DEVELOPMENT AND EVALUATION OF AN [CUSTOMIZABLE MOBILE] APPLICATION FOR BEHAVIORAL RESEARCH IN DATA ANALYTICS

Master Thesis
of

MAX DARMSTADT

June 29, 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

IEEE	Institute of Electrical and Electronics Engineers
ISO	International Organization for Standardization

Abstract

1 Introduction

1.1 Background and Motivation

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.

In the past couple of years this led many companies and researchers into spending an above-average share of resources into technical areas like data extraction, preparation or harmonization. While this has laid a very good foundation for the topic of data analytics both in businesses and in research, it has also led to topics relating to the use of data beyond technical aspects being neglected.

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.

Looking at how companies utilize their available data an above-average share of resources is dedicatded to actions like data extraction or preparation, whilste areas beyond these more technical processes hardly get any attention. In the past couple of years this led to big investments

1.2 Objective and methodology of the work

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 thesis develops a generic application that streamlines the process of conducting studies in the field of data analytics.

1.3 Contribution and Scope of the Study

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.

2 Theoretical foundations

This section represents the theoretical fundamentals of this elaboration by defining the terms “data analytic”, “information value chain”, and “boundaries and conflicts” as they are used in the context of this literature review.

2.1 Definition of terms

2.2 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 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. In summary, 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 literature review, 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).

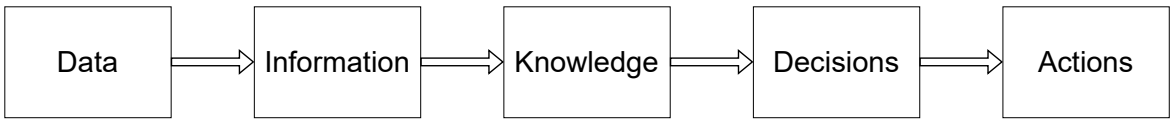

Figure 1: Information Value Chain

2.3 Information Value Chain

The information value chain (figure 1) 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 "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.

2.4 Boundaries and Conflicts in Organizations

This literature review uses the terms boundary and conflict interchangeably. In order to include as much literature as possible, the criteria for boundaries are kept very general. Prior to conducting the literature review, there was no formal definition of boundaries in the context of data analytics used for the selection of literature. Generally, boundaries are described as

"[...] a real or imagined line that marks the limits or edges of something and separates it from other things or places [...]" (Hornby, 2015). Based on this general description, the term boundary is defined in the context of this elaboration as any circumstance that leads to a reduction in the effectiveness or efficiency of an organization. Boundaries and conflicts are therefore used to describe any circumstance that hinders an organization from being perfectly productive. An example of such boundaries or conflicts would be communication issues between different departments, which lead to a reduction of productivity.

2.5 Design Science Research Methodology

2.6 Requirement Engineering

Die Methodik, um die Anforderungen der Endanwender in dieser Arbeit zu bestimmen, orientiert sich an dem sogenannten "Requirement Engineering" Institute of Electrical and Electronics Engineers (IEEE) Standard für die Analyse und Evaluation von Anforderungsspezifikationen. (Vgl. Alain Abran, James W. Moore, 2004, S.2) Der Begriff des "Requirement Engineering", frei übersetzt mit "Anforderungsentwicklung", ist dabei ein englischer Begriff aus der Systemanalyse und wird zur Analyse und Evaluierung von Endanwenderanforderungen genutzt (Vgl. Sommerville, 2011, S.82-111) Der IEEE Standard 610.12-1990 definiert eine Anforderung (englisch "Requirement") dabei folgendermaßen:

> ***"(1) Eine Bedingung oder Fähigkeit, die ein Benutzer benötigt, um ein Problem zu lösen oder ein Ziel zu erreichen. (2) Eine Bedingung oder Fähigkeit, die ein System oder eine Systemkomponente erfüllen oder besitzen muss, um einen Vertrag, eine Norm, eine Spezifikation oder andere formell auferlegte Dokumente zu erfüllen. (3) Eine Dokumentendarstellung einer Bedingung oder Fähigkeit wie in 1 oder 2."*** (IEEE, 1990, S.62)

