DEVELOPMENT AND EVALUATION OF A FRAMEWORK TO SUPPORT EXPERIMENTAL RESEARCH ON DATA ANALYTICS

Master Thesis
of

MAX DARMSTADT

July 17, 2023

Matriculation Number
1820000

Submitted at the Chair of Enterprise Systems
University of Mannheim

Reviewer: Prof. Dr. Hartmut Höhle
Supervisor: Jan Schilpp

Contents

List of Figures

List of Tables

List of Abbreviations

DSR	Design Science Research
IEEE	Institute of Electrical and Electronics Engineers
IS	Information Systems
IT	Informationstechnologie

Abstract

Over the past few years, data analytics has become increasingly important for companies across all industries. With the massive amount of data that is now available, companies can use data analytics to gain valuable insights into consumer behavior, market trends, and internal operations, among other things. As a result, data analytics has become a critical tool for companies looking to gain a competitive edge in today's rapidly evolving business environment. However, while data analytics has become an essential tool for businesses, there has been relatively little research done in the area of behavioral research. Specifically, there is a lack of research on the decision-making process involved in data analytics, and how individuals and organizations use data analytics to inform their decisions. One of the major challenges in conducting research in this area is the high cost of developing custom applications for each study. The development of such applications can be time-consuming, expensive, and often requires specialized expertise. To address this challenge, this master thesis develops a generic application that streamlines the process of conducting studies in the field of data analytics. This application enables researchers to design, conduct, and analyze studies more efficiently and cost-effectively, allowing them to explore the field in greater depth. This will be accomplished by using the design science research approach. Firstly, the problem of a lack of behavioral research in data analytics is identified. Then, the objectives for a solution are defined through a literature review and the use of requirement engineering to gather requirements for the application. Next, the application is design, implemented prototypically and its functionality demonstrated. Finally, solution is evaluated through the usages of the requirements.

1 Introduction

1.1 Background and Motivation

The introduction and widespread usage of computers has proven to be disruptive for all industries. Entire industrial sectors have been reshaped, whole professions made obsolete and new career opportunities have been created. This shift towards the adaption of Informationstechnologie (IT) has been necessary for businesses to stay competitive in the fast changing economic environment of the twenty-first century. Nowadays, the digitalization of organizations is viewed as a prerequisite for a successful business, rather than being an endeavored state and computers are indispensable for all industries. This digital revolution enabled the emergence of the widespread creation and collection of data. The amount of data generated globally is rising yearly (Seagate, 2018) and the pressure to use these data volumes effectively in order to gain a business advantage rises.

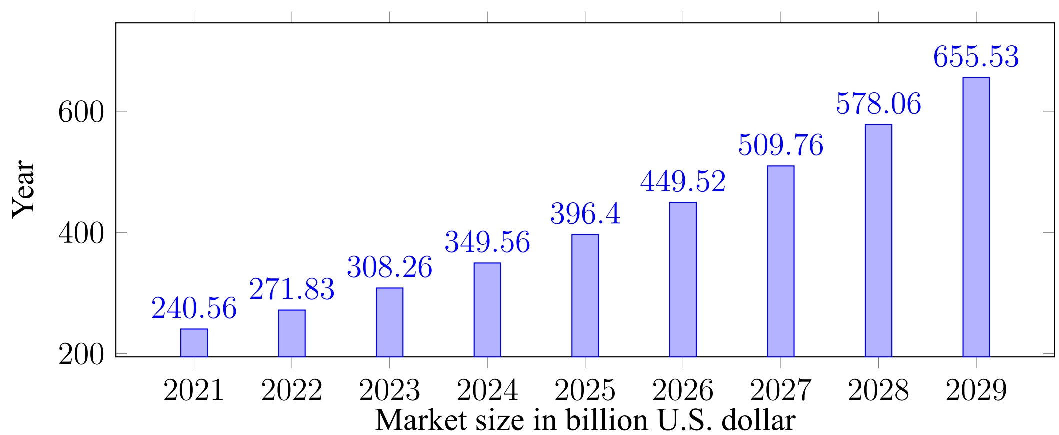

Figure 1: Size of the big data analytics market worldwide from 2021 to 2029 (Statista, 2022)

This new trend, often coined "big data" after the fact that never before seen amounts of data are generated and are available for processing, enables completely new business areas. This is reinforced, among other things, by the fact that many companies already view their data as a primary business asset (Redman, 2008). Simultaneously, the emergence of big data promises to completely reshape the decision-making process of traditional businesses through the adoption of data analytics. Although sales in the area of big data have risen significantly over the past years (BIS Research, 2018, Bitkom, 2018) and businesses already view big data as an important information technology trend (Bitkom, 2017) a lot of organizations struggle to effectivly utilize their data. Some 84% of industry-leading companies in the United States and around the world were already investing in big data analytics in 2019, according to their own statements, which only underlines the importance of data analytics for decision-making in the economy. (Statista, 2019). This is also reinforced by the market for big data analytics worlwide expected to more then double in size in the next 6 years (figure 1).

1.2 Applied & Theoretical Research Problem

In their article, Amankwah-Amoah and Adomako study the influence of big data usage on business failure. They come to the conclusion that the mere possession of big data as an asset has no positive effects on an organization (Amankwah-Amoah and Adomako, 2019) and that in order to prevent business failure, big data must be used effectively (Amankwah-Amoah and Adomako, 2019). Conducting research to resolve the underlying factors hindering the efficient utilization of data analytics in this particular context holds therefore crucial significance. However, although decision-making in data analytics has recently attracted scholars' attention (L. Chen et al., 2022) there is still a lack of research on non-technical aspects holding back the utilization of data analytics. Section 3 of this thesis uses the information-value-chain to look at the state of research in data analytics. Specifically, a literature review on boundaries and conflicts in the field of Data Analytics is conducted. The results of this literature review show that there is a lack of research on non-technical aspects of the information-value-chain. These literature gaps mainly consist of areas like decision-making and behavioral research. These results have been confirmed by studies in the past focussing on related fields (Trieu, 2017), which might indicate a persistant issues. This only becomes more apparent as new technologies, which overlap with data analytics like machine learning and artificial intelligence get more widespread. These "black-box"[1] technologies are already alternatives to human decision-making (Krakowski et al., 2023b). Moreover other studies confirm that aspects like company culture, business models and the overall commitment and strategy of organiza-

[1] Technologies whose exact internal sequence can hardly or not at all be explained, which therefore are acting like a "black-box"

tions have a big impact on the effectiveness of data analyitics (Holsapple et al., 2014). This lack of research on non technical regards could become a huge issue in the future, specifically for the decision-making process as firms' top priorities focus more and more on big data analytics for strategic decision making (Ghasemaghaei, 2019).

Furthermore, it is indicated in section 3 that the current state of research specifically lacks a variety of experiments conducted to confirm the validity of frameworks and hypothesis. The experiment being a particularly important tool for investigating causal relationships in research (Gniewosz, 2011). In addition other means of collecting information like inquest questionnaires, which are probably the most frequently used form of obtaining information (Mummendey and Grau, 2014) in quantitative research are not always the most suitable method. The behavior of people, for example, which comprises the literature gap found, can be better assessed by means of observational studies or experiments (Gniewosz, 2011). While prior research has conducted behavioral research of data analytics with surveys and case studies, little or no attention has been paid to the verification of research results and hypothesis through experiments. This general lack of experimental research in certain areas connects the applied business problem of better utilizing data analytics in organizations to the theoretical problem of lack of experimental research generally found in data analytics (refering to the results of the literature search in section 3). Improving the experimental research process in the field of data analytics could thereby significantly improve the future state of knowledge on said topic, whilste also allowing organizations to succeed in the fast paste economical environment of the digital-age.

