data Collection Map.

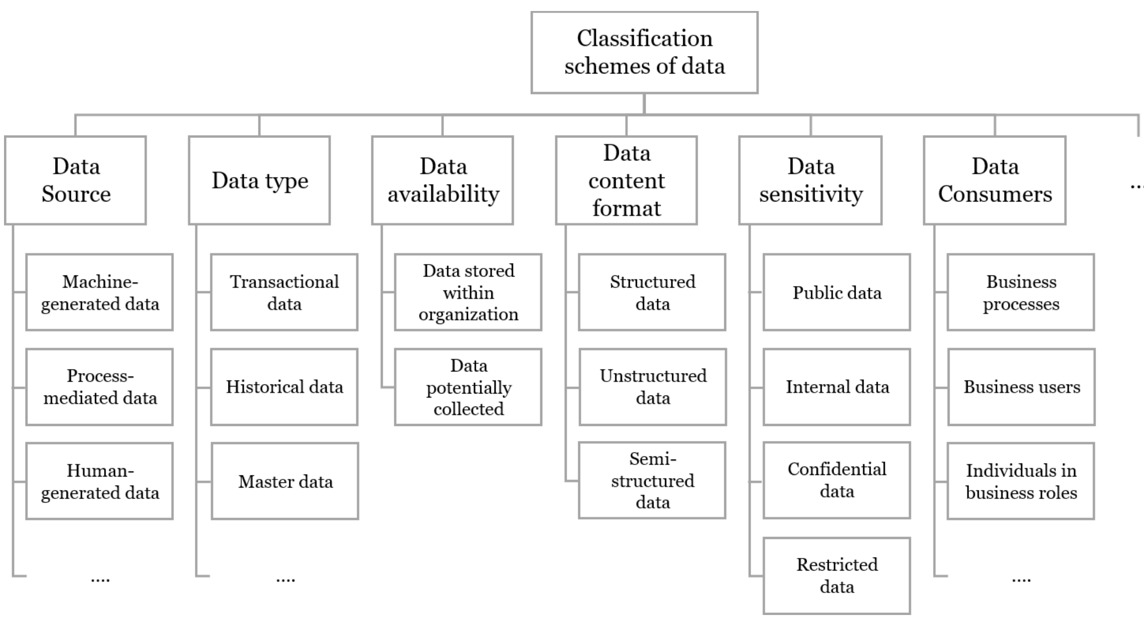

Figure 1. Data Classification Schemes

When it comes to building (big) data solutions, it is essential to classify nature and type of (big) data that is acquired, generated, stored, processed, and analyzed in many ways (Mysore et al. 2013). There are different classification schemes or rather patterns for data depending on the context in which data is examined (Hashem et al. 2015): Figure 1 only lists a selection of it. For instance, data classification can depend on the data generation process, i.e. where data is generated, distinguishing mainly between machine-generated data, process-mediated data and human-generated data (Devlin 2013). In addition, if the data source is taken into account, a distinction can be made between biometric and transaction data amongst others

(Mysore et al. 2013). Data classification can also be based on the type of data to be processed, including transactional data, historical data, and master data amongst others (Mysore et al. 2013). Furthermore, data can also be classified according to its availability within an organization, it can either be already available (within an organization) as it is already stored in existing systems (in whatever format) or could potentially be collected or even be accessed publicly (and hence from external sources respectively*)*. In the context of data analysis, data can be classified based on its content format, distinguishing primarily between structured data (e.g. numbers or geolocation), unstructured data (e.g. audio and images) and semi-structured data (Hashem et al. 2015; Mysore et al. 2013). Data classification in the context of information security is the classification of data based on its level of sensitivity, distinguishing between public, internal, confidential and restricted data[1]. Moreover, data can also be classified depending on the corresponding data consumers –possible consumers of the processed data– including business processes, business users, and individuals in various business roles amongst others (Mysore et al. 2013). Winter (2019) emphasizes data blackboxing, resulting from the need that IT-related concepts should be blackboxed in order to become understandable and usable by stakeholders with business background. With regard to creating data awareness and enhance collaboration between stakeholders with different backgrounds (Business, Data Science, etc.) in data-driven innovation projects, it is reasonable to classify data into categories which make data as a design material more tangible. It is important in the sense of data awareness that participants know where and in which systems data is to be located. Knowledge about the origin of the data facilitates the identification of stakeholders and those responsible for the respective systems (especially with regard to follow-up discussions). In addition, external partners can thus gain insights into the system landscape within the framework of innovation projects. The ontology developed with regard to data awareness in data-driven innovation projects distinguishes between five main types of data: (1) machine-generated data, (2) process-mediated data, (3) human-sourced data, (4) master and meta data, and (5) derived data and metrics (KPIs).

Data type	**Description**
Derived Data and Metrics	Data that results from other data and needs to be merged; in general: aggregated data, e.g. production volume per day
Master & Meta Data	Customer data, material data, etc., everything that rarely changes, e.g. Customer No 87, name, address, etc.; general descriptive data, e.g. "Which data records are stored with which provider in the cloud?"
Biometric Data	e.g. finger prints
Geo Data	e.g. tracking data, longitude and latitude
Measurement Data	Sensor data, machine data, e.g. Machine No 35, time stamp, temperature
Transaction Data	Transactions in the operative system, e.g. invoice details
Event Data	Data describing events, in general: log data, e.g. login attempt with wrong password
Interaction Data	Data from or for human users, e.g. E-mail, social media or wiki articles with best practices
Video Data	e.g. security video footage
Image Data	e.g. passport picture
Audio Data	e.g. audio recordings from virtual assistants

Table 2. Relevant Data Types

[1] https://edge.siriuscom.com/security/7-steps-to-effective-data-classification
https://www.srcsecuresolutions.eu/pdf/Data_Classification_Ownership.pdf

The selected types of data are not meant to be an exhausted list of all possible types but focus on the most frequently used data for innovation projects based on literature (Devlin 2013, p. 67) and our own previous data consulting experience. Table 2 shows the selected data types, partially based on (Devlin 2013, p. 67).

