Building a Connection Between Decision Maker and Data-Driven Decision Process

Felix Kruse, Viktor Dmitriyev and Jorge Marx Gómez

Abstract It is quite common that most of companies' decisions are made based on feelings, intuitions or personal experiences. The reasons for such patterns have organizational, technical and process oriented backgrounds. For instance, there is no structured way to deal with the analytical results on both sides simultaneously – organizational and technical. Usually, in case of analytics the ones doing analysis (e.g. data scientists) and the ones using results of analytics (e.g. decision makers) are different persons. As a result, such a structure leads to ambiguity and misunderstanding between the involved parties. In order to bridge the existing gap between data scientists and decision makers, we introduced the

---

Felix Kruse
Department of Computing Science, Carl von Ossietzky University of Oldenburg,
Ammerländer Heerstr. 114-118, 26129 Oldenburg, Germany
✉ felix.kruse@uni-oldenburg.de

Viktor Dmitriyev
Department of Computing Science, Carl von Ossietzky University of Oldenburg,
Ammerländer Heerstr. 114-118, 26129 Oldenburg, Germany
✉ viktor.dmitriye@uni-oldenburg.de

Jorge Marx Gómez
Department of Computing Science, Carl von Ossietzky University of Oldenburg,
Ammerländer Heerstr. 114-118, 26129 Oldenburg, Germany
✉ jorge.marx.gomez@uni-oldenburg.de

Archives of Data Science, Series A
(Online First)
KIT Scientific Publishing
Vol. 4, No. 1, 2018

DOI 10.5445/KSP/1000085951/03
ISSN 2363-9881

Data Product Profile which links both data science and data-driven decision processes.

1 Introduction

In order to facilitate a more structured way of communication between data scientists and decision makers, data products need to be the central result of every single data analysis process. *Data products* therefore, should support the achievement of corporate goals within companies. In companies, the data product can be used for decision support in different forms and formats, such as a report, a service for end-consumers, or simply as an analysis model, which is deployed in the process. The main goal of the data product is to generate value based on companies' data, that is essentially based on the analysis of raw data available to companies. On the one hand, such an approach aims to encourage decision makers to rely more on data-driven decisions. On the other hand, the data product serves as an intermediary between data scientists and decision makers (Anderson (2015b); Patil (2012); Provost and Fawcett (2013a)).

Despite the huge amounts of available raw data and various analytical tools to process them, most of the companies' decisions are based on certain individual factors. According to surveys conducted by Anderson (2015b); BARC (2017), around 58% of companies' decisions are made based on individualy feelings, intuitions or experiences, rather than being explicitly derived from corresponding data. Thus, companies' decisions are not yet data-driven and multiple studies have shown that a data-driven decision culture can provide a significant advantage over competitors (McElheran and Brynjolfsson (2017); Brynjolfsson and McElheran (2016)).

The realization of data products will open the door not only for data science experts, but also for domain knowledge experts, such as decision makers. In this way, companies will benefit from mechanisms of data science and move towards data-driven decision processes. The data product should be created in collaboration between data scientists and decision makers (Kowalczyk and Buxmann (2014a); Provost and Fawcett (2013a)).

According to the work of Provost and Fawcett (2013a), there are clear dependencies between data-driven decisions and data science. The starting point is data processing and data engineering. These steps serve as the basis for data science and analytical processes. Regardless if these dependencies' structure, it remains unclear how exactly an interaction between data-driven decisions and data science could be facilitated. Up to now, most companies have performed descriptive analysis in the form of reports. But most of them do not completely fulfill requirements of users, or are not explicitly used by decision makers. Therefore no value is created by using data (Anderson, 2015a). As the demand for predictive and prescriptive analysis increases, the need for data scientists in companies grows. Although one of the most important skill areas of data scientists is communication, the interaction and communication between data scientists and decision makers is not formally described nor properly addressed (Davenport and Patil (2012); Schmid and Baars (2016)).

In other words, when a company builds a particular data science process based on the widespread *Cross Industry Standard Process for Data Mining (CRISP-DM)* standard (Wirth and Hipp, 2000), it still remains unclear, how the CRISP-DM methodology could be adjusted to be able to reflect all information needs of decision makers. Currently, within the CRISP-DM model, the role of the decision maker is summarized in the *Business Understanding* step. It is the first step of the analytical process and it is not sufficient to integrate domain knowledge only at this stage. Decision makers may play also a valuable role during the further stages of the methodology as well.