Die hier verwendete Vorgehensweise orientiert sich allerdings nur in Teilen an diesem Standard, da das "Requirement Engineering" grundsätzlich ein Teil des Systementwicklungsprozesses ist und sich diese Arbeit nur mit dem Analysieren und Evaluieren von Standardinhaltsanforderungen in einem existierenden System befasst. Deshalb orientiert sich die Definition

im Kontext dieser Arbeit an Punkt (1) des IEEE Standard 610.12-1990 mit dem Ziel, dem Benutzer eine verbesserte Usability zu gewährleisten. Dabei kann das "Requirement Engineering" in die vier Teilbereiche der Anforderungserhebung, Anforderungsanalyse, Anforderungsspezifikation und Anforderungsbewertung aufgeteilt werden. Bei der Anforderungserhebung werden zunächst die Anforderungen an ein neues Softwaresystem definiert, diese dann analysiert und ausgearbeitet, um sie dann zu dokumentieren und zu evaluieren. (Vgl. Sommerville, 2011, S.82-111) Die Anforderungserhebung der Endanwender an das System wird in dieser Arbeit mithilfe einer Umfrage durchgeführt. Diese Umfrage wird anonymisiert mit Teilnehmern des SAP Enable Now Round Tables stattfinden. In der Anforderungsanalyse wird die momentane Situation der SAP Marketing Cloud betrachtet und mithilfe der Umfrageergebnisse verglichen, um die Anforderungen der Endanwender abzuleiten. Diese Anforderungen werden dann in einem Proof-of-Concept beispielhaft umgesetzt und dazu verwendet, die Anforderungsbewertung durchzuführen. In der Anforderungsbewertung wird dabei der International Organization for Standardization (ISO) 9241 Standard(ISO Org., 2019) genutzt, um die Verbesserung der Usability zu verdeutlichen. Das Ziel der Umfrage und dieser Arbeit ist es also herauszufinden, wie der SAP Enable Now Webassistent von den Endanwendern genutzt wird und wie der dafür bereitgestellte Standardinhalt in der SAP Marketing Cloud verbessert werden kann.

3 Identification of the Problem

3.1 Previous Studies and Gaps in the Literature

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 information technology (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. 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. 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. In order to prevent business failure, big data must be used effectively (Amankwah-Amoah and Adomako, 2019). This, however, might be hindered by boundaries and conflicts that have arisen during an organization's existence. Technical challenges are only some part of the underlying problem. Boundaries that have surfaced between employees or professional groups might also impact the effective usage of data analytics. This problem only becomes more apparent when the fact is taken into account that most companies are not built around processing data, but have subsequently implemented some sort of data analytics

system into their organization and processes.

The effective use of data and utilization of data analytic tools might be the next digital revolution, which businesses have to adapt to, in order to stay competitive. This literature review focusses on the salient boundaries and conflicts that potentially restrain companies from utilizing their data to its fullest potential. In an effort to achieve this, the current state of knowledge regarding boundaries in data analytics is analysed. For this, a literature review was conducted. The goal of this literature review, above all else, is to summarize the current knowledge of the topic, to find interrelationships that span the literature and to show potential literature gaps for future research on the topic.

In order to analyze and structure the existing literature effectively, this elaboration firstly defines the terms “data analytics”, “boundaries” and “conflicts”. In addition, the so-called “information value chain” is introduced, which will be used to structure the found literature further. Afterwards, the research method and procedure of the literature review are presented. Eventually, the findings and results are discussed and a conclusion about the current state of research on the topic is given.

In order to analyse the current state of knowledge regarding the topic of boundaries and conflicts in data analytics, a literature search was conducted. The main objective was to analyze the existing literature to find research gaps, particularities and interrelationships between literature. Consequently, relevant literature was identified and reviewed. Afterwards, the identified literature was categorized and analyzed.

Initially, it was assumed that the topic of boundaries and conflicts in 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 1 shows the databases that were used.

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 in-

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 1: Databases Used in the Literature Search

cludes 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 and by their respective industry and involved departments in order to find patterns and similarities in the literature. For example, the overaverage occurrence of boundaries in certain industries could indicate that certain businesses are more susceptible to the emergence of boundaries. In a second in-depth analysis, the different boundaries and

possible solutions that were proposed by these articles were categorized and summarized in order to draw overarching conclusions.

The findings of the literature search were structured in multiple ways in order to draw further conclusions than the individual publications allow for. Therefore, the literature search was structured according to the information value chain, the methodology used and the industry covered. Subsequently, the commonalities in content of the literature and the conclusion of the discussed articles are presented.

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 2.

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 2: Results Assigned to the Information Value Chain

Table 2 shows an overabundance of literature that got assigned to the “knowledge” phase of the information value chain. Among other things, this is due to the fact that the content of this literature deals with the construction and exchange of knowledge within certain groups. The context of this literature is mostly not directly written within the context of data analytics, but nonetheless deals with boundaries in a relevant context.