Machine-generated data are generated by sensors automatically and stored as measurements of these sensors (Devlin 2013, p. 67). This could be geo data based on, for example, a GPS sensor or biometric data based on, for example, a fingerprint sensor. Process-mediated data track the state of a business process. This could be data that describe an internal or external event like a worker strike or operational transactions data like a business order. Human-sourced data are created by and for users (Devlin 2013, p. 67). This could be interaction data like click data or social media messages. It could also be videos, images, or audio uploaded by users.

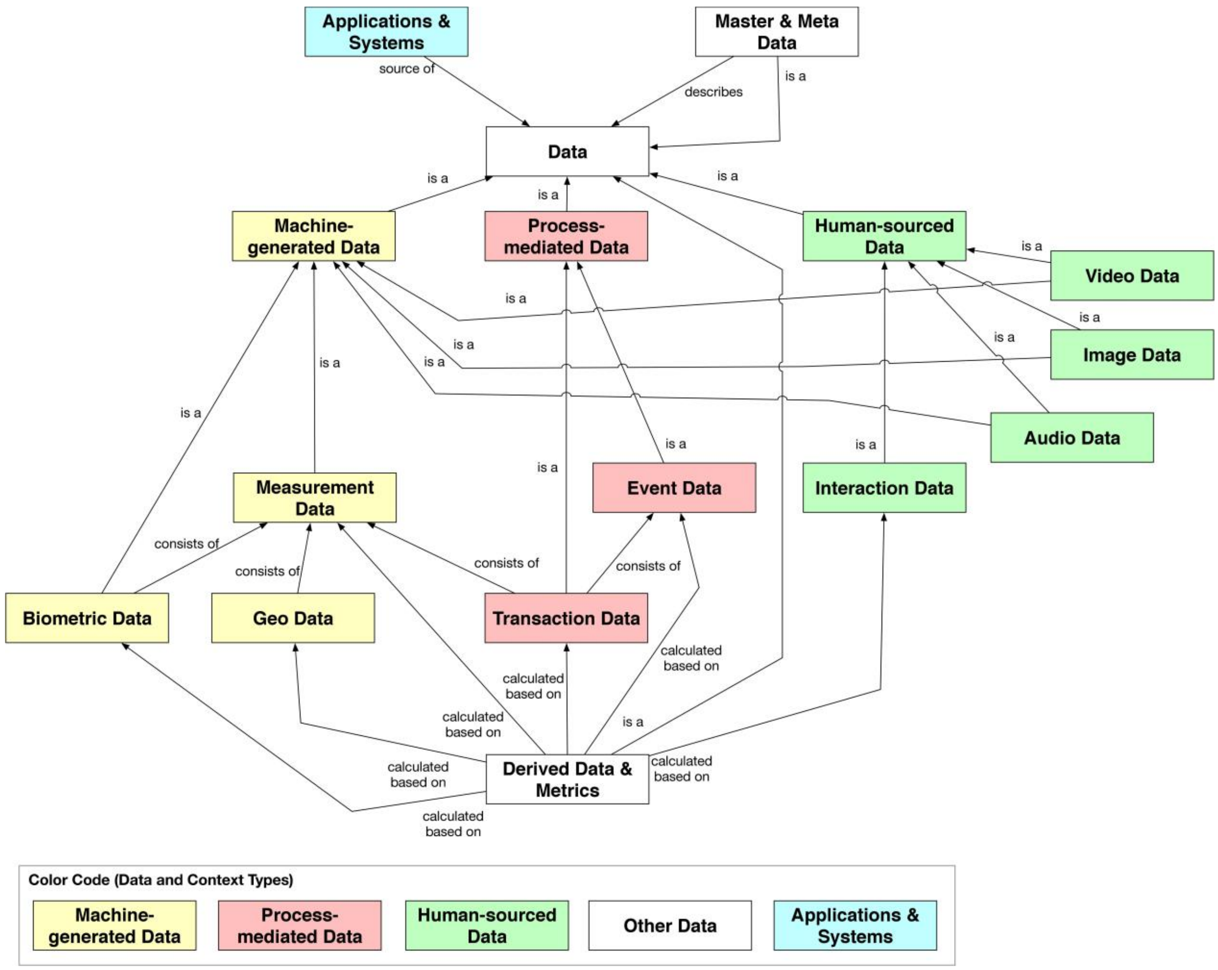

Figure 2. Ontology for the Data Collection Map

Human-sourced, process-mediated, and machine-generated data can be described by master data and meta data, that include, for example, who ordered the article (customer master data) or where a temperature sensor is located on a field (location meta data). Newly derived data and key performance indicators (KPIs) can be created by filtering, combining, and aggregating the data. Table 2 shows the definitions and examples of the different data types. All five main types of data are generated or stored by some application or IT system. Figure 2 shows the ontology with the relationships between the different elements.

