Junior Management Science 7(5) (2022) 1224-1241

Junior Management Science

journal homepage: www.jums.academy

Big Data Analytics Capabilities: A Systematic Literature Review on Necessary Skills to Succeed in Big Data Analytics

Marc A. Richly

Ludwig-Maximilians-Universität München

Abstract

While the amount of data keeps growing, managers ask themselves whether they already retrieve full value from their data. To maximize the value of big data, literature offers first insights in building BDA capabilities (Gupta & George, 2016; Mikalef, Framnes, Danielsen, & Krogstie, 2018). Nevertheless, BDA remains a new field to researchers and companies. BDA frameworks, still offered scarcely, discuss roughly the same dimensions (incorporating some technical, human, and cultural aspects), but are only superficially discussed. This thesis builds a framework of the different approaches offered in literature. Furthermore, it is important to distinguish whether a new development as BDA can be seen as a trend topic or rather a long-lasting game changer for businesses. Here, this thesis discusses differences among digital capabilities, IT capabilities, that research stared addressing by 1990, and BDA capabilities. A major finding is that building IT capabilities is considered as an isolated responsibility of IT departments by, i.e., offering IT infrastructure to the whole company. BDA capabilities, on the contrary, cannot be planned and rolled out from one specific department – those need to be developed in every organizational unit; therefore, a data-driven culture is a key element in building BDA capabilities.

Keywords: Big data analytics; Big data; Data analytics; Dynamic capabilities; Resource-based view.

1. Introduction

Several studies have shown that firms using Big Data Analytics (BDA) in their company are more successful. For instance, authors of a MIT Sloan Management article could show that "top-performing organizations use analytics five times more than low performers" (LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011, p. 22).

Moreover, research has shown in many studies that companies that run BDA activities reach stronger results (Günther, Huysman, & Feldberg, 2017; Gupta & George, 2016). Hao, Zhang, and Song (2019, p. 9-12) have found that big data improved sales growth and gross margins, whilst Wamba, Gunasekaran, Akter, Ren, and Dubey (2017, p. 9, 16) could show that BDA capabilities have direct, positive effects on firm performance. In general, many researchers see high potential of BDA enhancing innovation, competition, and productivity (Manyika et al., 2011).

While the amount of data keeps growing, managers ask themselves whether they already retrieve full value from their data. To maximize the value of big data, literature offers first insights in building BDA capabilities (Gupta & George, 2016; Mikalef et al., 2018).

Nevertheless, BDA remains a new field to researchers and companies. Thus, current results change quickly, and researchers still discuss capabilities a company needs for BDA. BDA frameworks, still offered scarcely, discuss roughly the same dimensions (incorporating some technical, human, and cultural aspects), but are only superficially discussed. Most BDA papers do simply orientate at IT capabilities and only offer marginal adoptions to the analytical component. Only recently, first papers start to quantify the proposed effects of BDA capabilities (and not analytics on firm performance in general) to enrich literature (Yasmina, Tatoglua, Kilicb, Zaimc, & Delen, 2020).

This leads to the first research question.

> RQ1: "Which capabilities are proposed in literature to conduct Big Data Analytics in firms and how do they influence success of Big Data Analytics?"

Furthermore, BDA was not situated in the context of other big technological developments as e.g., IT capabilities in the 1990s, yet. For managers and researchers, it is nevertheless important to be able to weigh different trends or to even iden-

DOI: https://doi.org/10.5282/jums/v7i5pp1224-1241

tify whether a new development as BDA can be seen as a trend topic or rather a long-lasting game changer for businesses.

Consequently, the second research question reads as follows:

> RQ2: "In what respect do Big Data Analytics Capabilities resemble and differ from IT and Digital capabilities?"

2. Conceptual Foundation

2.1. Definitions of Key Terminology

For the beginning, key terminology will be defined (Big Data, BDA, BDA capabilities), clearly delineated from each other, and situated in current research.

2.1.1. Big Data

Initially, Big Data was coined to reflect the "bigness" or *volume* of data "generated as a result of using new forms of technology (e.g., social media, radio-frequency identification (RFID) tags, smart phones, and sensors)" (Gupta & George, 2016, p. 1050). It is measured not only by a large set of observations itself, but also many variables.

With volume, also velocity and variety were introduced as the "three Vs" characteristics to define the term of Big Data (McAfee, Brynjolfsson, Davenport, Patil, & Barton, 2012, p. 4-5).

Velocity reverberates the speed at which data is collected (near to or at real time) from sales transactions, sensors, social media posts, and sentiment data from breaking news and social trends (Gupta & George, 2016).

The term *variety* describes the plurality of the data (as texts, graphics, videos, networks among others) as it can emerge in structured, partly structured, and unstructured sources (Gupta & George, 2016).

Today, most definitions of big data have shifted from including the analytical tools and visualization of results to only integrate the V-characteristics of Big Data, as Davis 2014 defined: "Big data consists of expansive collections of data (large volumes) that are updated quickly and frequently (high velocity) and that exhibit a huge range of different formats and content (wide variety)."

Some researchers have extended the existing body of definitions by further characteristics as *veracity* (unreliability inherent in some sources of data), *variability* (variation in the data flow rate), and *value* (showing the potential in transforming low value raw data into high value data via BDA) (Gandomi & Haider, 2015, p. 139). The term *value* is nonetheless highly firm-dependent as it is connected to the strategic goals of a company (Günther et al., 2017, p. 191).

2.1.2. Big Data Analytics

BDA definitions – in comparison to big data – focus on multiple analytics methods "that address the diversity of big data to provide actionable descriptive, predictive, and prescriptive results" (Lamba & Dubey, 2015, p. 5). The data can be analyzed with technologies (e.g., data mining tools or database) and different techniques (e.g., analytical methods) to generate insights of big data (Kwon, Lee, & Shin, 2014, p. 387).

Analytics and Data Style

As one of Big Data's characteristics is variety (video, text, audio, social media data etc.), also BDA must adopt with suitable techniques to the style of data (Choi, Wallace, & Wang, 2018; Mikalef et al., 2018).

Textual data

Firstly, textual data as texts, obtained from emails, social media posts, blogs, surveys, news etc., can be analyzed with analytical tools to extract information. So far, three approaches exist to work with textual data: Information Extraction (IE), Text summarization, and Question Answering (QA) (Gandomi & Haider, 2015, p. 140).

IE can, for instance, obtain specific data as drug name and dosage from a medical prescription. Text summarization are (or will be) useful in news articles and emails to provide a short overview with the important details. QA most important examples include Siri from Apple, or Alexa from Amazon. Siri can provide answers on orally asked questions (Gandomi & Haider, 2015, p. 140).