1.3 Objective and Expected Contribution

The objective of this thesis is to improve the experimental research process in the field of behavioral research in data analytics. This is done through the following three objectives: (1) the review of prior research on data analytics and their methodological procedure (2) the development of an artefact which improves the research process in the field of behavioral research in data analytics (3) the validation of the artefact through the exemplary realization of a study in said field utilizing said artefact. In order to accomplishe the creation of this artefact the Design Science Research (DSR) methodology is used. This methodology contains six steps, *Identification of the Problem*, *Definition of Objectives for a solution*, *Design and Dev of artefacts*, *Demonstration of the Artifact*, *Evaluation of the solution* and *Communication* (Peffers et al., 2006, Dresch et al., 2015). The *identification of the problem* is conducted in section 3 through a literature search. To be more specific the current state of research in the field of boundaries and conflicts that limit the utilization of data analytics is investigates. As stated before this results in the identification of a gap in literature in the field of behavioral research in data analytics. After identifying the problem, the next step of the DSR Methodology is the *definition of objectives for a solution*. This is accomplished by specifying requirements for the final artifact using the current state of literature in data analytics and the *requirement engineering* approach. For this purpose, a second literature search is utilized to identify literature in the field of data analytics that uses experiments or whose research object would in principle have permitted the use of experiments. These insights are then used to conceptualize, analyze and validate requirements for the final artefact (Sommerville, 2011, Alain Abran, James W. Moore, 2004). As the name of the next step suggests the artefact is designed and implemented in the *design and dev of artefacts* phase using the requirements specified in the last step. Subsequently, the resulting artefact is then demonstrated in the *Demonstration of the Artifact* step. The *Evaluation of the solution* phase then implements a real experimental study in data analytics as an example to evaluate the artefact and its benefits for the experimental research process. This examplary implementation is then also used to validate the afformentioned requirements. The last step of the DSR framework, which focusses on communicating the results to its stakeholders, is ensured by this thesis itself.

The practical contribution of this thesis to research is twofold. On the one hand an artefact is created which accelerates research in the field of data analytics, through the improvement of the research process. On the other hand meta-knowledge about the research process itself is created, which not only improves the conduct of research through said artefact, but can also be used in off-topic areas beyond the use cases of this thesis.

In summation, the goal of this thesis is to develop an artifact that improves the experimental research process in behavioral research in data analytics. For this purpose, the DSR method is used to create an artifact. The requirements for this artifact are established using past experimental and non-experimental studies in data analytics and the requirement engineering approach. The artifact is then validated by implementing a real study as an example, thereby demonstrating the improvements to the experimental behavioral research process in data analytics.

2 Theoretical foundations

This section represents the theoretical fundamentals of this elaboration by defining the term “data analytic”, the concept of the “information value chain”, the DSR methodology, the *requirement engineering* approach and the experimental research process in general.

2.1 Data Analytics

The term “data analytics” originated in the early 2000s and describes an interdisciplinary field that combines areas such as statistics, machine learning, pattern recognition, system theory, operations research and artificial intelligence (Runkler, 2020). It can be generally defined “[...] as the application of computer systems to the analysis of large data sets for the support of decisions.” (Runkler, 2020). This definition showcases the broadness of the topic, as most computer systems process some amount of data and thus theoretically allow for some kind of decision making. Due to this broad definition, data analytics can cover slightly different subject areas depending on the context it is discussed in. In this elaboration, data analytics refers to the processing of large amounts of data, also referred to as “big data”, through mathematical procedures or machine learning methods with the goal of creating new knowledge. Subsequently, processes that merely prepare or show data are not considered data analytics, but only processes that process data in such a way that new knowledge can be derived from it. This distinction is made to differentiate data analytics from traditional data processing areas like business intelligence. The goal of data analytics, as is discussed in this thesis, is to retrieve some kind of previously unknown knowledge from a set of data. This process can be generally described using the “information value chain” model. In their research, Abbasi et al. analyze this model in the context of big data in an effort to create an inclusive research agenda for big data in information system research (Abbasi et al., 2016).

2.2 Information Value Chain

The information value chain (figure 2) is a set of phases that define the transformation of raw data to information and eventually into knowledge. “Data” describes raw facts without any structuring. Once organized, the processed data represents “information”. This

Figure 2: Information Value Chain

"information" is then used to find patterns and draw conclusions. At this time, the information becomes knowledge (Fayyad et al., 1996a), Fayyad et al., 1996b. This knowledge is then used to make "decisions" and take corresponding "actions" (Sharma et al., 2014). Each phase of the information value chain also includes a different set of technologies and methodologies. For example, the "data" phase contains technologies and actions regarding the basic storage of data like database systems or data warehouses (Abbasi et al., 2016). The conventional version of this information value chain represents an approach that generally explains the processing of data. The main steps of this information value chain are also applicable for big data (Abbasi et al., 2016). This general structure of processing data is also supported by literature from the data analytics field (Runkler, 2020). In addition, the information value chain contains the further phases "decisions" and "actions", which deal with the influence of the processed data. These phases reflect the impact of data analytics, since data analytics is primarily a technology for the decision-making process (Runkler, 2020). For this reason, the information value chain is a suitable model to structure different phases in the processing of data in the context of data analytics.

2.3 Design Science Research Methodology

In fundamental terms, design science is a research approach that aims to develop and validate science-based design knowledge and guide research to problem solving (Hevner et al., 2004, Dresch et al., 2015). The goal of DSR is to gain prescriptive knowledge about the composition of various artifacts, including software, methods, models, and concepts. This particular design knowledge facilitates a systematic and scientific approach to the design of future projects. The design process and its practical implementation generate design-oriented knowledge, enriching the existing knowledge in DSR (Hevner et al., 2004). Thus the result of design-science, especially in Information Systems (IS), is the creation of an effective IT artefact which deals with a certain problem (Hevner et al., 2004), making the DSR a suitable approach for this thesis. Nevertheless, the exact activities of the design science model may differ from author to author to some extent (Fulcher and Hills, 1996). This thesis aligns itself with the phases and steps outlined by Peffers et al. in their 2006 article "The design science research process: A model for producing and presenting information systems research". In their article Peffers et al. analyze literature that implements design science in order to create a generally accepted process for research in IS (Peffers et al., 2006). As a result of their work, they describe the design science research approach using the following six steps: (Peffers et al., 2006)

1. ***Identification of the Problem***: In this phase, the problem fundamentally solved by the resulting artifact is described. At the same time the value of resulting solution is justified.

2. ***Definition of Objectives for a solution***: This phase focuses on the specific goal of the solution and how its solving can be measured. The objectives for the solution can usually be divided into qualitative and quantitative characteristics.

3. ***Design and Dev of artefacts***: This step covers the specific functional design and scope of the artifact and how it is to be achieved. The phase covers both the conceptual design and the specific practical implementation of the solution.

4. ***Demonstration of the Artifact***: The focus of this phase is to demonstrate that the artifact implemented in the previous step can actually solve the initial problem. For this purpose, various methods such as experiments, case studies or other test setups can be used.

5. ***Evaluation of the solution***: This step evaluates the artifact and measures how well it contributes to a solution to the problem, therefore this phase tests whether the developed artefact is a valid solution to the research problem. If the artifact does not meet the corresponding requirements, the process is iterated back to step number three. If the artifact meets all requirements and contributes significantly to the solution of the problem, the last step is conducted.