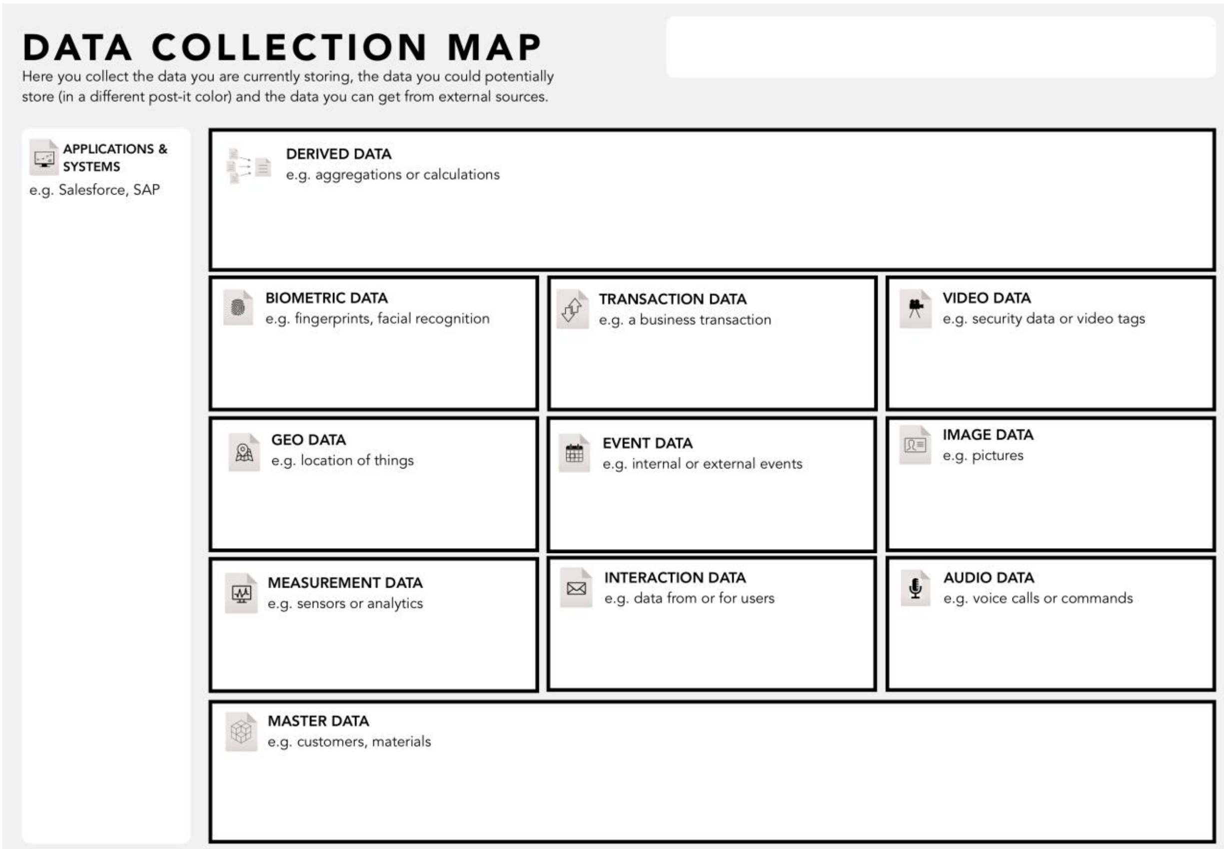

Figure 3. Data Collection Map (DCM)

Visual Canvas

Following the second design principle of Avdiji et al. (2018), the previously described ontology is converted into a shared visualization to enable collaboration within teams. The DCM as illustrated in Figure 3 consists of 11 building blocks. These building blocks resemble some of the most important data used by organizations (see Table 2 and Figure 2). The DCM was designed as an entry point in the ideation process of data-driven use cases. Hence, the purpose of the tool is to get people to think about data (e.g. clicks and engagement metrics) instead of IT systems (e.g. Google Analytics) and to raise the necessary data awareness about the available data resources within the organization.

The logic of the visual structure of the canvas is as follows: there are three main visual pillars of machine-generated, process-mediated, and human-sourced data next to each other, including the different data types (machine-generated data: measurement, geo, and biometric data; process-mediated data: event and transaction data; human-sourced data: interaction, audio, image, and video data).

At the bottom we placed master and meta data, because these data could relate to all three pillars. We placed derived data and metrics at the top because they can be generated by all three data pillars. At the left side we added a box for applications and IT systems.

Each building block has been given one or two examples to help the user deal with the different types of data. Also, each block includes a different icon to symbolize the data type.

The canvas depicted in Figure 3 shows the current iteration of the tool (final version) based on the ontology and the workshop findings. Some of these blocks were changed (i.e. moved or renamed) due to the findings of the validation workshops. Adaptations that have been conducted throughout previous iterations are described in the following sections. Furthermore, potential improvements and tests on the current iterations are described in a later section as well.

Evaluation Workshops

The ultimate goal of the conducted Data Thinking workshops was to design potential data use cases coming from data challenges and presenting them with creative (paper-based) comprehension prototypes. After the data challenge was specified, the participants were asked to collaboratively fill out a DCM. The outcome of this session eventually served as input for the subsequent ideation phase, in which data use cases were developed.

In the following subsection we will only describe a fraction of the workshops, because we were only interested in the part where the DCM was used. The DCM was tested during these workshops in order to (1) evaluate whether the objective of using the canvas and its structure was clear, and to evaluate its (2) usability and (3) perceived usefulness. The canvas was evaluated by triangulating the discussions, questionnaires, and observations (as described in the section *"Methodology"*).

Key Findings

The subsequent key findings follow the structure of the proposed evaluation criteria: (1) practicality of the DCM, (2) its usability, and (3) perceived usefulness.

The feedback from the focus group on the objective of raising data awareness was overall positively highlighted by the participants. This feedback was also confirmed by the quantitative analysis of the questionnaire: 75% of the participants either agreed or strongly agreed that the usage of the DCM has met its objective to its full extent, regardless of their data literacy level. 78% of the participants acknowledged that the expected outcome from using the DCM was completely understood. In general, participants from the focus groups positively mentioned the guidance provided by the tool. For instance, one participant highlighted that the tool would provide "a lot of initial help to even think in certain directions".

In addition, participants mentioned that the DCM supports discussions about the not so obvious data types. Since participants felt the urge to fill all the elements of the canvas, they already generated more potential data attributes than what was obvious. Hence, the DCM is of help to get an overview of what kind of data is currently stored or available, making transparent the amount and variety of the data. This is a good result in regards of generating new use cases in a later stage. Furthermore, feedback was given that the canvas would even inspire creativity in the sense of changing the perspective and thinking not only about collecting existing or currently stored data, but also data that could potentially be stored or even extracted from external sources. In this context one participant highlighted that: "[…] the creative freedom that you can give yourself is much wider if you simply change your perspective, and you do that with this method". Another participant noted that "the use of the canvas stimulates ideas as you already have a certain guideline with this sort of clustering and [hence] you do not just say that you write [the existing] data down [to paper], but again, as you have these [predefined] categories, it helps to develop ideas".

While Figure 3 shows the final version of the DCM after the conducted iteration cycles, the first layout of the canvas needed various adjustments. Some elements were often questioned by the workshop participants, which is why they were moved or renamed. In the first version, the block "master data" was placed at the top left of the canvas and the block "derived data" at the bottom below the remaining blocks. As a consequence of comprehensibility difficulties of some participants, we switched places of these two elements, arranging "master data" at the bottom functioning as the fundament of the above arranged data type blocks. The subordinate block, "derived data", is placed as a separate block above all the other data types, because it is derived from the canvas blocks below.