Video data

This form of data is not only more complex to analyze, also in terms of volume it is more difficult to process video analytics. Nevertheless, analytical methods for video data are progressing well (Gandomi & Haider, 2015, p. 141). Use cases for video analytics are broader, from closed-circuit television (CCTV) cameras to video sharing websites (as YouTube that automatically checks videos for e.g. copyright violations (Agrawal & Sureka, 2013)). Furthermore, video analytics can help for marketing purposes to - for instance - identify the demographics, gender and similar characteristics of people that go shopping (Hu, Xie, Li, Zeng, & Maybank, 2011).

Audio Data

For audio analytics (in most cases: speech analytics, when analysis is about spoken words), two main systems are in place: LVCSR systems (transcript -based approach) and Phonetic-based systems (Gandomi & Haider, 2015, p. 141).

LVCSR systems uses automatic speech recognition (ASR) to transcribe spoken word into text with the help of a dictionary. A popular application is the dictation function on smartphones. With the text output, standard text-based analytics can be conducted (Gandomi & Haider, 2015, p. 141).

Phonetic-based systems work with "perceptually distinct units of sound" (Gandomi & Haider, 2015, p. 141) that distinguish one word from another and thus create text. As with LVCSR systems, this text can then be analyzed with text-based analytics.

Main use cases for audio analytics are in call centers and hospital (for automatically documenting what a doctor has said) (Wang, Kung, & Byrd, 2018).

Social Media Data

Social media data can consist of varying online platforms, as social networks (e.g., Instagram, LinkedIn), blogs, social news (e.g., Digg and Reddit), media sharing (e.g., YouTube), wikis (e.g., Wikipedia), review sites (e.g., Trivago, Yelp), and questions-and-answer sites (e.g., Ask.com) (Gandomi & Haider, 2015, p. 142; Mikalef et al., 2018).

Social media analytics has emerged in the early 2000s and can be classified into two groups: *Content-based analytics* (focuses on data posted by users) and *Structure-based analytics* (synthesising structural attributes of a social network) (Gandomi & Haider, 2015, p. 142).

Technologies and Tools

Analytics itself can be conducted on different data types. But also different technologies are used to analyse data sets (here: descriptive, predictive, and prescriptive results (Lamba & Dubey, 2015, p. 5)) that will be discussed in the following paragraph.

Descriptive Analytics – What has happened? Raw data, as sales, customers, and operations data, from the past can be analyzed on insights. Organizations use this form of analytics to gasp a deeper understanding of their business and identify relationships that have happened previously (Soltanpoor & Sellis, 2016, p. 247). With that, insights can be obtained as: Which products sell better or worse compared to other products or how has the demand developed over a year.

To generate reports on the historical data or to extract information from raw data, techniques as statistical analytics, data integration, data augmentation, and data reduction can be used (Soltanpoor & Sellis, 2016, p. 247). In general, these techniques are rather simple to comply (creating graphs) and normally include basic descriptive statistics as "measures of central tendency (mean, median, mode), measures of dispersion (standard deviation), charts, graphs, sorting methods, frequency distributions, probability distributions, and sampling methods" (Ajah & Nweke, 2019, p. 7).

Statistics, nevertheless, were identified as one major technique in BDA by Choi et al. (2018). With this already well-established field, relationships as correlations or statistical regression are often used to generate insights of data sets.

Predictive Analytics – What will happen?

Prediction can vary a lot in big data use cases – "the failure of jet engines based on the stream of data from several thousand sensors, to predicting customers' next moves based on what they buy, when they buy, and even what they say on social media" (Gandomi & Haider, 2015, p. 143).

The aim here is to foresee opportunities and risks in the future. Two core techniques can be used to generate these insights: regression techniques (e.g., multinomial logit models) and machine learning techniques (Gandomi & Haider, 2015, p. 143).

Machine learning was identified as another major BDA technique by Choi et al. (2018). It compromises not only Artificial Intelligence, but also neural networks, support vector machines, and statistical machine learning. Even though machine learning takes time until it is trained and can return results, it provides helpful algorithms to capture complex behaviors.

With machine learning, the concept of "data lakes" also became relevant as a new technique. Data lakes were firstly defined about ten years ago by Dixon (2010) and compromise of four main characteristics: they are (1) massively scalable in terms of volume, and the data is (2) stored as raw data ("as is") and thus in an unstructured format (in contrast to data warehouses, where structured data is stored). Thereby, the lakes also use (3) dynamic analytics applications as machine learning (not like data warehouses: pre-build static). Lastly, data becomes (4) accessible, as soon as the lake is created (again, different to data warehouses which are designed for only slowly changing data) (Miloslavskaya & Tolstoy, 2016, p. 302). As the real time analysis has been identified as a major trend in Business Intelligence (BI) (Russom, 2011, p. 26), data lakes can be expected to play a big role in BDA.

Prescriptive Analytics – What should be done?

Prescriptive analytics "provides enterprises with adaptive, automated, time-dependent, and optimal decisions" (Soltanpoor & Sellis, 2016, p. 247).

The aim is to generate recommended business decisions, optimal courses of action or similar. These outputs are normally generated through techniques as optimization, simulation, operations research, or management science (Soltanpoor & Sellis, 2016, p. 247).

Optimization was also highlighted as another major BDA technique by Choi et al. (2018). This technique can be used in quantitative decision-making problems to find the (near) optimal solutions.

2.1.3. Big Data Analytics Capabilities

A business capability is defined as "the organization's capacity to successfully perform a unique business activity" (Keller, 2009, p. 2).

Looking at the field of digital transformation, Soto Setzke, Opderbeck, Böhm, and Krcmar (2020) identified several configurations and strategies that are needed to build capabilities. Their findings reach from the importance of centralized decision making to the rather negative impact of competitive pressure. Also, they suggest that smaller or medium-sized companies shall focus on strong partnerships that can (partly) replace missing capabilities (Soto Setzke et al., 2020, p. 14).

In IS literature, researchers have proposed to take a bird's eye view to better understand how IS investments enhance business value (Bharadwaj, 2000; Wamba et al., 2017). Hence, IT capabilities were introduced to investigate these connections. They are defined as the "firm's ability to mobilize and deploy IT-based resources in combination or co-

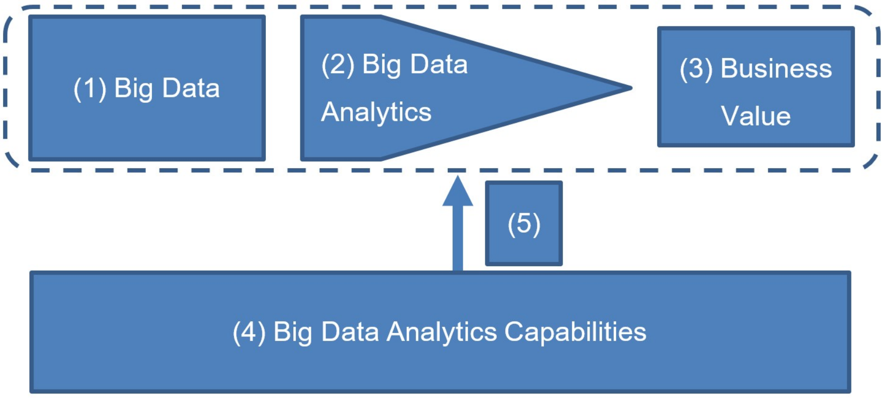

Figure 1: Delimitation of Terminology

present with other resources and capabilities" (Bharadwaj, 2000).