6. ***Communication***: The last step ensures that the developed artefact and its value is communicated to the scientific community. The goal of this phase can be achieved through a number of different ways. The choice for the right way of communicating the results thereby relies on the specific topic and problem at hand.

2.4 Functional and non functional Requirements

As a means to create any artefact, the determination of requirements are important(Seacord et al., 2003). Requirements can be classified according to ISO/IEC 25000, respectively the quality model from ISO/IEC 25010, as quality criteria for software and systems.(ISO/IEC 25010, 2011). The Institute of Electrical and Electronics Engineers (IEEE) defines requirements as:

> ***"(1) A condition or capability needed by a user to solve a problem or achieve an objective. (2) A condition or capability that must be met or possessed by a system or system component to satisfy a contract, standard, specification, or other formally imposed documents. (3) A documented representation of a condition or capability as in (1) or (2). See also: design requirement; functional requirement; implementation requirement; interface requirement; performance requirement; physical equirement."*** IEEE, 1990

According to these definitions, requirements can be generally defined as properties that need to be met in order to achieve an objective. For this reason, the requirements for the artefact are derived from properties and characteristics of past studies in the field of data analytics. In addition, general requirements for experimental research are also taken into account. These requirements are further divided into functional and non-functional requirements. A functional requirement describes a function that a system or system component must be able to perform (IEEE, 1990). An example of a functional requirement would be the calculation of a pricetag in euros and in dollars. Non-functional characteristics describe on the other hand the behavior of a system (Seacord et al., 2003) and go thereby beyond the functional characteristics. Thus functional requirements describe what a system must be able to do and non-functional requirements describe how this should be done. Non-functional requirements also often describe the quality of the individual functions and can influence several other requirements (Balzert, 2011). An example of non-functional properties would be that the conversion from euros to dollars must be performed in “a few seconds”.

2.5 Requirement Engineering

In order to define an objective for a possible solution to the research question requirements for the created artefact must be engineered. In order to accomplishe that the *requirement engineering* approach for analysis and evaluation of requirements is utilized (Alain Abran, James W. Moore, 2004, Sommerville, 2011). This approach has been shown to clearly contribute to software project successes in the past (Hofmann and Lehner, 2001) and is therefore a suitable approach to define the objectives for a solution. The exact individual phases and steps of the *requirement engineering* approach can vary from source to source and use case to use case. In general, however, all steps fall into one of three main categories, the *Requirements Elicitation*, *Requirements Specification* and *Requirements Validation* (Alain Abran, James W. Moore, 2004, Sommerville, 2011, Fernandes et al., 2009). In the first step, possible requirements and use cases are collected via a variety of sources like analyses, surveys, literature or interviews (Sommerville, 2011). This thesis utilizes a literature review in order to discover requirements. In the next step, the requirements are then specified and categorized. An important distinction being the difference between functional and non-functional requirements. In the *Requirements Validation* step, the elicited requirements are then tested for their validity. This phase emphasis the reviewing of the requirements in order to find out whether these requirements are actually representive of the desired artefact (Sommerville, 2011). This is accomplished througt Validity, Consistency, Completeness, Realism, and Verifiability checks in conjunction with protoyping and testing the requirements (Sommerville, 2011).

2.6 Experiments and their research methodology

In the context of quantitative empirical research, the starting point for experiments is a preliminary theoretical work that results in the formulation of hypotheses. These hypotheses are often found in the form that certain conditions or sets of conditions are put forward as influencing factors for a phenomenon and causal assumptions are made (Gniewosz, 2011). Experimental research and field research are often based on these causal assumptions. However, the procedure for testing these assumptions is different. In field research, phenomena (dependent variable) are supposed to be predicted by measuring influencing factors (independent variable). Independent variables or influencing factors could therefore be for example "good teaching" whereas the resulting phenomena or dependent variable would be "good grades" for the students. This posterior approach raises problems in deriving causal conclusions in many correlational studies, because effects of third variables can never be completely excluded. To address this problem, in experimental designs, the expression of the independent variable is not measured but generated by experimental manipulations. This means that the expression of the independent variable is triggered by a specific experimental design. Another key feature of experiments is the elimination of confounding variables. Confounding variables are influencing variables that can also affect the dependent variable and thus affect the relationship between the independent and dependent variable. This influence must be neutralized or controlled for the planned experiment in order to identify the effects of the independent variable (Gniewosz, 2011). There are two types of confounding variables, those that depend on the individual participants of the experiment and those that arise from the experimental design itself (Gniewosz, 2011). In order to neutralized these confounding variables a number of different approaches exist, which are listed in table 1.

If these confounding variables are fully removed an experiment can be conducted where the test subjects assignment to the experimental conditions is completely determined. If these confounding variables are not fully removed the effects of other influences cannot be completely ruled out.

Furthermore, there are other characteristics by which experiments are categorized. For instance, in individual experiments several dependent/independent variables can be tested as well as only one single variable. Single factorial and Multifactorial Designs describe experiments which explore single and multiple dependent variables respectively. If the independent variable is subject ot the experiment on the other hand, the experiment categories as an univariate or multivariate a design. On top of that the distinction between "between subject" and "within subject" designs are made. "Within subject" designs describe experiments where the

Confounding variable	Description
Parallelization of confounding variables of the test subjects	The confounding variables of all test subjects are measured beforehand. the participants are then divided among the experimental conditions in such a way that the mean values of the confounding variable are approximately equal in all experimental conditions.
Randomization of confounding variables of the subjects	If it is not known which confounding variables might be relevant, the test subjects can be randomly assigned to experimental conditions, which from a probabilistic point of view, results in an equal distribution over the experimental conditions.
Removing confounding variables of the experimental situation	If it is known which experimental conditions triggers the confounding variables, these conditions should be eliminated in the experimental situation
Keeping constant confounding variables of the experimental situation	All confounding variables for all test subjects should have the same form
Random variation of confounding variables of the experimental situation	If the confounding variables cannot be removed or kept constant they can be randomly assigned over the whole experimental process. The confounding variable would vary randomly between the experimental conditions and thus not be able to mask the effect of the manipulated independent variable.
Control group	Another option to rule out confounding variables is the usage of a control groups
Control of expectation effects	Not telling participants and personnel performing the experiment the final goal of the experiment can counteract expectations, which could otherwise influence the outcome as confounding variables

Table 1: Counter measures for confounding variables in experimental research (Gniewosz, 2011)

test subjects are tested multiple times with different experimental setups, whereas “between subject” experiments test different subjects with different experimental setups (Gniewosz, 2011).

3 Identification of the Problem

3.1 Previous Studies and Gaps in the Literature

In order to identify constrains on the research on data analytics a literature search is conducted. The main objective of it is to analyze the existing literature to find research gaps, particularities and interrelationships between literature. This is supposed to give insights into the current state of research and to find out which part of the research process on data analyitcs still has room for improvements. The literature search itself is conducted in the field of boundaries and conflicts that might hinder the usage of data analytics. As already presented, data analytics has become an extremely important topic for companies. Therefore, the identification and resolution of obstacles that limit the use of data analytics is fundamentally important for companies. For this reason and to expand the scope of the literature search, in addition to reviewing relevant literature on data analytics, an exploration of literature deaing with boundaries in data analytics is pursued. Consequently, relevant literature was identified and reviewed. Afterwards, the identified literature was categorized and analyzed. Initially, it was assumed that the topic of data analytics lies both in the field of information systems and business (Abbasi et al., 2016, Levina and Vaast, 2005). For this reason, the literature search was mainly conducted in literature databases that focussed on these topics. Table 2 shows the databases that were used.