Moreover, participants with lower experience working with data often mentioned IT systems instead of data. For example, often they would place Google Analytics into the block of measurement data. However, this is what the tool wants to solve by intentionally increasing the awareness of the data within these systems. In order to get away from "thinking in systems" towards "thinking in data" or "having data in mind", and to ensure that participants consciously distinguish between system and data, we added an "applications & systems" block to the canvas after the first iteration. This adaptation was confirmed by participants of the second iteration with high data literacy but low domain knowledge (e.g. they did not know the standard software of the challenge) who never used to work with this building block but went straight into inserting examples of data. Furthermore, in contrast to the first version, we brought the elements in the final version into a logic that is more closely related to the five data types introduced in subsection "*Ontology*": placing the three columns of (1) human-sourced data, (2) process-mediated data, and (3) machine-generated data next to each other, whereas master data and derived data blocks were placed as individual blocks either above or below.

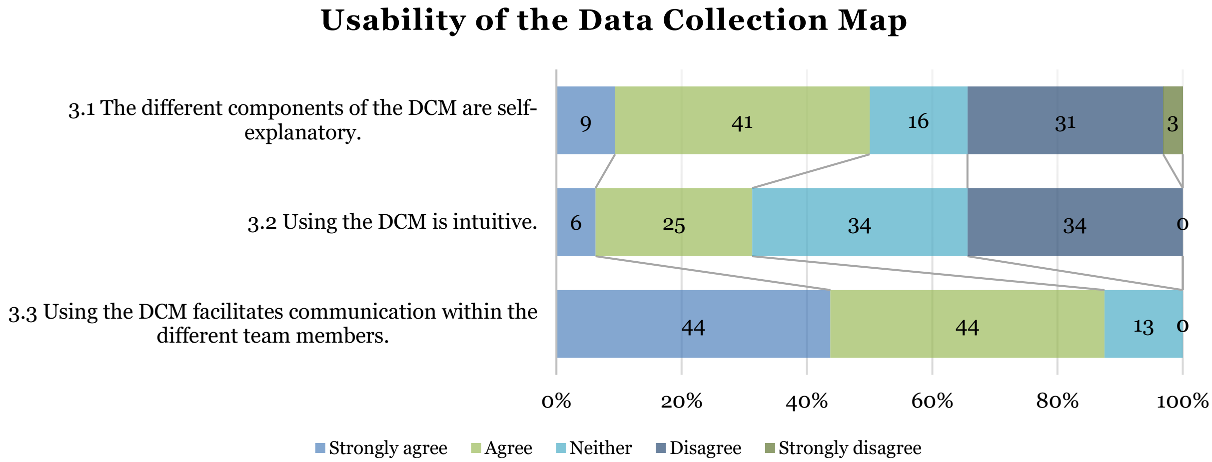

Figure 4. Usability of the Data Collection Map (N=32)

For all workshops, only 31% of the participants reported that using the DCM was intuitive (i.e. participants needed no additional explanation) and 34% claimed that the different components or rather building blocks of the DCM were not self-explanatory, regardless of whether having a low or high data literacy level (see Figure 4). Looking at the evaluation of the intuitively of the DCM across the three workshop iterations, the results of the survey show that, in percentage terms, intuitively was assessed more favorably after each iteration: none of the participants of the third workshop (strongly) disagreed that using the DCM is intuitive (see Figure 5). In general, the participants, especially those with a low data literacy, had a lot of questions about the meaning of some of the elements. Almost every participant has struggled with master data (especially during the first iteration cycle) and derived data. A possible hypothesis for this can be that these participants are on a different level of granularity and often did not know at what level of detail the data should be collected. In addition, the majority of the participants was not sure where to place certain elements as some data attributes could be allocated into more than just one element of the canvas. The stated examples below the blocks were intended to serve as a support for better understanding what the different types of data represent. As the goal of the tool is to increase data awareness, this situation is not regarded critical and can be counteracted by the coaches during the workshop.

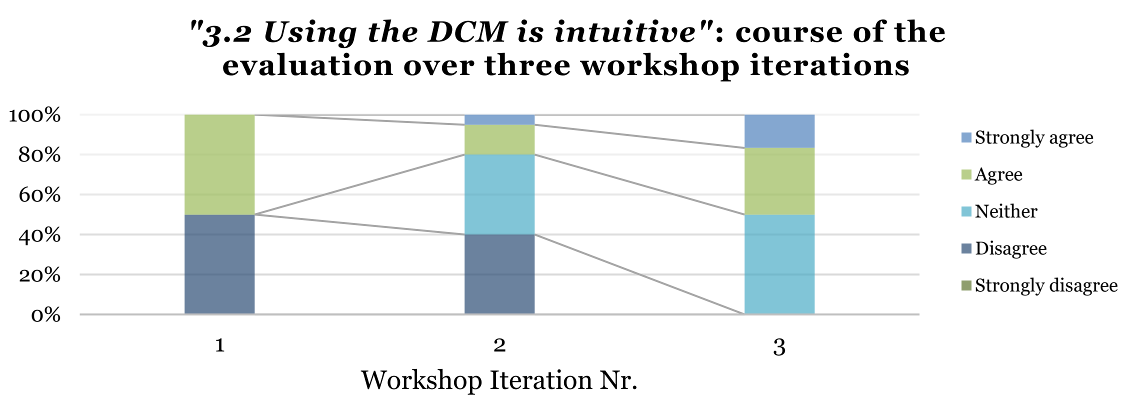

Figure 5: Evaluated Intuitively of the DCM over Three Iterations (in %)

In general, collaboration and communication were increased and further fostered: 88% of the participants stated that using the DCM facilitated communication within the different team members, the rest of the participants responded with neither or neutral (see Figure 4). Participants highlighted that using the DCM as a collaborative basis and having team members with a slightly different working background stimulated the creation of a communication platform where different ideas and perspectives can be discussed: “The exchange that you “look right and left”, ask "have you got that already?" and then maybe gets two or three more ideas about it. And then, of course, the exchange among each other, when we stood here in front of [the map] and looked at it, we pinned two, three, four [Post-its] on it, where we said “yeah, that is right, there is something else"”.