Also for BDA, capabilities were introduced. Akhtar, Frynas, Mellahi, and Ullah (2019, p. 252) defined BDA capabilities as a balanced combination of advanced technologies supported by large datasets to generate analytical reports and actionable insights utilized, produced, and processed by mathematical, statistical techniques, and machine learning tools for enhanced performance, but also essential human resources with big-data skills. In short – a balance must be found between the human part and the technical part of a business.

In regards to a successful implementation of BDA, Wang et al. (2018, p. 8-9) suggest five strategies in their analysis on the healthcare setting: (1) the implementation of big data governance, (2) the creation of an information-sharing culture, (3) the training of key personnel to use BDA, (4) the integration of cloud computing into the organization's BDA, and (5) the generation of new business ideas from BDA.

2.1.4. Delimitation of Terminology

While (1) big data is about data sets with exact characteristics (s. 2.1.1), (2) BDA encompasses the tools and technologies used to analyze big data (Mikalef et al., 2018, p. 555-556). Some researchers even include (3) business value in terms of presentation and performance impact of BDA in their definitions.

This paper will focus on (4) Big Data Analytics capabilities, to understand what firms need to work with Big Data and how firms can build BDA capabilities. Addressing the research question, firstly this paper will show current capabilities proposed by literature to conduct BDA (4). Secondly, this paper will show how these capabilities influence the success of BDA (5).

2.2. Definition and Types of Capabilities

2.2.1. Resource Based Theory

To describe why and how BDA is important and can enhance firms' performances, researchers use different strategic backgrounds. During the 1980s most strategic theories focused externally (e.g. Porter's five forces model, firstly published in 1979), in the 1990s researchers' frameworks have focused more internally as the resource-based view emerged (Barney, Ketchen Jr., & Wright, 2011, p. 1300). Also for BDA, internal frameworks are applicable, and thus, the resource-based view that later emerged in the resource-based theory (RBT) is commonly used by researchers to describe how BDA capabilities can add value (Gupta & George, 2016, p. 1050).

The RBT understands a company as a sum of different firm resources that are composed of "assets, capabilities, organizational processes, firm attributes, information knowledge etc." (Barney, 1991, p. 101). Initially three types of assets were proposed – physical (e.g., specialized equipment), human (e.g., expertise in technology), and organizational (e.g., superior sales force) – that can be used to implement value creating strategies (Eisenhardt & Martin, 2000, p. 1106-1107). Gupta and George (2016, p. 1054) argue that data can also be an important asset to an organization – if the firm has implemented a data-driven culture.

Firm resources must fulfil the characteristics of value, rareness, imperfect imitability, and substitutability (s. Tab. 1). With those a company can implement a value-generating strategy by creating a competitive advantage over competitors (Barney, 1991, p. 105-112).

Then, the firm resources can be used to archive a competitive advantage. A competitive advantage exists, when a firm has implanted a value creating strategy not used by any other firm (Barney, 1991, p. 102).

Later, the understanding of firm assets was split into two parts, resource-picking and capability-building to further explain the mechanism how assets can be turned into a competitive advantage (Makadok, 2001; Mikalef et al., 2018).

On the one hand, *resource-picking* describes the process of identifying (often also buying) resources that promise to be VRIN (valuable, rare, inimitable, and non-substitutable). Nevertheless, these resources are still rather easy replicable (Santhanam & Hartono, 2003).

On the other hand, *capability-building* facets are concerned with the orchestration of resources. Capabilities, in contrast to resources, are highly specific to an organization

Table 1: Conditions for sustainable advantage, (Barney, 1991, p. 105-106)

Conditions for sustainable advantage	Definition of condition
Value	Asset exploits opportunities / neutralizes threats in a firm's environment
Rareness	Asset is rare among current and potential competitors
Imperfect Imitability	Asset must be difficult (or impossible) to copy
Substitutability	No other asset that is rare and imperfect imitable can replace this asset

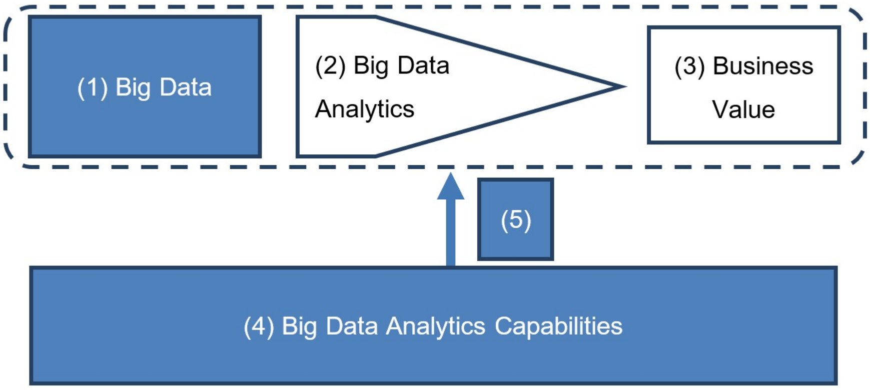

Figure 2: Resource-picking (1) and capability-building (4) in context of the BDA process

and thus is hard to replicate. With a distinctive set of capabilities, a sustained competitive advantage can be reached (Santhanam & Hartono, 2003).

In the context of BDA, big data can be thus seen as a *resource-picking activity* (s. (1) in Fig. 2) that can be bought from external suppliers or developed from internal data. BDA, in contrast, is the result of *capability-building activities* (s. (4) in Fig. 2). These capabilities are firm specific and cannot be bought or sold as they are in some way connected to the organization.

Research also suggests that firms do need a resource before they start building capabilities for it (Mikalef et al., 2018). This is important for firms as it shows that they first need some level of access to big data, before they should start building capabilities to analyse it.

2.2.2. Dynamic Capabilities View of an Organization

With the resource-based theory explaining how firms can build a competitive advantage against competitors, the RBV cannot explain how companies can defend its competitive advantage against competitors in rapidly changing environments. The Dynamic capabilities theory (DC) tries to fill this gap as an extension to the RBV (Mikalef et al., 2018, p. 560).