Online database	Subject Focus
ABI/INFORM Collection	Business and management
Business Source Premier	Accounting, business, economics, management
EconBiz	Business and economics
ProQuest One Business	Business
AIS Electronic Library	Informatics
MIS Quarterly Website	Business informatics
Web of Science	Multiple databases that provide access to different academic topics
Google Scholar	Web search engine for scholarly literature across an array of disciplines

Table 2: Databases Used in the Literature Search

The literature search was conducted using a keyword search. The used keywords consist of phrases like “Data Analytics”, “Data” and “Boundary”. A full list of keywords that were used is included in the appendix.

In order to ensure the quality of the identified literature initially, only publications from certain journals were considered. These journals consist of the *Senior Scholars' Basket of Journals* and the *UT Dallas Top 100 Business School Research Rankings*. The former includes journals in the area of information systems and the later includes journals in the area of business administration. A full list of included journals is listed in the appendix. Furthermore, only peer-reviewed articles were taken into account. This was done to ensure the quality of the found publications and to additionally exclude book reviews, editorials and opinion statements. Moreover, other 'non-scholarly' texts or publications that did not meet scientific requirements were also not considered in the search. Secondly, the abstracts of the particular articles were inspected to narrow the search further. Consequently, literature that did not meet the topic of boundaries in data analytics was excluded from the search. The literature found in the search was then used for a backward and forward search. During a backward search, all cited sources of an article are examined and during a forward search all the literature that cites the original article is examined (Webster and Watson, 2002). The backward search was conducted using Google Scholar. In addition to this, articles from other journals were, in a second step, reviewed and included as well if they met the scientific requirements, were officially published and relevant to the topic. This process yielded 35 research publications. The results were then assigned to different phases of the aforementioned information value chain, their content best represents. This was done to find literature gaps in the general process of data processing. Additionally, the identified literature was categorized by their research methodology in order to find patterns and similarities in the literature. As stated before, the information value chain consists of the phases "data", "information", "knowledge", "decisions" and "actions". The found literature was assigned to these phases, in order to structure and analyze the findings. By mapping the literature found, parts of the data processing process that are over- or under-represented may become visible. From this, conclusions can be drawn about the current state of research. Furthermore, the categories "overspanning" and "other" were introduced to represent literature that either fits multiple phases of the information value chain or none. Using this method leads to the results shown in the "First Search" column of table 3.

Information Value Chain	First Search	Additional Search
Data	4	
Information	3	
Knowledge	21	
Decisions	4	0
Actions	0	0
Overspanning	0	3
Other	3	
Total	35	3

Table 3: Results Assigned to the Information Value Chain

Table 3 shows an overabundance of literature that got assigned to the "knowledge" phase of the information value chain. The significantly fewer entries for the other phases indicate less research on these sub-parts of the information value chain. However, it cannot be concluded that this underrepresentation is due to the fact that these phases are less relevant in the context of data analytics. For this, more literature would have to exist confirming that these areas are less important for data analyitics or less prone to boundaries. The underrepresentation of the phases "data" and "information" could also be explained by the fact that these phases are more technology driven and therefore less researched in a bigger organizational data analytics context. In fact, the corresponding literature, which was assigned to these phases mainly consists of publications researching the technical possibilitties and application of data. Their main research object does not directly consist of any broader topics for companies or the application of Data Anayltics. The focus of this literature is largely to overcome technical hurdles or to show how individual technical functions can be implemented. This circumstance is particularly worrying in the context of data analytics, since Amankwah-Amoah and Adomako (Amankwah-Amoah and Adomako, 2019) has already established that the mere existence of data does not yet have any added value for the organization. Research in these very technical areas, such as how data is generated in the first place, is therefore fundamentally important, but without further research it does not contribute significantly to the effective use of data analytics. Nonetheless, in total, seven individual publications could be found that fit into these two phases. In addition, these two phases ("data" and "information") are mostly considered together in the further elaboration, since the literature which was assigned to these phases lies thematically very closely together.

Only four publications were assigned to the “decisions” phase and none to the “actions” phase. These results in particular call into question if the topic of behavioral research in data analytics has been extensively researched. The reason for this is the fact that data analytics is primarily a decision support method (Runkler, 2020). Therefore, an overabundance of literature delineating the decision-making process of data analytics should likely exist. This is compounded by the fact that no literature could be found that addressed overspanning issues, as no overarching theories could exist for an insufficiently studied topic. In order to ensure that the ratio of the literature found is based on the research state and not on the keyword search being biased in any way, a second literature search was conducted focussed on finding more literature that could be assigned to the “decisions” or “actions” phase. This was only done for these phases as these two are most relevant in the context of data analytics and because, in total, the least literature could be assigned to them (viewing “data” and “information” together). This second keyword search was conducted with the goal of finding more literature that could be assigned to the phases “decisions” and “actions”. Therefore, a new set of keywords including “decision”, “decision making” and “action” were added to the existing set of keywords. The full list of keywords is included in the appendix. Furthermore, the abstracts were examined with an emphasis on the aforementioned goal. The results of this second keyword search are represented in the “Additional Search” column of table 3. A total number of three additional publications were identified using this second search. These three publications were all assigned to the “overspanning” category. Consequently, no additional literature that could be assigned to the phases “decisions” or “actions” could be identified. This further indicates the fact that the topic of boundaries in data analytics is not researched extensively.

A total number of 38 publications were identified in these two searches and analyzed further.

In order to further analyze the literature and to potentially draw further conclusions, the found literature was also categorized regarding the research method that was used. This categorization is presented in table 4.

The distribution presented in table 4 show significant discrepencies in the number of research methods used. Methods such as case studys and surveys are used more frequently

Research approach	Method	Number
Qualitative (22)	Case Study	13
	Interviews	4
	Experiments	2
	Observation	3
Quantitative (16)	Survey	12
	Data Analysis	6

Table 4: Research Approach Used in the Literature

than average in contrast to other methods. The least frequently used methodology is the experiment. This fact further indicates an insufficient exploration of the field of behavioral research in data analytics, as experiments are the most suitable method for researching the behavior of people (Gniewosz, 2011). In this context experiments are a particularly important tool for investigating causal relationships in research (Gniewosz, 2011).

As already indicated in section 2.2, much of the literature found could be assigned to the "knowledge" phase of the information value chain. In addition, literature can be identified which was assigned to the information value chain phases "data" and "information". This literature mostly consists of technical publications, whose main goal is to research areas of application and advantages of data analytics. These publications, are for the most part, only concert with very specific technical issues. These two research areas, which do not quite fit the context of data analytics, reinforce the assumption that this field is not researched extensively. Other indications that have already been mentioned are the lack of literature that overspans the topic as a whole and the lack of literature that can be assigned to the "decisions" and "actions" phases of the information value chain. This circumstance is supported by the fact that hardly any experiments could be found as an used method in the current literature, even though experiments should be an integral part of behavioral research in the field of data analytics (Gniewosz, 2011). This has not only created a blind spot in the current state of knowledge, but also a permanent limitation for new research, which lacks existing knowledge as a foundation for new science. In summary, the literature search indicates a gap in the literature on behavioral research on data analytics as well as a lack of experimental research in this area. Although based on the current state of knowledge, both behavioral research and experimental research should be an integral part of research in data analytics.