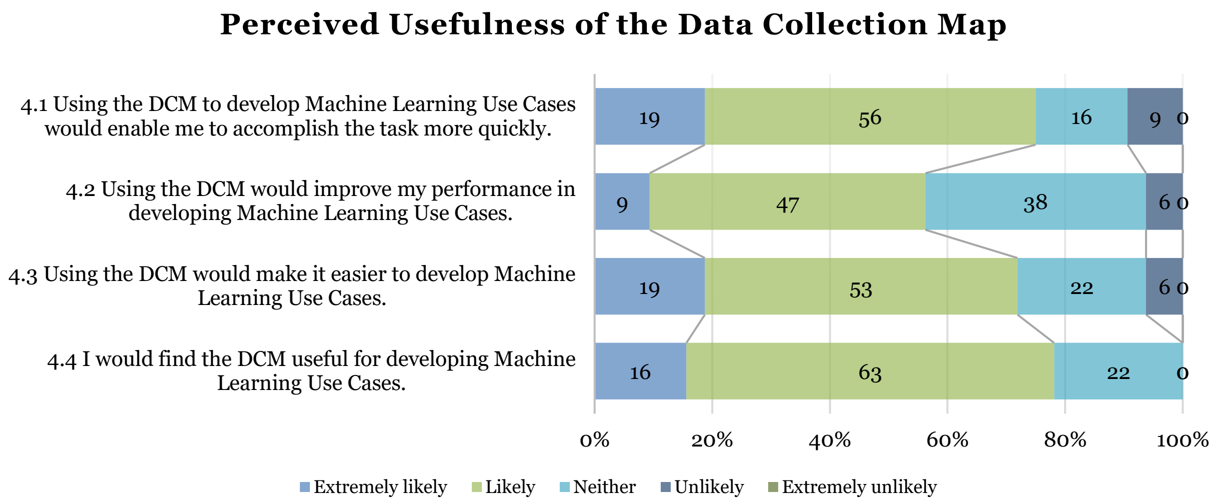

Figure 6. Perceived Usefulness regarding Development of ML Use Cases (N=32)

Figure 6 depicts the survey results regarding the perceived usefulness of the DCM. 79% of the participants either agreed or strongly agreed that they would find the DCM useful for the development of Machine Learning use cases regardless of their data literacy level (see question 4.4 in Figure 6). 72 % stated that using the DCM would facilitate the task of developing Machine Learning use cases (see question 4.3 in Figure 6). One participant highlighted that the DCM with its different types or branches of data fosters “thinking out of the box” and coming up with new ideas: “ […] it prepares you very well for the ideation process, […] just thinking about what kind of data exist, data that is important for hospital or doctors. And then for the customer side. You just dive into these and then of course it is much easier to come up with ideas at the second thought steps.” Other workshop participants noted that even ideation around possible data use cases was actually stimulated during the usage of the tool itself as they already started visualizing and ideating how the data placed on the map could be used in terms of data use cases. Moreover, working with the canvas was recognized as a “mind support” or “thinking aid” during the data collection phase: “[The Map] is a mental aid […]. There are a lot of things on it that you would not have thought of yourself. And when you bring that together in your head, you somehow come up with new thoughts, new steps, new processes or what is still missing. I think that is pretty good at this point. Because you can always check up and then draw lines in your head, so "that to that" and "that to that", what could you still need or what could still be useful”. In addition, one group noted that in a second step, the filled DCM can support the identification of relevant data for the development of a specific data use case: what data do we actually have? And where do we get the data from? etc.

Potential Improvements

The effect that the tool promotes creativity in the sense of thinking about data was overall positively noted by the workshop participants. However, feedback has been given that specifying categories may cause some ideas about data or data sources to be dropped by participants. A possible solution would be adding an “other non-related data” block to park data that cannot be clearly classified. In this regard the DCM, functioning as a thinking facilitator (Thoring et al. 2019), needs more customization or rather some further improvements to make sure that the ideation around collecting data not considered relevant at first glance (or rather not so obvious data types) is further facilitated or stimulated, and not thwarted which could lead to fixation (Thoring et al. 2019). Therefore, additional workshops are planned where the canvas shall be used and tested.

As mentioned in the previous subsection, often a lot of side discussions were held about the meaning of a specific block or where to place a certain data attribute. Hence, it needs to be further investigated through more workshops whether this fact has positive or negative implications for the tool.

Discussion

This paper presents the Data Collection Map (DCM), a visual collaboration tool that has been developed and evaluated using an ADR approach accompanied with qualitative focus group discussions and a quantitative questionnaire. The DCM is of benefit for teams that stand at the beginning of a data-driven design process and focus on the collaborative exploration of potentially available data assets. The DCM helps practitioners to sharpen their perspective on data assets and as a result raises data awareness along with enhanced data literacy. In doing so, the DCM complements the DIB developed by Kronsbein and Mueller (2019) by supporting and enabling the data exploration phase, which is the starting point for the subsequent ideation phase. The DCM would be most suitable to be used as an entry point prior to the ideation phase for the development of data-driven products, services, or processes. However, the canvas may also be useful later on in the design process, for example, when coming back to the canvas during the subsequent ideation or evaluation phase, in order to integrate more information about (potentially) available data assets and to further inform data use cases (Kronsbein and Mueller 2019; Wirth and Wirth 2017). In addition, the DCM supports collaboration and fosters communication between different stakeholders as it provides a visual platform to share and exchange knowledge as well as ideas around the (potentially) available data assets. In this regard, the tool allows the collection of the knowledge of (potentially) available data from variant stakeholders who are often driven by divergent interests and who have a different perspective of the data, for example influenced by diverse working backgrounds (IT department, data scientists and business teams).

The DCM can be considered as a part of a set of strategies within the framework of Data Thinking workshops, and thus supports the higher-level objective of developing data use cases in the context of data-driven innovation projects.

Additionally, the evaluation of the DCM demonstrated the need to further discuss and analyze the design principles of such visual collaborative tools which are rather data-driven. We have found that working hands-on with data in the design process can raise very different or new issues: an aspect of this research that could be further analyzed complementing related research attempts (Avdiji et al. 2018; Bhargava and D'Ignazio 2015; Thoring et al. 2019).

Limitations

Even though this research followed ADR in order to develop and evaluate a visual collaborative tool and hence holds a rather experimental character, there are also some clear limitations to consider.

We do not claim completeness of the data types listed in Table 2. The list was developed based on literature and our project experience but focuses on the most important types. In addition, it needs to be further evaluated or rather revised to what extent the format of a canvas might be too limiting for the complexity of the data and whether other formats such as card games might be more convenient or rather functional tool for classifying data. We decided to use the format of a canvas as we wanted to use this tool as a boundary object to facilitate first and foremost collaborative discussion within teams.