DC is defined in a seminal article as "the ability of a firm to integrate, build and reconfigure its resources and capabilities to address changing environments" (Teece, Pisano, & Shuen, 1997, p. 516). Researchers furthermore make a distinction between *ordinary capabilities* (also: "zero order" capabilities) that determine how "a firm makes a living at the moment" (Tiguint & Hossari, 2020, p. 2), in contrast to *dynamic capabilities* that enable an organization to change.

The impact of such dynamic capabilities depends on the environment. In moderately dynamic markets, dynamic capabilities are difficult to identify or to separate from ordinary capabilities. In high-velocity markets, routines (here: dynamic capabilities are simple, experiential, and iterative) are adaptive to changing circumstances – but at the cost of unstable processes and uncertain outcomes (Eisenhardt & Martin, 2000). This results in uncomfortable but still success bringing situations, as one manager put it:

"We do everything on the fly . . . I've done some things at IBM and other companies where there is a very structured environment—these companies are failing and we're leading the way. I'm not comfortable with the lack of structure, but I hesitate to mess with what is working." (Brown & Eisenhardt, 1997, p. 28)

To conclude, how can firms eventually ensure their long-term competitive advantage (especially in high-velocity markets)? They must use dynamic capabilities "sooner, more astutely, or more fortuitously than the competition to create resource configurations that have that advantage" (Eisenhardt & Martin, 2000, p. 1117).

2.3. Challenges in Building BDA Capabilities

First, research on BDA implications and capabilities is still scare, which also makes it difficult for companies to start with the implementation of BDA as they cannot linchpin on a broad base of literatures results (Gupta & George, 2016).

In terms of BDA, the proposed capabilities itself can also entail a challenge for companies. Big data – in its terms of volume, velocity, and variety - is still challenging to current IT architecture, networks, servers, and software (Ajah & Nweke, 2019, p. 22). Some research suggests that firms are in particular worried about the unstructured nature of data (the variety part) rather than the volume of data (Gupta & George, 2016, p. 1051). For that, new tools and technologies are required (e.g., Hadoop) that will be further discussed in the literature review (s. 4.1).

Moreover, before starting with the implementation, a company should make sure to have fully covered needed capabilities. If, for instance, measures to ensure privacy are not implemented in the BDA capability-building stage, it will hardly be considered in the BDA process (in this paper's BDA capabilities framework privacy and security are recognised under *Governance* in 4.3.2). This literature review will enrich current literature by discussing different proposed capability frameworks and showing, what impact the different characteristics have on the success of BDA (as described in RQ1).

2.4. Challenges in Building IT Capabilities

Research in IT Capabilities is much more mature than BDA capabilities – offering a better overview on the challenges that have occurred over time.

Nevertheless, also research on IT capabilities has fought some major problems. In the early 1990s, for instance, several studies have examined the so-called *IT productivity paradox* that referred to the failure to show a positive relationship between IT investments and firm productivity (Gupta & George, 2016, p. 1049). Also, renown economists as Robert Solow (holds a Nobel prize in economics) stated that "we see computers everywhere except in the productivity statistics" (Brynjolfsson, 1993).

In his paper, Brynjolfsson (1993) summarized four possible reasons for the occurrence of this paradox:

1. Mismeasurement of outputs and inputs,
2. Lags due to learning and adjustments,
3. Redistribution and dissipation of profits, and
4. Mismanagement of Information Technology.

Later, in some industries a positive correlation could be clearly shown, and the paradox was eventually resolved. Consequentially, it was suggested to take several resources besides pure investments into account, as managerial and technical IT skills, firm's intellectual capital, and IT infrastructure (Gupta & George, 2016, p. 1049). Also, so far no *BDA productivity paradox* has been identified.

Today, some researchers argue that IT capabilities no longer offer any competitive advantage (in contrast to BDA) but have turned to an organizational commodity (Chae, Koh, & Prybutok, 2014, p. 307).

2.5. Challenges in Building Digital capabilities

Digital capabilities are broadly defined and cover most of digital extensions that exceed IT capabilities (s. 5.2 - comparison of BDA and digital capabilities).

Thus, it is also difficult to define concrete challenges of digital capabilities. Mostly, organizations will work on concrete areas as BDA or Artificial Intelligence (AI) that have their respective own challenges.

3. Methodology

For this thesis, a structured literature review on existing papers in IS literature dealing with BDA capabilities will be conducted.

The literature review is divided into two parts, as we must expect that no broad use of BDA capabilities is proposed in current literature, yet. Thus, in the first stage, a structured literature review on existing papers dealing with BDA capabilities is conducted. The literature review is based on Webster and Watson (2002).

In the second stage, the basic capabilities framework will be discussed and for a deeper analysis further punctual research on specific parts of the proposed framework will be conducted.

3.1. Literature Review on Big Data Analytics Capabilities

The literature review in the first stage is conducted in four phases.

In stage one (1), articles of the AIS *Basket of Eight* IS journals were collected. As literature on BDA capabilities is still scarce, the search radius will also be extended to highly cited BDA publications from other journals (via Google Scholar and Web of Science). Additionally, three leading IS conferences (International Conference on Information Systems (ICIS), European Conference on information Systems (ECIS), and Americas Conference on Information Systems (AMCIS)) were considered to cover the latest research on BDA capabilities, too. For the research on journals, only papers published in the past eleven years (2010-2021) were considered, as this field is still relatively new and thus the emphasize lies on the latest developments in BDA research. For conferences, only the years 2018 to 2021 were considered, assuming good conference articles would have reached a journal outlet by today.

The following key words were used to find relevant studies (all words were also searched with their respective abbreviations): Big Data Analytics, Big Data Analytics Capabilities.

After stage one, 17 papers were found in the Basket of Eight, 15 highly cited journals from other journals, and eleven conferences could be identified.

In stage two (2), the titles, keywords, and abstracts were evaluated. Papers were excluded that focused too much on

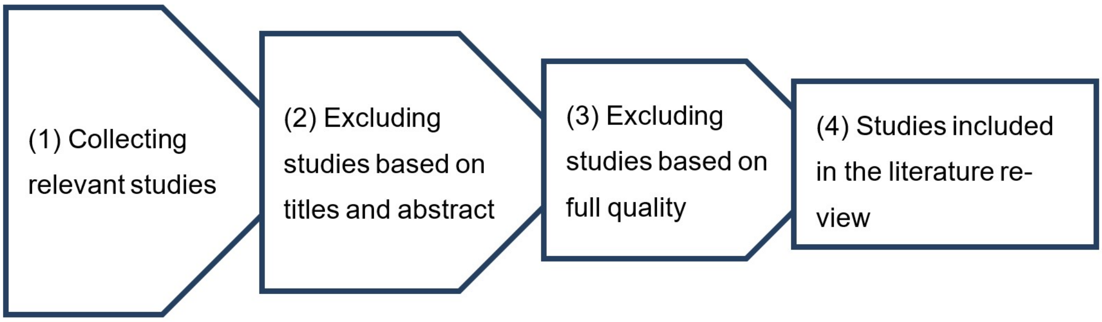

Figure 3: Process of literature review, based on (Webster & Watson, 2002)

the technical side of BDA or privacy. Papers with a focus on specific industry or case studies were not excluded, as still major concepts or the underlying theory could be useful for an evaluation in the literature review.