3.2 Applications (Anwendungen) for Behavioral Research

4 Definition of Objectives for a solution

4.1 Literature Review Studies in Data Analytics and General

4.2 requirements elicitation

4.2.1 Functional and non-functional requirements

Damit eine Anwendungsmodernisierung durchgeführt und eine modernisierte Architektur konzeptioniert werden kann, ist die Bestimmung der Anforderungen wichtig.(Vgl. Seacord et al., 2003, Kapitel 3) Anforderungen können nach ISO/IEC 25000, beziehungsweise dem Qualitätsmodell aus ISO/IEC 25010, als Qualitätskriterien an Software und Systeme klassifiziert werden.(Vgl. ISO/IEC 25010, 2011) Das IEEE definiert Anforderungen als eine Bedingung oder Eigenschaft, welche von einem System oder einer Systemkomponente erfüllt werden muss, um eine Problemstellung oder Zielsetzung eines Nutzers oder formalen Dokuments zu erfüllen.(Vgl. IEEE, 1990, S.62) Anforderungen können nach diesen beiden Definitionen als zu erfüllende Eigenschaften oder Qualitätskriterien einer Software oder eines Systems definiert werden. Aus diesem Grund werden die Anforderungen an den SAP Lean Catalog aus den von der SAP beschriebenen Produkteigenschaften und Merkmalen des SAP Lean Catalogs abgeleitet. Zusätzlich werden allgemeine Anforderungen an eine Softwaremodernisierung berücksichtigt, welche erfüllt sein müssen, um eine erfolgreiche Migration durchzuführen. Diese Anforderungen an den SAP Lean Catalog und an eine Softwaremigration im Allgemeinen werden weiter in funktionale und nichtfunktionale Anforderungen eingeteilt.

Eine funktionale Anforderung beschreibt eine Funktion oder Fähigkeit eines Systemes, die konkret von einem System oder einer Softwarekomponente durchgeführt werden können muss.(Vgl. IEEE, 1990, S.35) Ein Beispiel für eine funktionale Anforderung wäre die Berechnung des Bestellpreises in Euro und in Dollar. Nichtfunktionale Eigenschaften beschreiben hingegen Verhaltensweisen des Systems(Vgl. Seacord et al., 2003, Kapitel 3) und gehen damit über die funktionalen Eigenschaften hinaus. Damit beschreiben funktionale Anforderungen was ein System können muss und nichtfunktionale Anforderungen wie es funktionieren soll. Nichtfunktionale Anforderungen beschreiben außerdem häufig die Qualität der Funktionen und können mehrere andere Anforderungen beeinflussen.(Vgl. Balzert, 2011, S.109ff) Ein Beispiel für nichtfunktionale Eigenschaften wäre, dass die Umrechnung von Euro in Dollar in “wenigen Sekunden” durchgeführt werden muss.

4.3 Requirements analysis

Personen müssen aufgeklärt werden über das Experiment **Dresch.2011**

5 Design and Dev artefacts

5.1 System Architecture and Components

5.2 User Interface Design and Implementation

5.3 Prototype Development

6 Demonstration of the Artifact

7 Evaluation of the solution

7.1 Prototype Testing

7.2 Requirements validation

7.3 (App Performance and Usability / User Feedback and Satisfaction)

8 Conclusion

8.1 Summary of the Study

8.2 Contributions and Implications

8.3 Future Work and Recommendations

Bibliography

Abbasi, A., Sarker, S., & Chiang, R. H. L. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems*, *17*(2).

Abramson, C., Currim, I. S., & Sarin, R. (2005). An experimental investigation of the impact of information on competitive decision making. *Management Science*, *51*(2), 195–207.

Alain Abran, James W. Moore. (2004). *Swebok: guide to the software engineering body of knowledge*. IEEE Computer Society.

Allen, T. J., & Cohen, S. I. (1969). Information flow in research and development laboratories. *Administrative Science Quarterly*, *14*(1).

Amankwah-Amoah, J., & Adomako, S. (2019). Big data analytics and business failures in data-rich environments: An organizing framework. *Computers in Industry*, *105*.

Bag, S., Gupta, S., & Wood, L. (2022). Big data analytics in sustainable humanitarian supply chain: Barriers and their interactions. *Annals of Operations Research*, *319*(1), 721–760.

Balzert, H. (Ed.). (2011). *Lehrbuch der softwaretechnik: Entwurf, implementierung, installation und betrieb*. Spektrum Akademischer Verlag. https://doi.org/10.1007/978-3-8274-2246-0

Barrett, M., & Oborn, E. (2010). Boundary object use in cross-cultural software development teams. *Human Relations*, *63*(8), 1199–1221. https://doi.org/10.1177/0018726709355657

Bhatti, S. H., Hussain, W. M. H. W., Khan, J., Sultan, S., & Ferraris, A. (2022). Exploring data-driven innovation: What's missing in the relationship between big data analytics capabilities and supply chain innovation?. *Annals of Operations Research*, 1–26.

BIS Research. (2018). Umsatz mit big data im bereich healthcare weltweit nach anwendung in den jahren 2016 und 2025 (in milliarden us-dollar). https://de.statista.com/statistik/daten/studie/997352/umfrage/umsatz-mit-big-data-im-bereich-healthcare-nach-anwendung/

Bitkom. (2017). Welches sind die wichtigsten it-trends des jahres 2017? https://de.statista.com/statistik/daten/studie/675726/umfrage/die-wichtigsten-trends-in-der-itk-branche/

Bitkom. (2018). Umsatz mit big-data-lösungen in deutschland in den jahren 2016 und 2017 und prognose für 2018 (in milliarden euro). https://de.statista.com/statistik/daten/studie/257976/umfrage/umsatz-mit-big-data-loesungen-in-deutschland/

Carlile, P. R. (2002). A pragmatic view of knowledge and boundaries: Boundary objects in new product development. *Organization Science*, *13*(4), 442–455. https://doi.org/10.1287/orsc.13.4.442.2953

Chakraborty, I., Hu, P. J.-H., & Cui, D. (2008). Examining the effects of cognitive style in individuals' technology use decision making. *Decision Support Systems*, *45*(2), 228–241.

Chen, D. Q., Preston, S. D., & Swink, M. (2021). How big data analytics affects supply chain decision-making: An empirical analysis. *Journal of the Association for Information Systems*, *22*(5). https://doi.org/10.5194/gi-2016-11-RC2

Chen, L., Liu, H., Zhou, Z., Chen, M., & Chen, Y. (2022). It-business alignment, big data analytics capability, and strategic decision-making: Moderating roles of event criticality and disruption of covid-19. *Decision Support Systems*, *161*, N.PAG.

Chou, S.-W., & Chang, Y.-C. (2008). The implementation factors that influence the erp (enterprise resource planning) benefits. *Decision Support Systems*, *46*(1), 149–157.

Cross, R. L., & Parker, A. (2004). *The hidden power of social networks: Understanding how work really gets done in organizations*. Harvard Business School Press.

Currie, G., & Kerrin, M. (2004). The limits of a technological fix to knowledge management. *Management Learning*, *35*(1), 9–29. https://doi.org/10.1177/1350507604042281

Czekster, R. M., De Carvalho, H. J., Kessler, G. Z., Kipper, L. M., & Webber, T. (2019). Decisor: A software tool to drive complex decisions with analytic hierarchy process. *International Journal of Information Technology & Decision Making*, *18*(1), 65–86.

Demoulin, N. T., & Coussement, K. (2020). Acceptance of text-mining systems: The signaling role of information quality. *Information & Management*, *57*(1), N.PAG.