The practicality, usability, and perceived usefulness of the canvas has only been evaluated over a limited number of workshops. This fact also implicates the generalizability of the quantitative analysis. The results from the quantitative analysis of the questionnaire are not generalizable because of the small sample size (N=32). More iterations of the format with larger groups need to be conducted. However, the questionnaire results are only meant as an additional data source and are part of our data triangulation.

A large part of the workshop evaluation is based on observations and hence depends on the observer's impartiality. Therefore, it might be possible that the results of this research might be slightly biased. In order to compensate for possible subjectivity of the researchers, the research applied triangulation of different data sources.

Future research

Since iteratively designing and evaluating is key to ADR, next research steps will include further developing and testing the DCM during additional Data Thinking workshops. Over time, it would be interesting to see whether and how the usage of the DCM has an impact on generating data use cases or which concrete aspects of the canvas actually support data-driven innovation. Moreover, it might be of interest to study whether and how the DCM impacts the data literacy of its users. Since the DCM helps to raise awareness of data assets while looking from different perspectives (user, customer etc.), it might be worthwhile to study whether and to what extent the canvas might foster collaboration between the different stakeholders with varying backgrounds (IT, data scientists, and business people).

The focus of this paper was not on the process of generating data use cases itself, but on raising data awareness. Hence, future work may explore the challenges of data-driven innovation and how it can be furthermore supported in order to develop a framework and more suitable visual collaborative tools or canvases to facilitate the design of innovative data-driven use cases.

Building on this work, a more in-depth analysis and testing of the effect the DCM might have on the higher-level objective of Data Thinking workshops, (i.e. generating data use cases and fostering data-driven innovations), should be subject to future research. This next step would include the analysis and development of a detailed guideline or rather framework of how a data-driven design process could be guided.

Acknowledgments

We would like to thank the team from BMW around Jürgen Eder and Josef Schneider for making this project possible. A special thanks goes to Steffen Düring who actively supports our project and Dr. Uwe Friederich who figures as supervisor.

References

Agrawal, A., Gans, J., and Goldfarb, A. 2018. “A Simple Tool to Start Making Decisions with the Help of AI,” *Harvard Business Review Online*, , April 17. (https://hbr.org/2018/04/a-simple-tool-to-start-making-decisions-with-the-help-of-ai, accessed August 27, 2018).

Avdiji, H., Elikan, D., Missonier, S., and Pigneur, Y. 2018. “Designing Tools for Collectively Solving Ill-Structured Problems,” in *Proceedings of the 51st Hawaii International Conference on System Sciences (HICSS)*, , January 3, pp. 400–409. (https://doi.org/10.24251/HICSS.2018.053).

Bange, C. 2016. “Big Data And Digitalization Use Cases: How To Identify And Prioritize,” *The Digitalist by SAP*. (https://www.digitalistmag.com/future-of-work/2016/10/11/big-data-digitalization-use-cases-to-identify-and-prioritize-04506473).

Bauer, H., Richter, G., and Wüllenweber, J. 2017. “Smartening up with Artificial Intelligence (AI) - What`s in It for Germany and Its Industrial Sector?,” Report, Report, Digitial McKinsey.

Bhargava, R., and D’Ignazio, C. 2015. “Designing Tools and Activities for Data Literacy Learners,” in *Workshop on Data Literacy, Webscience*, Oxford, UK.

Burn-Murdoch, J. 2012. “Study: Less than 1% of the World’s Data Is Analysed, over 80% Is Unprotected,” *The Guardian*. (https://www.theguardian.com/news/datablog/2012/dec/19/big-data-study-digital-universe-global-volume).

Cao, L. 2018. “What Is Data Science,” in *Data Science Thinking* (Vol. 28), Data Analytics, L. Cao (ed.), Cham: Springer International Publishing, pp. 29–58. (https://doi.org/10.1007/978-3-319-95092-1_).

Carlile, P. R. 2002. “A Pragmatic View of Knowledge and Boundaries: Boundary Objects in New Product Development,” *Organization Science* (13:4), pp. 442–455. (https://doi.org/10.1287/orsc.13.4.442.2953).

Cave, A. 2017. “What Will We Do When The World’s Data Hits 163 Zettabytes In 2025?,” *Forbes*. (https://www.forbes.com/sites/andrewcave/2017/04/13/what-will-we-do-when-the-worlds-data-hits-163-zettabytes-in-2025/#16c19095349a).

Creswell, J. W., and Creswell, J. D. 2017. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, (Fifth.), Sage Publications. (https://www.sagepub.com/hi/nam/research-design/book255675#preview).

Creswell, J. W., and Plano Clark, V. L. 2017. *Designing and Conducting Mixed Methods Research*, (3rd), Sage Publications.

Cronholm, S., Göbel, H., and Rittgen, P. 2017. “Challenges Concerning Data-Driven Innovation,” in *The 28th Australasian Conference on Information Systems, Hobart Australia, December 4-6, 2017*. (http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1169951&dswid=-1004).

Curley, M., and Salmelin, B. 2018. “Data-Driven Innovation,” in *Open Innovation 2.0*, Innovation, Technology, and Knowledge Management, M. Curley and B. Salmelin (eds.), Cham: Springer International Publishing, pp. 123–127. (https://doi.org/10.1007/978-3-319-62878-3_).

Davis, F. D. 1989. “Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology,” *MIS Quarterly* (13:3), pp. 319–340. (https://doi.org/10.2307/249008).

Devlin, B. 2013. *Business UnIntelligence: Insight and Innovation Beyond Analytics and Big Data*, Technics Publications, LLC.

Dorard, L. 2016. *From Data to AI with the Machine Learning Canvas (Part I): A Framework to Connect the Dots between Data Collection, Machine Learning, and Value Creation*. (https://medium.com/louis-dorard/from-data-to-ai-with-the-machine-learning-canvas-part-i-d171b867b047).

Dorard, L. 2018. *The Machine Learning Canvas - A Handbook for Innovators and Visionary Managers Striving to Design Tomorrow`s Machine Learning Systems*. (https://gumroad.com/l/mlcanvas/gbi40xs?mc_cid=55c0a91d08&mc_eid=[UNIQID]).