After stage two, 24 papers and seven conferences were considered for the next stage.

At stage three (3), the full paper was accessed. Here, papers were excluded that did not discuss actual capabilities needed to perform BDA projects in firms and where BDA only played a minor role.

Ultimately, twelve papers and two conferences were accessed and included in the literature review of this thesis.

3.2. Discussion on Big Data Analytics Capabilities versus IT and Digital Capabilities

After the literature review, an extension in form of an argumentative evaluation of the different capabilities a company must have to perform BDA compared to IT and Digital capabilities will be provided. For this, a small research via Google Scholar and Science of Web was conducted to retrieve most influential papers dealing with IT and digital capabilities.

Conclusively, five papers on IT capabilities and nine papers on digital capabilities were evaluated.

4. Literature-Review on Big Data Analytics Capabilities

To build a BDA capability a firm needs different resources. One of the first frameworks "to put analytics to work in [. . .] a business" (Davenport, Harris, & Morison, 2010, p. 19) was offered by Davenport et al. (2010) that was later further discussed by Seddon, Constantinidis, Tamm, and Dod (2017). To implement analytics, they suggested a framework consisting of five pieces, called DELTA (Davenport et al., 2010, p. 19):

- *D*ata: accessible and high quality
- *E*nterprise orientation
- *L*eadership in analytics
- *T*argets: Strategic goal
- *A*nalysts: the human input needed

Nevertheless, this framework was not further discussed or developed in the specific BDA literature. Moreover, Davenport et al. (2010) focus on analytics in general, but not discuss the more specific BDA. Thus, this framework will not be investigated further.

In a highly influential paper, Gupta and George (2016) have built a framework of needed resources, dividing them by tangible, human and intangible resources.

Initially, Gupta and George (2016) derived this separation from research on IT capabilities, using then the framework to expand it on BDA capabilities (Bharadwaj, Sambamurthy, & Zmud, 1999; Chae et al., 2014; Gupta & George, 2016; Santhanam & Hartono, 2003).

Gupta and George's three pillars framework was taken as a foundation by Mikalef et al. (2018), who have adjusted the framework by some characteristics. As this literature review by Mikalef et al. (2018) can be seen as an extension of the Gupta and George (2016) paper, the adopted version of Mikalef et. al. will be taken as a basis for further analysis.

The other big framework used by researchers for BDA capability distinguishes them among (sometimes called slightly different but entailing the same ideas) *BDA technology / infrastructure / data capabilities*, *BDA management capabilities*, and *BDA talent capabilities* (Akter, Wamba, Gunasekaran, Dubey, & Childe, 2016; Ransbotham, Kiron, & Prentice, 2015; Wamba et al., 2017).

Also Yasmina et al. (2020) defined three similar pillars (similar differentiation: Infrastructure capabilities, Human Resource Management Capabilities, and Management Capabilities) in one of the latest researches on BDA.

Both, Mikalef and Yasminas' frameworks, cover same areas, as technical requirements for BDA, human knowledge, managerial roles, culture, and alignment with a firm's business strategy (s. Fig. 4).

Nevertheless, both frameworks set different emphasis. While Mikalef et al. (2018) summarize analytical skills of employees and managerial skills in one pillar "Human Resources", Yasmina et al. (2020) highlight employees and managerial skills by dividing both on two of the three pillars. Looking closer at differences between the third cluster of

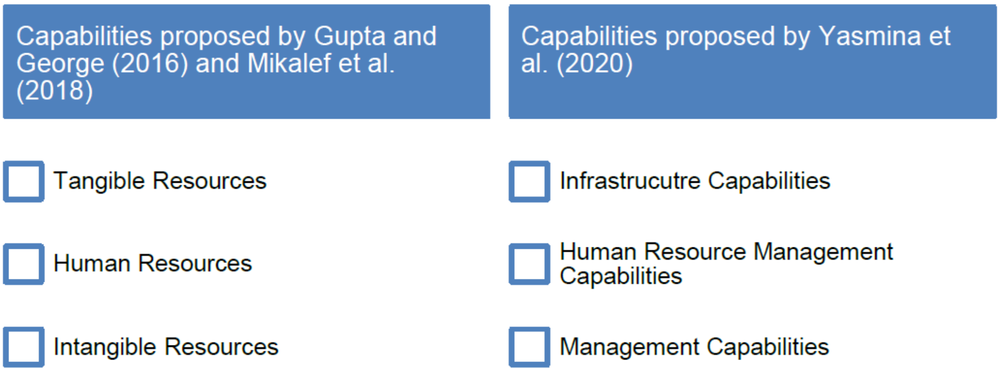

Figure 4: Proposed BDA capability framework by Gupta and George (2016) and Mikalef et al. (2018) (left side) and Yasmina et al. (2020) (right side)

Mikalef et al. (2018) – Intangible resources – and the third cluster of Yasmina et al. (2020) – Management Capabilities – reveals that both discuss cultural and strategy alignment aspects, which straightens out the expected major difference of both researchers.

A detailed comparison can be seen in the following table 2.

Nonetheless, for the following analysis Yasmina et al. (2020) will not be taken as a foundation (even though it is the most recent paper published on BDA capabilities), mainly because the paper does not clearly distinguish IS from BDA capabilities (e.g., in the detailed description of the pillars they refer to IS, but not to BDA characteristics). Furthermore, the paper does not provide deep theory backgrounds, but focuses on testing its hypothesis. Also other papers' frameworks (Akter et al., 2016; Ransbotham et al., 2015; Wamba et al., 2017) with similar separation of capabilities did often not clearly separate BDA from IS capabilities. Thus, in this thesis the provided framework from Mikalef et al. (2018) that clearly focuses on BDA and its respective theories will be used.

4.1. Tangible Resources

The first pillar in the BDA capabilities framework is about *tangible resources* (Gupta & George, 2016; Mikalef et al., 2018). Initially, Gupta and George (2016) have defined the following characteristics: *Data* (internal, external, merging of both), *Technology* (Hadoop, NoSQL), and *Basic Resources* (time, investment).

The characteristic *data* was established by Mikalef et al. (2018), too. Nevertheless, they have exchanged the two other terms with *Software and IS*, and *Infrastructure*.