Donghyuk, S., Shu, H., Gene Moo, L., Whinston, A. B., Cetintas, S., & Kuang-Chih, L. (2020). Enhancing social media analysis with visual data analytics: A deep learning approach. *MIS Quarterly*, *44*(4), 1459–1492.

Dresch, A., Lacerda, D. P., & Antunes, J. A. V. (2015). Design science research. In *Design science research: A method for science and technology advancement* (pp. 67–102). Springer International Publishing. https://doi.org/10.1007/978-3-319-07374-3_4

Du, Wnyu, Pan, S. L., Xie, K., & Xiao, J. (2020). Data analytics contributes to better decision-making beyond organizational boundaries. *MIS Quarterly Executive*, *19*(2).

Elgendy, N., & Elragal, A. (2016). Big data analytics in support of the decision making process. *Procedia Computer Science*, *100*, 1071–1084. https://doi.org/10.1016/j.procs.2016.09.251

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996a). From data mining to knowledge discovery in databases. *AI Magazine*, *17*(3).

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996b). The kdd process for extracting useful knowledge from volumes of data. *Communications of the ACM*, *39*(11), 27–34. https://doi.org/10.1145/240455.240464

Fernandes, J., Machado, R.-J., & Seidman, S. (2009). A requirements engineering and management training course for software development professionals. *Software Engineering Education Conference, Proceedings*, 20–25. https://doi.org/10.1109/CSEET.2009.24

Fink, L., Yogev, N., & Even, A. (2017). Business intelligence and organizational learning: An empirical investigation of value creation processes. *Information & Management*, *54*(1), 38–56.

Foerderer, J., Kude, T., Schuetz, S. W., & Heinzl, A. (2019). Knowledge boundaries in enterprise software platform development: Antecedents and consequences for platform governance. *Information Systems Journal*, *29*(1), 119–144. https://doi.org/10.1111/isj.12186

Fosso Wamba, S., Queiroz, M. M., Wu, L., & Sivarajah, U. (2020). Big data analytics-enabled sensing capability and organizational outcomes: Assessing the mediating effects of business analytics culture. *Annals of Operations Research*, 1–20.

Fulcher, A. J., & Hills, P. (1996). Towards a strategic framework for design research. *Journal of Engineering Design*, *7*(2), 183–193. https://doi.org/10.1080/09544829608907935

Ghasemaghaei, M. (2019). Does data analytics use improve firm decision making quality? the role of knowledge sharing and data analytics competency. *Decision Support Systems*, *120*, 14–24.

Ghasemaghaei, M., Ebrahimi, S., & Hassanein, K. (2017). Data analytics competency for improving firm decision making performance. *The Journal of Strategic Information Systems*, *27*(1), 101–113. https://doi.org/10.1016/j.jsis.2017.10.001

Ghasemaghaei, M., Hassanein, K., & Turel, O. (2017). Increasing firm agility through the use of data analytics: The role of fit. *Decision Support Systems*, *101*, 95–105.

Gniewosz, B. (2011). Experiment. In H. Reinders, H. Ditton, C. Gräsel, & B. Gniewosz (Eds.), *Empirische bildungsforschung: Strukturen und methoden* (pp. 77–84). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-93015-2_6

Goodhue, D. L., Kirsch, L. J., Quillard, J. A., & Wybo, M. D. (1992). Strategic data planning: Lessons from the field. *16*(1).

Guven–Uslu, P., Blaber, Z., & Adhikari, P. (2020). Boundary spanners and calculative practices. *Financial Accountability & Management*, *36*(4), 439–460. https://doi.org/10.1111/faam.12266

Han, S., Datta, A., & Joshi, K. D., Chi, Lei. (2017). Innovation through boundary spanning: The role of it in enabling knowledge flows across technological and geographical boundaries. *International Journal of Knowledge Management*, *13*(4), 90–110. https://doi.org/10.4018/IJKM.2017100105

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, *28*(1), 75–105. Retrieved July 16, 2023, from http://www.jstor.org/stable/25148625

Hofmann, H., & Lehner, F. (2001). Requirements engineering as a success factor in software projects. *IEEE Software*, *18*(4), 58–66. https://doi.org/10.1109/MS.2001.936219

Holsapple, C., Lee-Post, A., & Pakath, R. (2014). A unified foundation for business analytics. *Decision Support Systems*, *64*, 130–141.

Hornby, A. S. (2015). *Oxford advanced learner's dictonary of current english* (9th). Cornelsen.

IEEE. (1990). Ieee standard glossary of software engineering terminology - ieee std 610.12-1990.

Inman, J. J., Winer, R. S., & Ferraro, R. (2009). The interplay among category characteristics, customer characteristics, and customer activities on in-store decision making. *Journal of Marketing*, *73*(5), 19–29.

Işık, Ö., Jones, M. C., & Sidorova, A. (2013). Business intelligence success: The roles of bi capabilities and decision environments. *Information & Management*, *50*(1), 13–23.

ISO Org. (2019). *Norm iso 9241*. Retrieved September 30, 2019, from https://www.iso.org/standard/77520.html

ISO/IEC 25010 (Ed.). (2011). *Systems and software engineering - systems and software quality requirements and evaluation (square)*. ISO/IEC. Retrieved March 23, 2021, from https://www.iso.org/standard/35733.html

Jha, A. K., Agi, M. A., & Ngai, E. W. (2020). A note on big data analytics capability development in supply chain. *Decision Support Systems*, *138*, N.PAG.

Ji, G., Yu, M., Tan, K. H., Kumar, A., & Gupta, S. (2022). Decision optimization in cooperation innovation: The impact of big data analytics capability and cooperative modes. *Annals of Operations Research*, 1–24.

Johnson, J. P., Lenartowicz, T., & Apud, S. (2006). Cross-cultural competence in international business: Toward a definition and a model. *Journal of International Business Studies*, *37*(4), 525–543.

Kankanhalli, A., Ye, H. (, & Teo, H. H. (2015). Comparing potential and actual innovators: An empirical study of mobile data services innovation. *MIS Quarterly*, *39*(3), 667–682. Retrieved July 6, 2023, from https://www.jstor.org/stable/26629625

Karhade, P., & Dong, J. Q. (2021). Innovation outcomes of digitally enabled collaborative problemistic search capability. *MIS Quarterly*, *45*(2), 693–718. https://doi.org/10.25300/MISQ/2021/12202

Kim, B. J., & Tomprou, M. (2021). The effect of healthcare data analytics training on knowledge management: A quasi-experimental field study. *Journal of Open Innovation: Technology, Market, and Complexity*, *7*(1), 60. https://doi.org/10.3390/joitmc7010060

Klingebiel, R., & Meyer, A. D. (2013). Becoming aware of the unknown: Decision making during the implementation of a strategic initiative. *Organization Science*, *24*(1), 133–153. Retrieved July 6, 2023, from http://www.jstor.org/stable/23362104

Korschun, D. (2015). Boundary-spanning employees and relationships with external stakeholders: A social identity approach. *Academy of Management Review*, *40*(4), 611–629. https://doi.org/10.5465/amr.2012.0398

Kotlarsky, J., Scarbrough, H., & Oshri, I. (2014). Coordinating expertise across knowledge boundaries in offshore-outsourcing projects: The role of codification. *MIS Quarterly*, *38*(2).

Krakowski, S., Luger, J., & Raisch, S. (2023a). Artificial intelligence and the changing sources of competitive advantage. *Strategic Management Journal (John Wiley & Sons, Inc.)*, *44*(6), 1425–1452.