For instance, according to Kowalczyk and Buxmann (2014a), in order to have a successfully deployed data-driven decision process, an interaction between the decision maker and the data scientist should be facilitated and supported by heterogeneous methodologies. It is necessary to reach a common understanding of a problem and its solution from both sides, the ones performing data analysis, and obtaining insights, as well as the ones using these insights to make decisions and, therefore, create corporate value.

To bridge the gap between data scientists and decision makers, we introduce the *Data Product Profile* (DPP) as a method for communication optimization. The DPP interlinks both data science and data-driven decision processes. According to Loukides (2010); Patil (2012); Stockinger and Stadelmann (2014), the data product is defined as follows: "A data product is the central result of the analysis process of data science. The data product supports the achievement of business goals. (...) The main objective is for the data product to generate added value from the analysis of the data". It allows decision makers to participate more actively during important steps of a data analysis process, without the need of deeper understanding of data mining, machine learning and deep learning techniques.

By encouraging decision makers to participate in a data analysis process, the quality of the obtained results will improve, as the experts have a lot to offer in terms of domain expertise. In addition of obtained results quality, such an approach will also increase the acceptance rate of the resulting analytical model by decision makers, since decision makers will be a part of their creation. By taking part in generating such models, decision makers will be more confident about the obtained results. This will increase the probability of integrating the results into the decision making process.

To determine the requirements for the DPP, a systematic literature review has been carried out. After the requirements have been determined, the DPP was developed. To meet the requirements, the decision process and the CRISP-DM were combined and the DPP as a process-supporting artefact was developed. The last step includes the evaluation of the developed artefacts in the frame of five case studies and two workshops in an SDAX company.

2 Related Work: Data Science and Data-Driven Decisions

In this section, the current state of research as well as the relationship between data science and data-driven decisions is discussed. Furthermore, the works of Provost and Fawcett (2013a); Kowalczyk and Buxmann (2014b); Elgendy and Elragal (2016); Cato (2016) are discussed.

Figure 1: The Relationship between Data Science and Data-Driven Decision Making, adapted from (Provost and Fawcett, 2013a).

The publication of Provost and Fawcett (2013a) shows the relationship between data science and decision making (see Figure 1). This demonstrates that data science can help to support the data-driven decision making process.

The contribution of Kowalczyk and Buxmann (2014a) represents the actual state of business intelligence and data science based on decision making processes. The authors describe the relationship between data science and data-driven decision making processes in a more detailed way in comparison to Provost and Fawcett (2013a). Kowalczyk and Buxmann (2014a) describe the current state of data-driven decision making processes and postulate important aspects to be taken into consideration, but they do not show how a company can begin to use data science and promote data-driven decision making processes.

The contribution of Elgendy and Elragal (2016) has a more technical view than Provost and Fawcett (2013a) and Kowalczyk and Buxmann (2014a). An interdisciplinary team in an analytical project, and the mentioned gap between data scientists and decision makers do not play a role in this article. They present their "Big - Data, Analytics and Decisions" (B-DAD) framework which links the decision making process together with the analytical process which includes all required technologies.

Cato (2016) focuses on big data systems, which often have the goal to enable and promote data-driven decisions. The work shows, that in addition to technical requirements, organizational requirements play an important role for analytical projects and corporate culture. However, the dissertation does not describe how exactly any company could use data science to promote data-driven decisions.

The presented and discussed contributions show the connection and importance of data science and data-driven decision processes. The contributions highlight some of the obstacles that should be overcome when using data science to support data-driven decision making processes. However, none of the papers provide a concrete guidance or concept of how data science should be used to support data-driven decisions. The current work is bridging this gap by introducing the new concept, the Data Product Profile (DPP).

3 Requirements

The requirements for the DPP to build a connection between decision maker and data-driven decision process, were determined with the help of a qualitative literature review. The publications used are shown in table 1. With the qualitative literature review the following research question should be answered:

What are the requirements for data science in relation to the decision making process?

Table 1: Literature used in the requirement analysis.