4.1.1. Data

The most obvious ingredient to conduct BDA is data itself. As a continuously trend, the cost for storage of data is still (even exponentially) declining and consequently data storage is not a main challenge for companies for conducting BDA anymore (Rosenthal et al., 2012). Nevertheless, to recap, big data is not only about its volume, but comes along with other characteristics as velocity, variety, value, and veracity (s. 2.1.1).

Whereas it is easy for companies to measure the size of their data, it already gets more complicated to evaluate the data in terms of quality. Research shows that improved quality helps to turn data into a competitive advantage (Ransbotham et al., 2015). Consequently, leaders in organizations ask themselves "whether they are getting full value from the massive amounts of information they already have within their company" (LaValle et al., 2011, p. 21).

Moreover, organizations have moved on from focusing on enterprise-specific structured data to an increasing amount of unstructured data (variety characteristic of big data) – that does increasingly include external data from outside the organization - to make business decisions (Gupta & George, 2016, p. 1051; Manyika et al., 2011). Thus, not only the amount of data increases (volume), but data also gets more diverse (variety).

Therefore, it is important to evaluate available data in terms of quality, for which research offers some standards to evaluate data sets. Cai and Zhu (2015, p. 4) propose a data quality framework that includes five dimensions (s. Tab 3).

4.1.2. Software and Information System

Next to the increasing amount of data (volume) and the growth in unstructured data (variety), also the major trend in analyzing data in real-time (velocity) have led to the development of new software and Information Systems (IS) (McAfee et al., 2012).

Today, most prominent example of a BDA software is Hadoop. Introduced as an open source project in 2007, it has evolved to support the whole BDA workflow, "including data collection, storage, processing, and much more" (Landset, Khoshgoftaar, Richter, & Hasanin, 2015, p. 5). Besides, research states that the real value of Hadoop does not only lie in its ability to handle large volumes of data, but to manage

Table 2: Detailed comparison of BDA capabilities framework by Mikalef et al. (2018) and Yasmina et al. (2020)

Framework by (Mikalef et al., 2018)	**Similarities**	**Differences**	**Framework by** (Yasmina et al., 2020)
Tangible Resources	Describes technical needs	Yasmina focuses solely on IS infrastructure, not on special infrastructure needed for analytics	**Infrastructure Capabilities**
Human Resources	Technical knowledge needed from employees	Only Mikalef describes managerial skills (Yasmina does so in third pillar)	**Human Resource Management Capabilities**
Intangible Resources	Focus on cultural alignment	Yasmina connects cultural alignment and management capabilities	**Management Capabilities**

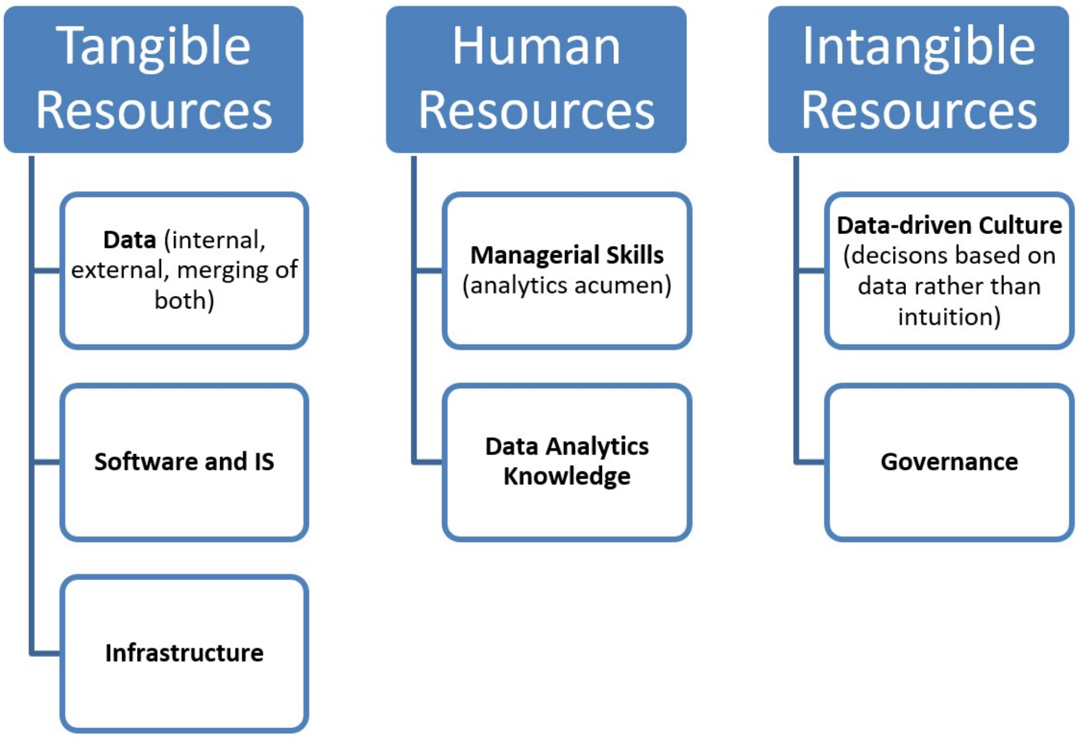

Figure 5: BDA resources to build BDA Capabilities (obtained from Mikalef et al. (2018))

data's variety - a broad range of data types (Russom, 2011, p. 7).

Hadoop combines commodity hardware with its open-source software. Large data streams can then be distributed (also in real-time) onto cheap disks for the analysis part (McAfee et al., 2012, p. 8). The distribution is done by the Hadoop distributed file system (HDFS), while the MapReduce data processing engine is the heart of the analytics part.

MapReduce consists of two primary parts, a mapper and a reducer (Ajah & Nweke, 2019, p. 13). The mapper filters and transforms data by cleaning out dirty data (also duplicates). After this, an intermediate output file is generated that is then distributed to reducers. Here, the reducer phase begins, sorting each file by key and aggregating them into one large file again (Landset et al., 2015; White, 2012). This process allows to run through large and heterogenous data sets in a very effective way.

Other software solutions of nonrelational databases to parallel process large unstructured datasets and to visual-

Table 3: Data quality characteristics (Cai & Zhu, 2015, p. 5-6)

Term	Meaning
Availability	Having access to data in a reasonable amount of time
Usability	Data can satisfy the user's needs
Reliability	Data is complete, from a trusted source, consistent, and meets integrity criteria
Relevance	"Fitness of data" – data can possibly answer what an organization wants to know
Presentation Quality	Format of data is clear / understandable

ize the results are Apache Cassandra, MongoDB, Monet, and Hazelcat (Mikalef et al., 2018, p. 563).