Etventure. 2018. “Data Thinking: Das Potenzial von Machine Learning mit Data Thinking richtig nutzen,” Trendreport, Trendreport. (https://www.etventure.de/data-thinking/).

Flick, U., Kardoff, E., and Steinske, I. 2004. *A Companion to Qualitative Research*, Los Angeles: Sage Publications.

de Graaf, R. 2019. “Data Science Strategy for Projects,” in *Managing Your Data Science Projects* (Vol. 2), R. de Graaf (ed.), Berkeley, CA: Apress, pp. 25–49. (https://doi.org/10.1007/978-1-4842-4907-9_).

Haj-Bolouri, A., Purao, S., Rossi, M., and Bernhardsson, L. 2018. “Action Design Research in Practice: Lessons and Concerns,” in *26th European Conference on Information Systems: Beyond Digitization - Facets of Socio-Technical Change*, Portsmouth, UK.

Harris, J. 2012. “Data Is Useless Without the Skills to Analyze It,” *Harvard Business Review Online*. (https://hbr.org/2012/09/data-is-useless-without-the-skills, accessed May 22, 2018).

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., and Ullah Khan, S. 2015. “The Rise of ‘Big Data’ on Cloud Computing: Review and Open Research Issues,” *Information Systems* (47), pp. 98–115. (https://doi.org/10.1016/j.is.2014.07.006).

Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. “Design Science in Information Systems Research,” *MIS Quarterly* (28:1), pp. 75–106. (https://doi.org/10.2307/25148625).

Hornick, M. 2018. “Data Science Maturity Model - Data Awareness Dimension (Part 6),” *Best Practices, News, Tips and Tricks - Learn about Oracle’s R Technologies for Oracle Database and Big Data*, , June 8. (https://blogs.oracle.com/r/data-science-maturity-model-data-awareness-dimension-part-6, accessed June 9, 2019).

Hunke, F., and Engel, C. 2018. “Utilizing Data and Analytics to Advance Service: Towards Enabling Organizations to Successfully Ride the Next Wave of Servitization,” in *Exploring Service Science* (Vol. 331), G. Satzger, L. Patrício, M. Zaki, N. Kühl, and P. Hottum (eds.), Cham: Springer International Publishing, pp. 219–231. (https://doi.org/10.1007/978-3-030-00713-3_17).

Hussein, A. 2009. “The Use of Triangulation in Social Sciences Research: Can Qualitative and Quantitative Methods Be Combined?,” *Journal of Comparitive Social Work* (1), pp. 1–12.

Kayser, V., Nehrke, B., and Zubovic, D. 2018. “Data Science as an Innovation Challenge: From Big Data to Value Proposition,” *Technology Innovation Management Review* (8:3), pp. 16–25. (https://doi.org/10.22215/timreview/1143).

Kleinsmann, M., and Valkenburg, R. 2008. "Barriers and Enablers for Creating Shared Understanding in Co-Design Projects," *Design Studies* (29:4), pp. 369–386. (https://doi.org/10.1016/j.destud.2008.03.003).

Kollwitz, C., Mengual, M. P., and Dinter, B. 2018. "Cross-Disciplinary Collaboration for Designing Data-Driven Products and Services," *2018 Pre-ICIS SIGDSA Symposium on Decision Analytics Connecting People, Data & Things*, pp. 1–15.

Kreutzer, R. T., and Sirrenberg, M. 2019. "KI-Challenge – Wie Künstliche Intelligenz Im Unternehmen Zu Verankern Ist," in *Künstliche Intelligenz Verstehen*, R. T. Kreutzer and M. Sirrenberg (eds.), Wiesbaden: Springer Fachmedien Wiesbaden, pp. 271–315. (https://doi.org/10.1007/978-3-658-25561-9_).

Kronsbein, T., and Mueller, R. M. 2019. "Data Thinking: A Canvas for Data-Driven Ideation Workshops," *52nd Hawaii International Conference 2019*, pp. 561–570.

Kun, P., Mulder, I., and Kortuem, G. 2018. *Design Enquiry Through Data: Appropriating a Data Science Workflow for the Design Process*, presented at the British HCI 2018, pp. 1–12. (https://doi.org/10.14236/ewic/HCI2018.32).

La Valle, S., Lesser, E., Shockley, R., Hopkins, M., and Kruschwitz, N. 2011. *Big Data, Analytics and the Path From Insights to Value*, (MIT Sloan Management Review:Winter issue 2011).

Manyika, J. 2017. "What`s Now and next in Analytics, AI and Automation," Briefing note, Briefing note, , May. (https://www.mckinsey.com/featured-insights/digital-disruption/whats-now-and-next-in-analytics-ai-and-automation.).

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," , May. (https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation).

Marr, B. 2018. "How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read," *Forbes*. (https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#baf2eb260ba9).

Meierhofer, J., and Herrmann, A. 2018. "End-to-End Methodological Approach for the Data-Driven Design of Customer-Centered Digital Services," in *Exploring Service Science* (Vol. 331), Lecture Notes in Business Information Processing, G. Satzger, L. Patrício, M. Zaki, N. Kühl, and P. Hottum (eds.), Cham: Springer International Publishing, pp. 208–218. (https://doi.org/10.1007/978-3-030-00713-3_).

Mueller, R. M., and Thoring, K. 2012. "Design Thinking vs. Lean Startup: A Comparison of Two User-Driven Innovation Strategies," in *Proceedings of the DMI 2012 International Research Conference: Leading Innovation through Design*, Design Management Institute, pp. 151–161.

Mysore, D., Khupat, S., and Jain, S. 2013. *Introduction to Big Data Classification and Architecture: How to Classify Big Data into Categories*.

Nagle, T., and Sammon, D. 2017. "The Data Value Map: A Framework for Developing Shared Understanding on Data Initiatives," in *25th European Conference on Information Systems*, AIS Electronic Library (AISeL), pp. 1439–1452.

Newman, D. 2018. "Top 10 Digital Transformation Trends For 2019," *Forbes*. (https://www.forbes.com/sites/danielnewman/2018/09/11/top-10-digital-transformation-trends-for-2019/#7a71fdab3c30).

OECD. 2015. *Data-Driven Innovation: Big Data for Growth and Well-Being*, Paris: OECD Publishing. (https://doi.org/10.1787/9789264229358-en).