Author	**Title**
Baars (2016)	Predictive Analytics in der IT-basierten Entscheidungsunterstützung - methodische, architektonische und organisatorische Konsequenzen
Cato (2016)	Einflüsse auf den Implementierungserfolg von Big Data Systemen
Davenport et al (2010)	Analytics at work: Smarter decisions, better results
Davenport (2012)	Business Intelligence and Organizational Decisions
Dimitri Gross and Opitz Consulting (2016)	Big Data organisieren - Erste Schritte zum Competence Center
Duan and Cao (2015)	An Analysis of the Impact of Business Analytics on Innovation
Elgendy and Elragal (2016)	Big Data Analytics in Support of the Decision Making Process
Gross and Thomsen (2016)	Advanced Analytics: Die konsequente Antwort auf Big Data
Guerra and Borne (2016)	10 signs of data science maturity
Howard et al (2012)	Designing great data products: The Drivetrain Approach: A four-step process for building data products
Kim (2016)	Five steps for success
Kowalczyk and Buxmann (2014a)	Big Data und Informationsverarbeitung in organisatorischen Entscheidungsprozessen
Miller and Mork (2013)	From Data to Decisions: A Value Chain for Big Data
Patil (2012)	Data Jujitsu: The art of turning data into product: Smart data scientists can make big problems small
Provost and Fawcett (2013b)	Data science for business: [what you need to know about data mining and data-analytic thinking]
Provost and Fawcett (2013a)	Data Science and its Relationship to Big Data and Data-Driven Decision Making
Schmarzo (2016)	Big data MBA: Driving business strategies with data science

A total of 16 requirements are divided into two groups. The first group, with an organizational focus coded as ORx in table 2, has 13 requirements, differencing the second group with a technical focus coded as TRx in table 2 has 3. Besides organizational or technical focus, requirements are also divided according to implementation horizon. 7 requirements have a short-term implement (S) horizon, 7 requirements have an implementation horizon between short-term and long-term (S, L) and 2 requirements have a long-term (L) implementation horizon. In Table 2 the requirements are shown in descending order, with the number of apperances in the literature and their implementation horizon (S > S,L > L).

Table 2: Overview of the requirements lexicographically sorted by number of entries and implementation horzion (S > S,L > L).

Code	**Requirement**	**Type**	**Amount**
OR11	Support communication and cooperation between data scientist and decision makers	S	11
OR13	Focus on company's benefits	S	8
OR3	Data laboratory	S, L	7
OR1	Analytical approaches and decision process combined together	S	6
OR9	Data science team	S, L	6
OR8	Data science competences	S, L	5
OR5	Data product	S	5
OR10	Data-driven company culture	L	5
OR2	Data analytics thinking	S, L	5
TR1	Analytical infrastructure	S, L	5
TR2	Analytical and statistical functionalities	S, L	3
TR3	Information and data quality	S, L	2
OR12	Management support	L	2
OR4	Data product as an alternative for action in a decision making process	S	1
OR6	Secure data product insights	S	1
OR7	Identify data product stakeholders	S	1

Organizational requirements are mostly related to processes and procedures, whereas technical requirements are relatively general and represent basic requirements to be able to perform data science in any company at all.

The main objective of the concept is to implement requirements with a short-term implementation horizon (S), as these must be fulfilled in order to use data science to support the data-driven decision making process. Short-term requirements include OR11 with 11, OR13 with 8, OR1 with 6, and OR5 with 5 responses; OR4, OR6, and OR7 each with one entry (see Figure 2).

4 Towards the Data Product Profile Concept

In this section, the development of the Data Product Profile (DPP) is described. The DPP could be defined as an instrument that creates a common knowledge base of the participating stakeholders and enhances the approach advocated by the CRISP-DM methodology. The DPP is described in a way that the provided description supports the previously determined requirements. In order to operationalize the DPP, it should be integrated into an analytical process. For this reason, the CRISP-DM was chosen as the most common and well established generic methodology to perform an analytic process in companies (Brown (2016)).

In order to connect data science and data-driven decision, the data product is placed between them as a central instrument (see Figure 2). By placing the data product between data science and data-driven decision, the importance of the data is to be promoted as an important business resource (OR5).

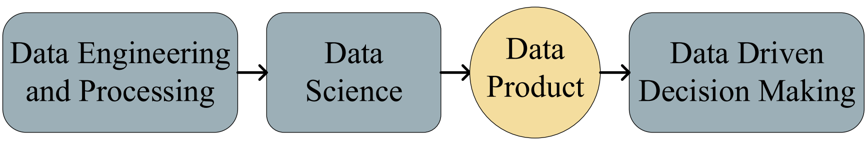

Figure 2: Relationship between Data Science, Data-Driven Decision Making and Data Product, based on (Provost and Fawcett, 2013a, p.54).