4.1.3. Infrastructure

Wamba et al. (2017, p. 17) suggest three elements that BDA infrastructure is composed of: *connectivity*, *compatibility*, *modularity*. Here, *connectivity* grasps that all offices and departments have access to / can provide their data for BDA. *Compatibility* is a further step of connectivity, which means that not only the data flow is ensured among departments/offices, but also that analysis can be conducted on different platforms and consequently results are accessible for all. Lastly, *modularity* describes an efficiency component: can software modules be re-used or adjusted for different tasks?

Sometimes, also software solutions as Apache Hadoop and cloud-based computing are called in regards to infrastructure (Akter et al., 2016, p. 13). In this thesis, this falls in the category of Software and IS (s. 4.1.2).

ϕ

According to LaValle et al. (2011, p. 23) the most difficult adoption that organizations face are not rooted in tangible assets as data and technology, but in managerial and cultural dimensions. This will be further investigated in the following two pillars.

4.2. Human Resources

The second pillar in the BDA capabilities framework is about *human resources* (Gupta & George, 2016; Mikalef et al., 2018). Between both papers only a minor difference occurred, as Gupta and George (2016) named the knowledge characteristic toward employees *Technical Skills* that was specified by Mikalef et al. (2018), calling it *Data Analytics knowledge*.

4.2.1. Managerial Skills

Companies do not just perform better because of more or better data, but "because they have leadership teams that set clear goals, define what success looks like, and ask the right questions" (McAfee et al., 2012, p. 8). Gupta and George (2016, p. 1053) explained further that it is in the manager's responsibility to foresee potential of generating insights from data and they then must have a good understanding of how and where to apply the insights retrieved from their technical teams.

Also, Mikalef et al. (2018) stated that it is more important for managers to learn how to access companies' big data sets in the best possible way and how to perform BDA that is aligned with an organization's competitive strategy than to "simply perform raw data analysis on large data sets without a clear direction" (Mikalef et al., 2018, p. 572) or unclear contribution to overall firm strategy.

Moreover, also in contrast to employees' technical skills that can be developed by hiring new staff, managerial skills are highly firm specific and thus must be internally developed over time (Gupta & George, 2016, p. 1053).

To enhance this management capability, research suggests improving the quality of planning, investment, coordination, and control (Akter et al., 2016, p. 41).

4.2.2. Data Analytics Knowledge

A challenging question for managers is how to structure technical and human resources in a firm. The implementation of *competency centres* as a centralized approach to bundle analytics competence in one place have be considered as an effective concept, even though the idea of *competency centres* is also controversial, because of the difficulty to reach an alignment between technical and business units (Günther et al., 2017, p. 197).

In terms of *Data Analytics Knowledge* different roles with respective skills are proposed in literature. For that, Ajah and Nweke (2019) describe three distinct teams. They suggest an (1) *Analytics Team*, compromised of *Data Scientist* and *Business Analysts / Data Analysts*. While the *Data Scientist* is required to have more the technical background with "advanced skills in mathematics and statistical modelling" (Ajah & Nweke, 2019, p. 20), a *Business Analysts* should be aligned with business goals and strategic directions. They also have technical knowledge and are skilled in varying data modelling tools, nevertheless, they primarily keep a bird's-eye view to oversee the alignment of analytics with business strategy / performance outcome.

Additionally, Ajah and Nweke (2019, p. 21) describe the (2) *Big Data Architect Team*, compromising of *Global Architects / Platform Engineers* and *Data architect / Data Wranglers*. Global architects are experts in supercomputing platforms, focusing their work on performance tuning and root cause analysis.

Data architects are important because of their industry knowledge and further for their strong mathematics/statistics skills and thus are mostly working on special use cases (Ajah & Nweke, 2019, p. 21).

Lastly, Ajah and Nweke (2019, p. 21) describe the (3) *Big Data Hadoop operators team* consisting of *Hadoop Engineers*, responsible for cluster performance, debugging etc., while *Hadoop Operators* take care of the quality of results. To emphasize Hadoop teams that much might be surprising first, but it is nevertheless the most prominent BDA tool and needs knowledge and skills that exceed the abilities found in most IT departments (McAfee et al., 2012, p. 8).

Overall, McAfee et al. (2012, p. 8) suggest that next to statistical knowledge for the core analytics part, also skills as cleaning and organizing large data sets will become increasingly demanded as well as tools and techniques for visualising results.

Mikalef et al. (2018, p. 565-566) have proposed an additional dimension to the otherwise homogenous dimensions of skills and knowledge that are offered in literature (*Technical Knowledge*, *Business Knowledge*, and Business Analytics Knowledge):

Since teams with an analytics background must work well together with teams that have a business background, "Relational Knowledge" as the "communication and collaboration skills between employees of different backgrounds" (Mikalef et al., 2018, p. 565-566) do play an important role in ensuring strong results from the different teams.

4.3. Intangible Resources

The third and last pillar in the BDA capabilities framework is about *intangible resources* (Gupta & George, 2016; Mikalef et al., 2018). While Gupta and George (2016) have also included the pillar *Intensity of Organizational Learning* this was not considered by Mikalef et al. (2018). Reason for this could be that already existing aspects on learning or knowledge of employees are already covered in the second *Human resources* pillar.

4.3.1. Data-driven Culture

For the implementation of a successful BDA process, researchers suggest a change in the organization's culture to reach a data-driven culture. The baseline is to build a culture where "decisions [are taken] based on data rather than intuitions" (Gupta & George, 2016, p. 1051).

Here it is important to distinguish between the core BDA teams that possess extensive analytics knowledge and are responsible for developing insights from big data and must identify which parts of an organization have the highest potential when analytics is brought in. Whereas other business units' employees from HR, marketing etc. must adopt to a certain degree to a data-centric thinking (a) to increase their knowledge on analytical models and technologies that are or will be used in their department, and (b) to question their daily work routine: Where can analytics be helpful for my work or department? When do we base our decisions on intuition that could be better evaluated with data?

To implement a data-driven culture, it is not enough to support data-to-knowledge initiatives, an organization must ensure leaders behave as a role model and insist that also others take decisions based on data rather than intuition (Davenport, Harris, De Long, & Jacobson, 2001, p. 136).

Regarding culture, a further extension grasps the alignment of IT and business strategy. Akter et al. (2016) have investigated a mediating effect of business strategy alignment on firm performance via BDA capabilities for which they have found a positive significant relationship. Besides, Akhtar et al. (2019) could also show a positive significant relationship among big-data-savvy teams, big-data-driven actions, and firm performance. Moreover, Court (2015) found that organizations can increase operating margins by more than 60% when they can ensure a good alignment between analytics efficiency and business strategy.

4.3.2. Governance

In newer research, experts underscore the importance of data governance (Llave, Hustad, & Olsen, 2018). In literature, two distinctive specifications are offered regarding governance in BDA.