The significantly fewer entries for the other phases could be explained due to these phases being researched less. However, it cannot be concluded that this underrepresentation

is due to the fact that these phases are less susceptible to boundaries. For this, more literature would have to exist confirming that these areas are 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 the context of boundaries. In fact, the corresponding literature, which was assigned to these phases mainly consists of publications researching the application of big data. Their main research object does not directly consist of the identification or resolution of boundaries. 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 boundaries 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 2. 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 3.

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

Table 3: Research Approach Used in the Literature

The distribution presented in table 3 does not show any significant results. Methods such as case studys and surveys are used more often, this however might be due to the fact that these research methods are easier to implement or more common. Experiments for example might be harder to justify and less effective than surveys, in the context of data analytics.

Table 4 shows the assignment of the identified literature to the corresponding area that was the topic of the particular publication. Specifically, the literature was divided by areas or objects in which or by which boundaries occurred and in which industry this was studied. This was done to assess whether boundaries do occur above average in certain areas or industries. The distribution of the location boundaries are occurring in, shown in table 4 (column “Occurrence of boundaries”), is relatively balanced. Boundaries occurring between different organizations is slightly less often topic of publications and boundaries between local teams slightly more often. However, this circumstance is likely due to the fact that studies spanning multiple organizations are comparatively more difficult to conduct and that local teams were often the topic of literature that deals with boundaries in knowledge management, which was more common in the results of the literature search. In fact, most boundaries seem

Occurrence of boundaries	Industry										**Total**
	Primary Sector	Secundary Sector		Tertiary Sector							
	Oil and Gas	Production	Software	Financial	Transport	Insurance	Consulting	Education	Healthcare	Not specified	
Upper Management		2			1					4	7
Different Organizations				1					1	1	3
Knowledge Sharing	1		1	2		1	1		1		7
Different Departments		2							2	2	6
Information Systems		1	1		1					2	5
Local Teams		2	1	1		1	1	1	1	2	10
Total	1	7	3	4	2	2	2	1	5	11	38

Table 4: Industry and Subject Matter of the Literature

to occur between groups that are part of a clearly separate affiliation. For example, upper management can be prone to lead to boundaries as this group of people is separated from the workforce by organizational structures. Simultaneously, local teams within the workforce can also be separated by organizational structures and thus be prone to boundaries. The rows "Knowledge Sharing" and "Information Systems" that are displayed in table 4 show that this exceeds the occurrence of boundaries between groups. Results from the literature indicate that the detachment of processes and systems from the natural workflow or groups in general also encourages boundaries.

The distribution of literature regarding the corresponding industry is less evenly divided. Industries in the primary sector are the most underrepresented in this context, only being the focus of one publication in the oil and gas industry. The most common industry, on the other hand, is the traditional production industry, which includes all businesses that are engaged in the production or manufacturing of goods. A reason for this could be the widespread availability of these traditional industries. More patterns like the distribution of capital in different industries were tried to be applied to the results displayed in table 4, but none resulted in any significant results.

In summary, no further particularities can be drawn from the distribution of literature among the corresponding industries.

After analyzing the literature in terms of its methodology and field of application, the specific results of each publication were analyzed in-depth. Consequently, in the following, both commonalities in boundaries and solutions to solve them are presented. These are structured according to common features and show the number of publications that support these findings. Table 5 shows the most common boundaries that are researched in literature and table 6 is the most common means to resolve them.

In general, all results regarding the occurrence of boundaries can be divided into the three categories indicated in table 5. These three categories all have in common that some kind of separation takes place between groups. Another finding that has emerged from the literature is that the occurrence of boundaries between separate parties does not only affect groups of people. Separating people and processes through poor tasks or separating people and IT systems through poor processes can also lead to boundaries. The literature indicates

that any form of unclean separation between parties inevitably leads to boundaries. This is confirmed by literature which shows that boundaries occur even when processes or systems whose task it is to minimize boundaries are too isolated. On the other hand, tightly integrated processes show little or no boundaries. The literature gives the supply chain as an example of this. Although a large number of different groups and departments have to work together in it, there are (usually) hardly any major boundaries, since the individual parties are closely connected by the optimized affiliation to the supply chain (Chen et al., 2021).