Krakowski, S., Luger, J., & Raisch, S. (2023b). Artificial intelligence and the changing sources of competitive advantage. *Strategic Management Journal (John Wiley & Sons, Inc.)*, *44*(6), 1425–1452.

Lancelot Miltgen, C., Popovič, A., & Oliveira, T. (2013). Determinants of end-user acceptance of biometrics: Integrating the "big 3" of technology acceptance with privacy context. *Decision Support Systems*, *56*, 103–114.

Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To engage or not to engage with ai for critical judgments: How professionals deal with opacity when using ai for medical diagnosis. *Organization Science*, *33*(1), 126–148.

Lee, G. M., He, S., Lee, J., & Whinston, A. B. (2020). Matching mobile applications for cross-promotion. *Information Systems Research*, *31*(3), 865–891.

Lehmann, D., Fekete, D., & Vossen, G. (2016). *Technology selection for big data and analytical applications* (ERCIS Working Paper No. 27). Münster, Westfälische Wilhelms-Universität Münster, European Research Center for Information Systems (ERCIS). http://hdl.handle.net/10419/156084

Leidner, D. E., & Elam, J. J. (1995). The impact of executive information systems on organizational design, intelligence, and decision making. *Organization Science*, *6*(6), 645–664.

Levina, N., & Vaast, E. (2005). The emergence of boundary spanning competence in practice: Implications for implementation and use of information systems. *MIS Quarterly*, *29*(2).

Levina, N., & Vaast, E. (2006). Turning a community into a market: A practice perspective on information technology use in boundary spanning. *Journal of Management Information Systems*, *22*(4), 13–37. https://doi.org/10.2753/MIS0742-1222220402

Levina, N., & Vaast, E. (2008). Innovating or doing as told? status differences and overlapping boundaries in offshore collaboration. *MIS Quarterly*, *32*(2).

Li, H., Lu, K., & Meng, S. (2015). Bigprovision: A provisioning framework for big data analytics. *IEEE Network*, *29*(5), 50–56.

Li, M. (, Huang, Y., & Sinha, A. (2020). Data-driven promotion planning for paid mobile applications. *Information Systems Research*, *31*(3), 1007–1029.

Liberatore, M. J., & Stylianou, A. C. (1995). Expert support systems for new product development decision making: A modeling framework and applications. *Management Science*, *41*(8), 1296–1316. Retrieved July 6, 2023, from http://www.jstor.org/stable/2632787

Lindgren, R., Andersson, M., & Henfridsson, O. (2008). Multi-contextuality in boundary-spanning practices. *Information Systems Journal*, *18*(6), 641–661. https://doi.org/10.1111/j.1365-2575.2007.00245.x

Lucas, H. C. (1981). An experimental investigation of the use of computer-based graphics in decision making. *Management Science*, *27*(7), 757–768. Retrieved July 6, 2023, from http://www.jstor.org/stable/2630917

Lukyanenko, R., Parsons, J., Wiersma, Y. F., & Maddah, M. (2019). Expecting the unexpected: Effects of data collection design choices on the quality of crowdsourced user-generated content. *MIS Quarterly*, *43*(2), 623–647. https://doi.org/10.25300/MISQ/2019/14439

Lurie, N. H., & Mason, C. H. (2007). Visual representation: Implications for decision making. *Journal of Marketing*, *71*(1), 160–177.

Majchrzak, A., More, P. H. B., & Faraj, S. (2012). Transcending knowledge differences in cross-functional teams. *Organization Science*, *23*(4), 951–970. https://doi.org/10.1287/orsc.1110.0677

Mäkelä, K., Barner-Rasmussen, W., Ehrnrooth, M., & Koveshnikov, A. (2019). Potential and recognized boundary spanners in multinational corporations. *Journal of World Business*, *54*(4), 335–349. https://doi.org/10.1016/j.jwb.2019.05.001

Marchena Sekli, G. F., & de La Vega, I. (2021). Adoption of big data analytics and its impact on organizational performance in higher education mediated by knowledge management. *Journal of Open Innovation: Technology, Market, and Complexity*, *7*(4), 221. https://doi.org/10.3390/joitmc7040221

Mell, J. N., Knippenberg, D., Ginkel, W. P., & Heugens, P. P. M. A. R. (2022). From boundary spanning to intergroup knowledge integration: The role of boundary spanners' metaknowledge and proactivity. *Journal of Management Studies*, *Not yet published*. https://doi.org/10.1111/joms.12797

Minbaeva, D., & Santangelo, G. D. (2018). Boundary spanners and intra-mnc knowledge sharing: The roles of controlled motivation and immediate organizational context. *Global Strategy Journal*, *8*(2), 220–241. https://doi.org/10.1002/gsj.1171

Montgomery, A. L., Li, S., Srinivasan, K., & Liechty, J. C. (2004). Modeling online browsing and path analysis using clickstream data [Copyright - Copyright Institute for Operations Research and the Management Sciences Fall 2004; Document feature - references; tables; graphs; equations; Last updated - 2023-07-03; CODEN - MARSE5]. *Marketing Science*, *23*(4), 579–595. https://www.proquest.com/scholarly-journals/modeling-online-browsing-path-analysis-using/docview/212288093/se-2

Mueller, O., Fay, M., & vom Brocke, J. (2018). The effect of big data and analytics on firm performance: An econometric analysis considering industry characteristics. *Journal of Management Information Systems*, *35*(2), 488–509. https://doi.org/10.1080/07421222.2018.1451955

Mummendey, H., & Grau, I. (2014). *Die fragebogen-methode: Grundlagen und anwendung in persönlichkeits-, einstellungs- und selbstkonzeptforschung*. Hogrefe Verlag GmbH & Company KG. https://books.google.de/books?id=6aYVBAAAQBAJ

Nutt, P. C. (1998). How decision makers evaluate alternatives and the influence of complexity. *Management Science*, *44*(8), 1148–1166. Retrieved July 6, 2023, from http://www.jstor.org/stable/2634692

Pawlowski, S. D., & Robey, D. (2004). Bridging user organizations: Knowledge brokering and the work of information technology professionals. *28*(4).

Peffers, K., Tuunanen, T., Gengler, C., Rossi, M., Hui, W., Virtanen, V., & Bragge, J. (2006). The design science research process: A model for producing and presenting information systems research. *Proceedings of First International Conference on Design Science Research in Information Systems and Technology DESRIST*.

Peng, Y., & Sutanto, J. (2012). Facilitating knowledge sharing through a boundary spanner. *IEEE Transactions on Professional Communication*, *55*(2), 142–155. https://doi.org/10.1109/TPC.2012.2188590

Pil Han, S., Park, S., & Oh, W. (2016). Mobile app analytics: A multiple discrete-continuous choice framework. *MIS Quarterly*, *40*(4), 983–A42.

Po-An Hsieh, J. J., Rai, A., & Xin Xu, S. (2011). Extracting business value from it: A sensemaking perspective of post-adoptive use. *Management Science*, *57*(11), 2018–2039.

Popovič, A., Hackney, R., Coelho, P. S., & Jaklič, J. (2012). Towards business intelligence systems success: Effects of maturity and culture on analytical decision making. *Decision Support Systems*, *54*(1), 729–739.

Pour, M. J., Abbasi, F., & Sohrabi, B. (2023). Toward a maturity model for big data analytics: A roadmap for complex data processing. *International Journal of Information Technology & Decision Making*, *22*(1), 377–419.