Paar, A., and Blankenburg, S. 2017. "Big Data for Engineers – Human-Centered Perspectives," in *17. Internationales Stuttgarter Symposium*, M. Bargende, H.-C. Reuss, and J. Wiedemann (eds.), Wiesbaden: Springer Fachmedien Wiesbaden, pp. 1071–1076. (https://doi.org/10.1007/978-3-658-16988-6_81).

Panetta, K. 2019. "A Data and Analytics Leader`s Guide to Data Literacy," *Gartner*. (https://www.gartner.com/smarterwithgartner/a-data-and-analytics-leaders-guide-to-data-literacy/, accessed June 9, 2019).

Parmar, R. 2014. "Driving Innovation through Data," Executive report, Executive report, IBM Institute for Business Value, p. 16. (https://www.ibm.com/big-data/au/en/big-data-and-analytics/Driving_innovation_through_data_Exec_Report_0520.pdf).

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).

Pettey, C. 2019. "Why Data and Analytics Are Key to Digital Transformation," *Gartner*. (https://www.gartner.com/smarterwithgartner/why-data-and-analytics-are-key-to-digital-transformation/).

Pigneur, Y., and Fritscher, B. 2014. "Business Model Design An Evaluation of Paper-Based and Computer-Aided Canvases:," in *Proceedings of the Fourth International Symposium on Business Modeling and Software Design*, Luxembourg, Luxembourg: SCITEPRESS - Science and and Technology Publications, pp. 236–244. (https://doi.org/10.5220/0005426602360244).

Ransbotham, S., and Kiron, D. 2017. "Analytics as a Source of Business Innovation," *MIT Sloan Management Review* (3).

Rittel, H. W. J., and Webber, M. M. 1973. "Dilemmas in a General Theory of Planning," *Policy Sciences* (4:2), pp. 155–169. (https://doi.org/10.1007/BF01405730).

Rogers, B., Maguire, E., and Nishi, A. 2017. "Data & Advanced Analytics - High Stakes, High Rewards," Forbes Insights, EY, p. 58.

Seidelin, C., Grönvall, E., and Dittrich, Y. 2018. *Data Work in a Knowledge-Broker Organisation: How Cross-Organisational Data Maintenance Shapes Human Data Interactions*, presented at the British HCI 2018, pp. 1–12. (https://doi.org/10.14236/ewic/HCI2018.14).

Sein, Henfridsson, Purao, Rossi, and Lindgren. 2011. "Action Design Research," *MIS Quarterly* (35:1), pp. 37–56. (https://doi.org/10.2307/23043488).

Speed, C., and Oberlander, J. 2016. "Designing from, with and by Data: Introducing the Ablative Framework," *Proceedings of DRS 2016*, pp. 2991–3004.

Stadelmann, T., Stockinger, K., Heinatz Bürki, G., and Braschler, M. 2019. "Data Scientists," in *Applied Data Science* (Vol. 16), M. Braschler, T. Stadelmann, and K. Stockinger (eds.), Cham: Springer International Publishing, pp. 31–45. (https://doi.org/10.1007/978-3-030-11821-1_).

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

Steen, M. 2013. "Co-Design as a Process of Joint Inquiry and Imagination," *Design Issues* (29:2), pp. 16–28. (https://doi.org/10.1162/DESI_a_00207).

Sternkopf, H., and Mueller, R. M. 2018. "Doing Good with Data: Development of a Maturity Model for Data Literacy in Non-Governmental Organizations," in *Proceedings of the 51st Hawaii International Conference on System Sciences (HICSS)*, Hawaii, USA: IEEE Computer Society, pp. 5045–5054.

The unbelievable Machine Company. 2016a. "Data Thinking: Erfolgsrezept für den digitalen Wandel - Eine Handlungsanleitung vom grundlegenden Denken zum digital souveränen Unternehmen," Report.

The unbelievable Machine Company. 2016b. "Data Thinking: New Thinking and Processes for the Digital Goals of Our Time (Part 1)," , March 10. (https://www.unbelievable-machine.com/en/blog/data-thinking-new-thinking-and-processes-for-the-digital-goals-of-our-time-part-1/, accessed May 22, 2018).

The unbelievable Machine Company. 2016c. "Data Thinking: New Thinking and Processes for the Digital Goals of Our Time (Part 2)," , March 14. (https://www.unbelievable-machine.com/en/blog/data-thinking-data-leadership-process-model/?__hstc=753710.efe422f5585e3cf5db570ec39fc1313b.1559901998737.1560256680408.1561799993912.6&__hssc=753710.1.1561799993912&__hsfp=2125903568, accessed June 9, 2019).

Thoring, K., Mueller, R. M., and Badke-Schaub, P. 2019. "Exploring the Design Space of Innovation Canvases," in *Proceedings of the Academy of Design Innovation Management Conference (ADIM)*, London, UK.

Torkzadeh, G., and Doll, W. J. 1999. "The Development of a Tool for Measuring the Perceived Impact of Information Technology on Work," *Omega* (27:3), pp. 327–339. (https://doi.org/10.1016/S0305-0483(98)00049-8).

Vanauer, M., Bohle, C., and Hellingrath, B. 2015. "Guiding the Introduction of Big Data in Organizations: A Methodology with Business- and Data-Driven Ideation and Enterprise Architecture Management-Based Implementation," in *48th Hawaii International Conference on System Sciences*, HI, USA: IEEE, pp. 908–917. (https://doi.org/10.1109/HICSS.2015.113).

Wache, H., Dinter, B., and Kollwitz, C. 2019. "Analytics Use Cases for Mass Customization – A Process-Based Approach for Systematic Discovery," in *52nd Hawaii International Conference on System Sciences (HICSS)*, IEEE, pp. 5868–5877.

Winter, R. 2019. "Blackboxing Data—Conceptualizing Data-Driven Exploration from a Business Perspective," in *The Art of Structuring*, K. Bergener, M. Räckers, and A. Stein (eds.), Cham: Springer International Publishing, pp. 153–163. (https://doi.org/10.1007/978-3-030-06234-7_15).

Wirth, N., and Wirth, R. 2017. "How to Meet the Four Key Challenges to Turn Big Data into Smart Data-Driven Solutions," *Research World* (2017:64), pp. 31–36. (https://doi.org/10.1002/rwm3.20516).