Theoretical construct	Meaning	Articles
Between local groups and professional teams	Boundaries often occur between groups that are part of a clearly separated affiliation, like management and workforce, offshore teams, different organizations or even employees who are supposed to interact as boundary spanners (Levina and Vaast, 2008, van Osch and Steinfield, 2016, Mäkelä et al., 2019).	16
Information Technology and available data	The IT systems or data do not meet the quality required for effective use (Wixom and Watson, 2001). Additionally, in some cases employees are not able to use the full potential of the systems due to lack of training (Goodhue et al., 1992).	10
Too much investment in dedicated boundary spanning systems	The investment into dedicated systems or actions that are meant to span boundaries can in of itself lead to boundaries (Levina and Vaast, 2005). A critical condition for this to occur is the isolated un-integrated implementation of such systems. Examples found in the literature range from IT infrastructure (Currie and Kerrin, 2004) to boundary spanning staff).	3

Table 5: Boundaries Found in the Literature

Table 6 shows possible solutions to overcome boundaries that are suggested by the literature. These are more varied and numerous, but can also be summarized. The most important means of counteracting boundaries is the close collaboration between different groups, processes, and systems, as well as intrinsic motivation of stakeholders and quality of data and the right usage of systems. A large part of literature which directly deals with the occurrence of boundaries also mentions the implementation of so called boundary spanners and boundary objects as a way to reduce boundaries. Cross and Parker define boundary spanners as

Theoretical construct	Meaning	Articles
Integration of personnel and processes	The integration of people and information technology can effectively prevent boundaries from forming (Kotlarsky et al., 2014). Information systems should be naturally used in different tasks, instead of requiring the additional use of a dedicated system (Levina and Vaast, 2005, Wook et al., 2021). The close collaboration between teams and organizations can prevent boundaries at the same time (Zhang and Li, 2021).	14
Intrinsic motivation of individuals and teams	Boundary spanning employees need to be encouraged in order to make them more effective. This can be done through organizational design of processes or tasks or reward and feedback mechanisms (Minbaeva and Santangelo, 2018). The boundary spanners should aim for the larger mutual benefit of organizations (Mäkelä et al., 2019). One example is that semi-formal boundary spanners are often more effective than formal ones, because of their intrinsic motivation to share knowledge (Guven–Uslu et al., 2020).	10
Right usage of big data analytics	In order to unlock the full potential of (big) data analytics it is important to utilize it right (Amankwah-Amoah and Adomako, 2019). This can be, for example, archieved through the usage and data scientists (Kim and Tomprou, 2021) or the usage of the right application architecture (Goodhue et al., 1992).	8
Quality of data	Deployed information system must allow for a seamless data collection (Lukyanenko et al., 2019). Especially the data quality in the area of Big Data must be guaranteed (Wook et al., 2021).	5
Usage of boundary objects	The usage of boundary objects, such as information technology or documentation of all kinds can improve boundary spanning capabilities (Pawlowski and Robey, 2004).	5
Legitimacy of boundary spanning personnel	The fact that boundary spanners are formally nominated and socially accepted is an important part of their effectiveness (Levina and Vaast, 2005). Based on this, it is important that employees know who can be addressed about certain topics (Mell et al., 2022).	2

Table 6: Solutions for Boundaries Found in the Literature

"[...] critical links between two groups of people that are defined by functional affiliation, physical location, or hierarchical level [...]" (Cross and Parker, 2004). Thus, boundary spanners represent individuals that unite any two groups that are separate from each other. This concept is, for example, used in knowledge management to explain the distribution and cooperation of knowledge sharing in organizations (Levina and Vaast, 2005). Managers can be one example of these kinds of individuals, as they are supposed to bridge the gaps between multiple teams through a hierarchical manner (Allen and Cohen, 1969). IT employees can be another more informal example of boundary-spanning employees as they have to interact with a number of different teams which maybe would not share a connection otherwise. In this case, the IT-system itself can act as a so-called boundary-object (Pawlowski and Robey, 2004). Boundary-Objects are any kind of object that is used by a variety of distinct groups. An example of a boundary object would be the sketch of a car, which is created by the design team and shared with an engineering team to calculate wind resistance. The main property of boundary-objects that makes them help resolve boundaries is that they are used by different groups which otherwise do not share any extensive and direct interactions with each other. The presence of both boundary spanners and boundary-objects in literature indicates that close collaboration and cooperation can lead to the resolving of boundaries. This is supported by the other findings of table 6.

As already indicated in section **??**, much of the literature found could be assigned to the "knowledge" phase of the information value chain. This literature mostly consists of publications which research boundaries in certain professional groups, but not within the context of data analytics. 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, not concerned with potential boundaries that could arise within data analytics. These two research areas, which do not quite fit the context of boundaries in 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.