In this sense, any data product should be seen as a result of data science activities. Thus, it should be considered as an alternative action in the decision making process in order to carry out data-driven decisions (OR4).
By combining the CRISP-DM with the classic decision-making process, a data-driven decision-making process emerges (see Figure 3).

This combination fulfills the requirement OR11. Thus, the decision process can be combined with the CRISP-DM. Both have an iterative nature and may have some parallels in their process steps. The data-driven decision-making process should also be iterative.

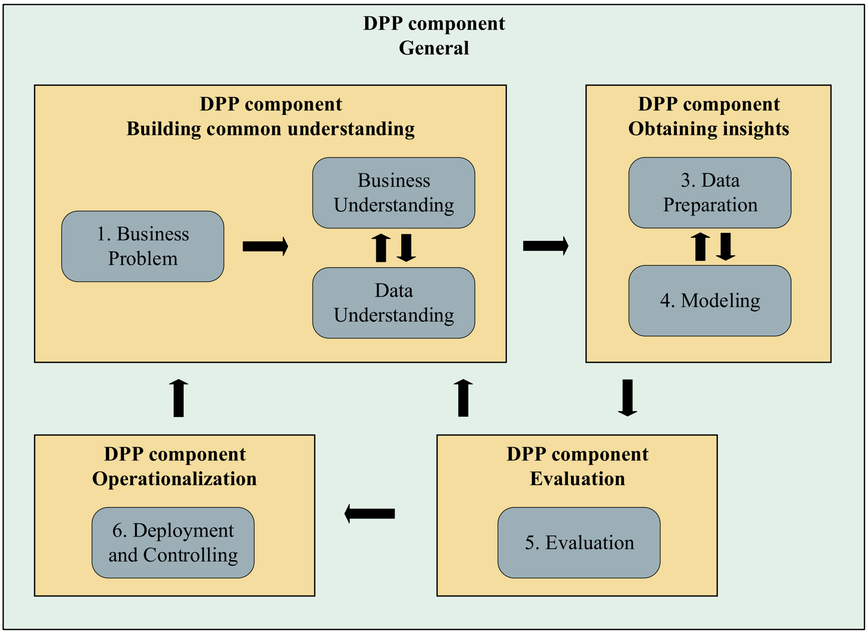

Figure 3: The Design of the DPP Components.

The process depicted in Figure 3 is to be carried out by a data science team. In the team, data engineers, data scientists and decision-makers (e.g. business owners, project sponsors, etc.) should work together to develop a suitable data product. In this process, communication and collaborative work should be ensured (OR11). To ensure this, the DPP will be integrated into the data-driven decision process (see Figure 3). The DPP is intended to support the communication and cooperation of the team members in all process steps. Among other things, it provides a guideline for structurally processing all essential topics in the individual process steps and for building a common understanding. Next process steps should only be started once all relevant topics of the previous process step have been completed. This is intended to promote a structured transition of the data-driven decision making process.

The DPP serves as a supporting tool in the data-driven decision making process. Three existing concepts were used for the content design of the DPP. Because a data product can be seen as kind of company internal start-up idea, the content from the Business Model Canvas, a widely used start-up support framework, was used. The Drivetrain Approach Framework influenced the design of the DPP content, as it represents a rough data product development process. The third existing concept used is the CRISP-DM (Howard et al, 2012; Osterwalder and Pigneur, 2013). The DPP should create the following added value during the process:

- The participation of all data science team members in the creation of the Data Products reduces the resistance to data-driven decision making processes and promotes their acceptance at the same time. The DPP is the central medium to support all involved stakeholders in their communication and collaboration (OR11).
- The participation and the collaborative development of the DPP lead to an exchange of knowledge between the team members. This knowledge exchange is intended to foster the team members data analytic thinking. Data scientist and decision makers should learn from each other.
- An often complex analytical project gets a structured guide through the DPP to master the complexity.
- There is a central place for securing knowledge about the data product (OR6).
- A common unified knowledge base is built within the data science team, which is always visible through the DPP and can be shared with other employees. The knowledge base can also be used in future projects.