Qiqi, J., Chuan-Hoo, T., Choon Ling, S., & Kwok-Kee, W. (2019). Followership in an open-source software project and its significance in code reuse. *MIS Quarterly*, *43*(4), 1303–1319.

Redman, T. C. (2008). *Data driven: Profiting from your most important business asset*. Harvard Business Review Press.

Richter, A. W., West, M. A., van Dick, R., & Dawson, J. F. (2006). Boundary spanners' identification, intergroup contact, and effective intergroup relations. *Academy of Management Journal*, *49*(6), 1252–1269. https://doi.org/10.5465/amj.2006.23478720

Runkler, T. A. (2020). *Data analytics*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-29779-4

Russell, S., Gangopadhyay, A., & Yoon, V. (2008). Assisting decision making in the event-driven enterprise using wavelets. *Decision Support Systems*, *46*(1), 14–28.

Ryder, B., Gahr, B., Egolf, P., Dahlinger, A., & Wortmann, F. (2017). Preventing traffic accidents with in-vehicle decision support systems - the impact of accident hotspot warnings on driver behaviour. *Decision Support Systems*, *99*, 64–74.

Seacord, R. C., Plakosh, D., & Lewis, G. A. (2003). *Modernizing legacy systems: Software technologies, engineering processes, and business practices*. Addison-Wesley. Retrieved February 21, 2021, from https://learning.oreilly.com/library/view/modernizing-legacy-systems/0321118847/

Seagate. (2018). Prognose zum volumen der jährlich generierten digitalen datenmenge weltweit in den jahren 2018 und 2025 (in zettabyte). https://de.statista.com/statistik/daten/studie/267974/umfrage/prognose-zum-weltweit-generierten-datenvolumen/

Sharda, R., Barr, S. H., & McDonnell, J. C. (1988). Decision support system effectiveness: A review and an empirical test. *Management Science*, *34*(2), 139–159. Retrieved July 6, 2023, from http://www.jstor.org/stable/2632057

Sharma, R., Mithas, S., & Kankanhalli, A. (2014). Transforming decision-making processes: A research agenda for understanding the impact of business analytics on organisations. *European Journal of Information Systems*, *23*(4), 433–441. https://doi.org/10.1057/ejis.2014.17

Simon, H. A. (1996). *The sciences of the artificial (3rd ed.)* MIT Press.

Sommerville, I. (2011). *Software engineering* (9th ed.). Pearson.

Song, P., Zheng, C., Zhang, C., & Yu, X. (2018). Data analytics and firm performance: An empirical study in an online b2c platform. *Information & Management*, *55*(5), 633–642.

Spiller, S. A., Reinholtz, N., & Maglio, S. J. (2020). Judgments based on stocks and flows: Different presentations of the same data can lead to opposing inferences. *Management Science*, *66*(5), 2213–2231.

Statista. (2019). *Big data projects success rate among corporations in the united states and worldwide as of 2019, by area*. Retrieved July 14, 2023, from https://www.statista.com/statistics/742935/worldwide-survey-corporate-big-data-initiatives-and-success-rate/

Statista. (2022). *Size of the big data analytics market worldwide from 2021 to 2029*. Retrieved July 15, 2023, from https://www.statista.com/statistics/1336002/big-data-analytics-market-size/

Tang, H., Liao, S. S., & Sun, S. X. (2013). A prediction framework based on contextual data to support mobile personalized marketing. *Decision Support Systems*, *56*, 234–246.

Trieu, V.-H. (2017). Getting value from business intelligence systems: A review and research agenda. *Decision Support Systems*, *93*, 111–124.

Trieu, V.-H., Burton-Jones, A., Green, P., & Cockcroft, S. (2022). Applying and extending the theory of effective use in a business intelligence context. *46*(1).

van Osch, W., & Steinfield, C. W. (2016). Team boundary spanning: Strategic implications for the implementation and use of enterprise social media. *Journal of Information Technology*, *31*(2), 207–225. https://doi.org/10.1057/jit.2016.12

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, *26*(2).

Wixom, B. H., & Watson, H. J. (2001). An empirical investigation of the factors affecting data warehousing success. *MIS Quarterly*, *25*.

Wook, M., Hasbullah, N. A., Zainudin, N. M., Jabar, Z. Z. A., Ramli, S., Razali, N. A. M., & Yusop, N. M. M. (2021). Exploring big data traits and data quality dimensions for big data analytics application using partial least squares structural equation modelling. *Journal of Big Data*, *8*(1). https://doi.org/10.1186/s40537-021-00439-5

Wright, R. T., Jensen, M. L., Bennett Thatcher, J., Dinger, M., & Marett, K. (2014). Influence techniques in phishing attacks: An examination of vulnerability and resistance. *Information Systems Research*, *25*(2), 385–400.

ZAVADSKAS, E. K., VAINIŪNAS, P., TURSKIS, Z., & TAMOŠAITIENĖ, J. (2012). Multiple criteria decision support system for assessment of projects managers in construction. *International Journal of Information Technology & Decision Making*, *11*(2), 501–520.

Zhang, Q., & Li, J. (2021). Can employee's boundary-spanning behavior exactly promote innovation performance? the roles of creative ideas generation and team task interdependence. *International Journal of Manpower*, *42*(6), 1047–1063. https://doi.org/10.1108/IJM-06-2019-0302

Zhao, Z. J., & Anand, J. (2013). Beyond boundary spanners: The 'collective bridge' as an efficient interunit structure for transferring collective knowledge. *Strategic Management Journal*, *34*(13), 1513–1530. https://doi.org/10.1002/smj.2080

Literature search: Boundaries and conflicts in data analytics

List of Keywords (First Search)

- Data Analytics
- Data AND Boundary
- Organization AND Data Analytics
- Big Data
- boundary
- boundary theory
- boundary spanning
- boundary objects
- boundary spanner

List of Keywords (Second Search)

- Data Analytics
- Data AND Boundary
- Organization AND Data Analytics
- Big Data
- boundary
- boundary theory
- boundary spanning
- boundary objects
- boundary spanner
- Decision
- Decision Making
- Action

Senior Scholars' Basket of Journals

- European Journal of Information Systems
- Information Systems Journal
- Information Systems Research
- Journal of AIS
- Journal of Information Technology
- Journal of MIS
- Journal of Strategic Information Systems
- MIS Quarterly

UT Dallas Top 100 Business School Research Rankings

- The Accounting Review
- Journal of Accounting and Economics
- Journal of Accounting Research
- Journal of Finance
- Journal of Financial Economics
- The Review of Financial Studies
- Information Systems Research
- Journal on Computing
- MIS Quarterly
- Journal of Consumer Research
- Journal of Marketing
- Journal of Marketing Research
- Marketing Science
- Management Science
- Operations Research
- Journal of Operations Management
- Manufacturing and Service Operations Management
- Production and Operations Management
- Academy of Management Journal
- Academy of Management Review
- Administrative Science Quarterly

- Organization Science
- Journal of International Business Studies
- Strategic Management Journal

Affidavit

I hereby declare that I have developed and written the enclosed master thesis entirely on my own and have not used outside sources without declaration in the text. Any concepts or quotations applicable to these sources are clearly attributed to them. This master thesis has not been submitted in the same or a substantially similar version, not even in part, to any other authority for grading and has not been published elsewhere. This is to certify that the printed version is equivalent to the submitted electronic one. I am aware of the fact that a misstatement may have serious legal consequences.

I also agree that my thesis can be sent and stored anonymously for plagiarism purposes. I know that my thesis may not be corrected if the declaration is not issued.

Mannheim, July 17, 2023

Max Darmstadt