Nevertheless, the literature that can be identified regarding boundaries in organizations shows that a lot of boundaries occur between groups that are part of a clearly separated affiliation. As already analyzed, this circumstance applies not only to groups of people but also to processes and systems. An interesting conclusion from this would be to no longer consider boundaries as an event that occurs, but as a default state that can be contained by certain measures. This would lead to the assumption that boundaries exist between each employee or asset in a company and that these can be mitigated through good processes, tight integration or IT systems, instead of having the assumption that boundaries only occur after the fact. This would shift the research focus from identifying reasons for boundaries to studying how to actively solve them, which might be more effective.

The remaining literature that could be identified deals less with boundaries and more with data analytics itself. In particular, big data analytics is a rather new topic and thus a lot of research focusses on its potential rather than on what kind of boundaries or conflicts hold it back. In addition, most data analytic studies start out with a set of data and then use it to explore the possibilities of big data, without regard to the whole process. This general approach might not be suited to show boundaries, because the used data already exists. An approach where a certain desired outcome is specified first and then data is collected, preferably in an organizational context, might be better suited. This approach, however, was not implemented in any of the literature, possibly due to the fact that research is still trying to determine what capabilities the usage of big data and data analytics has. This also reflects the aforementioned fact that no literature could be found that could have been assigned to the "decisions" and "actions" phases of the information value chain. An interesting approach could therefore be to research the occurrence of boundaries in the decision-making process. Under certain circumstances, it would be possible that decisions based on data analytics are trusted differently than the ones made by more traditional means. The predictions of a machine learning algorithm are maybe more trusted than the ones by experienced managers. This might lead to some internal conflicts and would therefore be an example of a currently insufficiently studied part of the topic of boundaries in data analytics.

Furthermore, it should be again critically noted that although a large number of boundaries and solutions for them could be found in the literature, almost all of these were not ex-

amined in the context of data analytics. The results are nevertheless transferable to the topic of data analytics, although slightly different actual results in the area of data analytics are conceivable and will only become apparent with further research into this area.

All of this leads to the fact that boundaries in a traditional organizational context are researched quite extensively. However, boundaries in the context of data analytics are not. In general, although findings can be transferred thematically, the exact facts still contain many gaps in the literature.

As mentioned before, the three goals of this literature review were to summarize the current knowledge of the topic of boundaries in data analytics, to find interrelationships that span the literature and to show potential literature gaps for future research on the topic.

Even though data analytics might be the next big thing and shape the future of decision making in business, the topic in of itself is only researched fragmentarily. The literature of data analytics mainly focusses on its application. Possible boundaries that limit the effectiveness of data analytics are not discussed. Generally, the occurrence of boundaries in organizations and possible counter measures to resolve them are presented in literature. These, however, focus mostly on knowledge sharing in a traditional setting and do not treat the subject matter in the context of data analytics. These two different literature paths narrow down the topic of boundaries in data analytics, but do not explore it directly. This shows a gap in the current state of research that could be subject to future investigation. Simultaneously, even though literature could be found discussing boundaries and challenges in data analytics, almost no literature could be found that focusses on the decision-making process itself. This is suprising as data analytics is first and, for most, a decision-making tool. This circumstance again clearly shows the lack of research done in this area and another research gap. Future investigations could, for example, focus on the effects of data analytics on the decision-makers themselves. Nevertheless, it was possible to find literature whose findings can be transferred thematically to the field of data analytics. This literature clearly indicates that boundaries tend to occur between groups that do not share an affiliation. Simultaneously, it was found that information technology or other processes can generally lead to boundaries if they are not integrated well enough. These boundaries can, however, be overcome through the tight integration of the different parties. Finally, intrinsic motivation and the quality of processes and data are further

factors that can influence the occurrence of boundaries and conflicts. These general findings reflect and summarize the current state of knowledge regarding boundaries in data analytics.

In summary, it was found that the topic has not yet been extensively researched. At the same time, commonalities in terms of boundaries and solutions were found in the literature and transferred to the topic of data analytics. This is highly relevant, because data analytics is an important topic for companies to remain competitive. For this reason, the area of boundaries in data analytics should be further investigated.

The three main goals of this literature review, summarizing the current state of knowledge, finding interrelationships that span literature and to show literature gaps could therefore be met. In particular, identifying gaps in the literature for future research is an important milestone in order to fully understand how salient boundaries and conflicts in data analytics can be resolved.

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 Qual-

ität 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

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).

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*.

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

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/

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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/

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

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

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

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

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, June 29, 2023

Max Darmstadt