The DPP should support the individual process steps of the developed data-driven decision process (see Figure 3), so that it can be executed in a structured manner. The process is divided into 5 components that can be assigned to the process steps accordingly (see Figure 3). Component *building common* *understanding* should support the process steps business problem, business understanding and data understanding. If the component of the DPP is sufficiently processed, component *obtaining insights* should support the process steps of data preparation and modelling. After component *obtaining insights* of the DPP has been completed, component *evaluation* for the evaluation process step and component *operationalization* for the process step deployment and controlling follow. Component *general* contains topics such as tasks, efforts and show stoppers and should be available in addition to the other components in each process step.The five DPP components are briefly described below:

Building common understanding: The business problem is used as the trigger to start the development of a data-driven decision making process. The DPP is intended to help with an initial understanding of the data, the definition of the company benefit and the decision to be supported.

Obtaining insights: The 2nd component of the DPP is intended to support the process steps such as data preparation and modelling. Once all essential component 1 content of the DPP have been populated, the data product can be developed in these two process steps.

Evaluation: After the potential data products have been developed and component 2 of the DPP has been documented, the evaluation can begin. The evaluation process step is supported by component 3. Important information for the evaluation has already been documented in component 1. The information on "how" and "when" the success of the data product should be measured is used here.

Operationalization: The 4th component of the DPP supports the deployment and controlling process step. In this process step, the previously evaluated and selected data product is operationalized.

General: The 5th component of the DPP is to be used throughout the entire analytical process. The component consists of general content that is relevant at all times during the process.

5 Prototypical Implementation through Data Product Profile Guideline

The evaluation of the DPP was carried out with the qualitative methods case study and expert's workshop. The business partner provides both the case studies and the experts for the evaluation of the concept. Case studies are used to evaluate developed artefacts under real conditions. For the execution of the case studies, the DPP was prototypically implemented in the form of a guideline (see Figure 4).

The guideline includes short content key points or questions for every DPP component which should help the data science team to develop a common understanding and a persistent documentation of the data product. For the five case studies, the guideline was presented to the participants in a Power Point presentation. The topics discussed on the respective contents of the guideline were documented separately in protocols. Table 3 summarizes the results of the evaluation and shows which requirements have been confirmed as implemented in the case studies and expert workshops.

Table 3: Summary of the Evaluation Results.

Code	**Requirement**	Type	Horizon	Case Study 1	Case Study 2	Case Study 3	Case Study 4	Case Study 5	Workshop 1	Workhsop 2	∑
OR1	Analytical approaches and decision process combined together	O	S	X	X	X	X	X			5
OR4	Data product as an alternative for action in a decision making process	O	S		X	X			X		3
OR5	Data product	O	S	X	X	X	X	X	X	X	7
OR6	Secure data product insights	O	S	X	X	X	X	X	X		6
OR7	Identify data product stakeholders	O	S	X	X	X	X	X			5
OR11	Support communication and cooperation between data scientist and decision makers	O	S	X	X	X	X	X	X	X	7
OR13	Focus on company's benefits	O	S	X	X	X	X				4

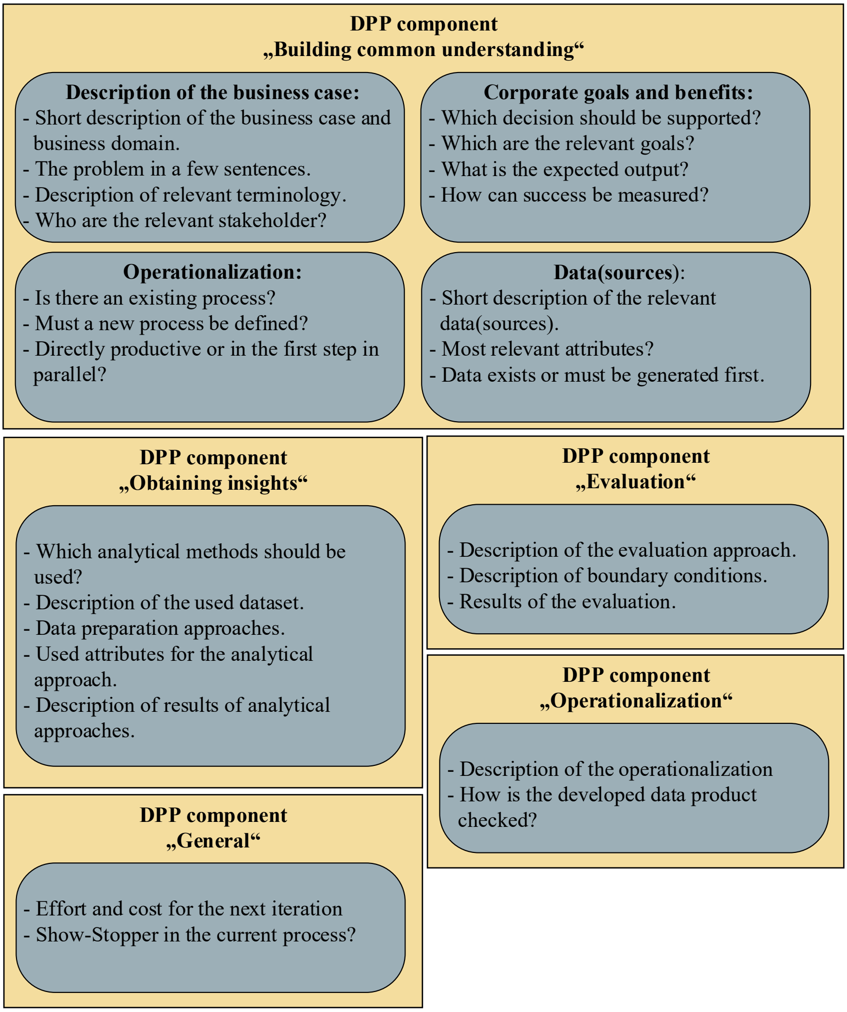

Figure 4: DPP guideline for the five components.

In Table 3 it can be seen that requirements with a short-term implementation horizon (**S**) could be confirmed after being implemented in the case studies and expert workshops. The most frequent requirement, OR11 was confirmed as implemented in all case studies and expert workshops. Also the requirements

OR5 with 7, OR6 with 6, OR1 with 5 and OR7 with 5 confirmations were implemented during the evaluation. The second most frequent requirement OR13 has been confirmed as implemented in four case studies. The requirement OR4 could be confirmed as implemented in two case studies and one expert workshop. The evaluation showed, that the concept implements the short-term requirements for the use of data science in support of data-driven decision making process.

6 Conclusion and Future Work

Despite the fact that there is a lot of work done in data science on organizational and technological levels. The work demonstrates that there are a number of requirements not fully addressed yet. Thus, in order to address most crucial gaps existing on the intersection of organizational, procedural and technical aspects of data science and data-driven decisions, a new concept, the data product profile (DPP), was introduced. The DPP was prototypically implemented through a guideline, in order to evaluate the DPP by five case studies. The evaluation shows that the DPP enables a structured, target-oriented and thus faster development of data products. In addition, a common understanding among the participants was fostered.

The evaluation was only qualitative. Future work could quantitatively evaluate the DPP, for example by conducting surveys before and after a development of a data product to measure the benefits. Furthermore, the guideline could be supported by software.

References

Anderson C (2015a) Being data-driven: It's all about the culture: Data analysis can change an organization's future – but only if it's used in the present, every day, by everyone., URL `https://www.oreilly.com/ideas/being-data-driven-its-all-about-the-culture`

Anderson C (2015b) Creating a Data-driven Organization: Practical Advice from the Trenches. O'Reilly Media, Inc., Sebastopol. ISBN: 978-1-491916-91-9

Baars H (2016) Predictive Analytics in der IT-basierten Entscheidungsunterstützung - methodische, architektonische und organisatorische Konsequenzen. Controlling 28(3):174–180, DOI 10.15358/0935-0381-2016-3-174

BARC (2017) 14 Survey-Based Recommendations on How to Improve Data-Driven Decision-Making, URL https://bi-survey.com/data-driven-decision-making-business

Brown M (2016) The Right Process For Big Data Analytics Profit, . URL https://www.forbes.com/sites/metabrown/2016/03/31/open-standard-process-yields-best-big-data-analytics-results/

Brynjolfsson E, McElheran K (2016) Data in Action: Data-Driven Decision Making in US Manufacturing. US Census Bureau Center for Economic Studies, DOI 10.2139/ssrn.2722502, URL http://dx.doi.org/10.2139/ssrn.2722502

Cato P (2016) Einflüsse auf den Implementierungserfolg von Big Data Systemen. Dissertation, Verlag Dr. Kovač

Davenport T, Patil D (2012) Data Scientist: The Sexiest Job of the 21st Century, URL https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

Davenport TH (2012) Business Intelligence and Organizational Decisions. In: Organizational applications of business intelligence management, Herschel RT (ed), IGI Global (701 E. Chocolate Avenue, Hershey, Pennsylvania, 17033, USA), Hershey, Pa., p. 1–11, ISBN: 14-6660-280-5

Davenport TH, Harris JG, Morison R (2010) Analytics at work: Smarter decisions, better results. Harvard Business Press, Boston, Mass.

Dimitri Gross, Opitz Consulting (2016) Big Data organisieren - Erste Schritte zum Competence Center. IT-Management, p. 2–5

Duan Y, Cao G (2015) An Analysis of the Impact of Business Analytics on Innovation. In: Twenty-Third European Conference on Information Systems (ECIS), p. 26–29, URL http://www.ecis2015.eu/conference/program.html

Elgendy N, Elragal A (2016) Big Data Analytics in Support of the Decision Making Process. Procedia Computer Science 100:1071–1084, DOI 10.1016/j.procs.2016.09.251

Gross D, Thomsen C (2016) Advanced Analytics: Die konsequente Antwort auf Big Data. ONLINE-THEMENSPECIAL PREDICTIVE & ADVANCED ANALYTICS,

Guerra P, Borne K (2016) 10 signs of data science maturity, URL https://www.oreilly.com/ideas/10-signs-of-data-science-maturity

Howard J, Zwemer M, Loukides M (2012) Designing great data products: The Drivetrain Approach: A four-step process for building data products, URL https://www.oreilly.com/ideas/drivetrain-approach-data-products

Kim M (2016) Five steps for success. In: Perspectives on data science for software engineering, Williams L, Menzies T, Zimmermann T (eds), Morgan Kaufmann, Cambridge, MA, p. 245–248, ISBN: 978-0-128042-06-9, DOI 10.1016/B978-0-12-

804206-9.00044-1
Kowalczyk M, Buxmann P (2014a) Big Data und Informationsverarbeitung in organisatorischen Entscheidungsprozessen. WIRTSCHAFTSINFORMATIK 56(5):289–302, DOI 10.1007/s11576-014-0430-6
Kowalczyk M, Buxmann P (2014b) Big Data und Informationsverarbeitung in organisatorischen Entscheidungsprozessen. Wirtschaftsinformatik 56(5):289–302, Springer
Loukides M (2010) What is data science?, URL `https://www.oreilly.com/ideas/what-is-data-science`
McElheran K, Brynjolfsson E (2017) The rise of data-driven decision-making is real but uneven. IEEE Engineering Management Review 45(4):103–105, DOI 10.1109/EMR.2017.8233302
Miller HG, Mork P (2013) From Data to Decisions: A Value Chain for Big Data. IT Professional 15(1):57–59, DOI 10.1109/MITP.2013.11
Osterwalder A, Pigneur Y (2013) Business model generation: A handbook for visionaries, game changers, and challengers. Wiley&Sons, New York. ISBN: 978-0-470876-41-1
Patil DJ (2012) Data Jujitsu: The art of turning data into product: Smart data scientists can make big problems small, URL `http://radar.oreilly.com/2012/07/data-jujitsu.html`
Provost F, Fawcett T (2013a) Data Science and its Relationship to Big Data and Data-Driven Decision Making. Big data 1(1):51–59, DOI 10.1089/big.2013.1508
Provost F, Fawcett T (2013b) Data science for business: [what you need to know about data mining and data-analytic thinking], 1st edn. O'Reilly, Beijing. ISBN: 978-1-449361-32-7
Schmarzo B (2016) Big data MBA: Driving business strategies with data science. Wiley, Indianapolis, Indiana. ISBN: 978-1-119181-38-5
Schmid B, Baars H (2016) Programmierer oder Statistiker? Anforderungen an Data Scientists und Data Analysts - Ergebnisse einer Studie. Online Themenspecial Predictive & Advanced Analytics, URL http://www.sigs.de/publications/bi/2016/Predictive$_A$dvancedAnalytics/schmid$_b$aars$_B$IS$_O$TS$_2$016.pdf
Stockinger K, Stadelmann T (2014) Data Science für Lehre, Forschung und Praxis. HMD Praxis der Wirtschaftsinformatik 51(4):469–479, DOI 10.1365/s40702-014-0040-1
Wirth R, Hipp J (2000) CRISP-DM: Towards a standard process model for data mining. Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining, p. 29–39