For the first specification, governance grasps compliance with privacy regulations (legal boundaries of how much data may be collected / what degree of personalisation is allowed) and ensuring security of data (data must be secured against hacker attacks). Ajah and Nweke (2019, p. 23) state that organizations often (if so) only have immature tools and practices in place to ensure privacy laws and data security of BDA. For Hadoop for instance, extensions as Kerberos, an open-source project initiated at the MIT, are needed to store data encrypted on Hadoop clusters.

This specification of governance only came later to a scientific discussion. Also both papers of Gupta and George (2016); Mikalef et al. (2018) have not mentioned this characteristic, but solely focused on the second specification in governance:

The second specification focuses on internal rules to ensure an effective use of BDA. Specifically for BDA, Tallon, Ramirez, and Short (2013) propose a governance framework consisting of three types: (a) structural (defining responsibilities, directing, and planning), (b) procedural (cost control, resource allocation, and shaping user behaviours trough value analysis), and (c) relational (alignment / partnerships between IT and business units, idea exchange and conflict resolution).

To ensure that governance is taken to action, literature suggest that an organization must follow a top-down approach, which also implies that top management must commit to data-driven decisions (Vidgen, Shaw, & Grant, 2017) (this is also important for a successful IT/business alignment, s. 4.3).

Concluding, many researchers stress the importance of having governance schemes implemented when working with BDA. Some researchers even see a missing or badly implanted governance model as a main reason why companies cannot extract full value of big data (Mikalef et al., 2018, p. 565).

5. BDA Capabilities Compared to IT and Digital Capabilities

5.1. Overview IT Capabilities

In the 1990s first papers were published in which IT capabilities (Sabherwal & Kirs, 1994) were defined – in contrast to BDA capabilities in the early 2010s (LaValle et al., 2011). On the first sight, IT capabilities are not only older than BDA capabilities, but also facilitate “day-to-day running of the firm” (Bharadwaj, 2000, p. 175). Moreover, BDA specialists need significantly different skills to IT specialists in a firm (Gupta & George, 2016, p. 1061).

Nevertheless, researchers of BDA and IT capabilities both use the resource-based view to describe the mechanism how these capabilities can enhance value creation in a company (e.g., in terms of overall performance, innovation performance, (Hao et al., 2019; Wamba et al., 2017)). This is not too surprising as both – IT and BDA – are a firm resource that, if it is valuable, rare, inimitable, and non-substitutable, and if this resource is managed well (s. 2.2.1), can lead to a competitive advantage over competitors.

While the theoretical mechanism relies on the same theory, it is not trivial to distinguish the frameworks. Even though Gupta and George (2016, p. 1050) have adopted their framework (tangible-human-intangible resources) from the IS literature, content-wise they do differ in some aspects.

For this thesis, the following IT capabilities papers are considered for a comparison to the BDA capabilities (s. Tab. 4).

5.1.1. Tangible Resources

Structure-wise, in both capabilities frameworks data and technology are defined as important tangible resources. Nevertheless, they differ content-wise.

BDA capabilities are far more specified than IT capabilities. For BDA, exact technologies are defined as machine learning or optimization (s. 2.1.2). Moreover, precise software solutions as Hadoop (s. 4.1.2) are described. Still, the biggest difference between both capabilities lies in data itself. In BDA, data is of major importance, while it is not important per se in a company’s IT.

In addition, “big data IT fundamentally differ from traditional IT” (Miloslavskaya & Tolstoy, 2016, p. 301). While for traditional IT a processing device (computer, cloud) is put at the centre of a data processing process, big data IT is seen as a “continuous flowing substance” (Miloslavskaya & Tolstoy, 2016, p. 301). The key lies in the constant float rates of data (which is not a priority for traditional IT), otherwise a theoretical infinite overload of data could occur.

IT capabilities, in contrast, are much broader regarding its tangible elements. Here, the digital support of initial analogous processes is of central importance to an IT department (e.g., CRM, ERP systems).

5.1.2. Human Resources

In all papers the human factor was identified to play an important role in building capabilities. On the one hand, it was described that employees with IT knowledge are needed (for IT capabilities, e.g.: “firm -relevant IT knowledge and competence” (Ross et al., 1996, p. 33)). On the other hand, also support from top management was described to be important.

IS literature, nevertheless, limits the need for IT competence to IT staff and suggests that through the mix of IT and business units the IT staff can gain a “business understanding” (Ross et al., 1996, p. 33). Especially for the development of IT accompaniment of business processes (e.g., digitalization of an ordering process, implementation of an ERP system, or similar), it is understandable why a good skilled IT staff is recommended to understand the needs of business units to then offer suitable IT solutions to the business units.

This stands in wide contrast to BDA capabilities. Certainly, researchers suggest specific knowledge of some employees to perform analytics on big data. But literature also states the need of a comprehensive change in culture – from intuitive decision making to a data-driven culture (s. 4.3.1).

5.1.3. Intangible Resources

In BDA research the focus of intangible resources lies on a data-driven culture and the intensity of organizational learning (s. 4.3). For IT capabilities, intangible resources are focused differently on knowledge assets, customer orientation and synergies (Bharadwaj, 2000, p. 174).

Similarly, BDA and IT capabilities encompass characteristics of a learning organization to enhance adoption to novel BDA or IT solutions (BDA: as a capability: organizational learning, IT, as a resource: knowledge assets).

In fact, IT capabilities stress two other components: *customer orientation* and *synergies*.

In general, *customer orientation* has been found to positively impact firm performance (Jaworski & Kohli, 1993). To satisfy a customer orientation in a firm, often IT is used (e.g., for a classic CRM system). But also, to better understand and predict customer needs, IT relates to a firm’s strategy at its foundation.

Synergies describe the idea that IT systems can be re-used in a company (e.g., in different departments, countries, similar cases).

To sum up, BDA intangible resources point on organizational aspects, while IT intangible resources pin on efficiency and quality aspects (being customer-centric with IT applications, re-using technology in a company).

5.2. Overview Digital Capabilities

In 2016, Junior, Maçada, Brinkhues, and Montesdioca (2016, p. 2) conducted a literature review on *Digital Ca-*

Table 4: Overview of proposed IT capabilities

Author(s)	Capabilities proposed
Bharadwaj et al. (1999)	IT business partnership, External IT linkages, Business IT strategic thinking, IT business process integration, IT management, IT infrastructure
Bharadwaj (2000)	IT infrastructure, Human IT Resources, IT enabled Intangibles
Chae et al. (2014)	Rebottle of Bharadwaj (2000)
Ross, Beath, and Goodhue (1996)	Technology Asset, Human Asset, Relationship Asset
Santhanam and Hartono (2003)	Rebottle of Bharadwaj (2000)

pabilities, searching the Basket of Eight as well as Ebscohost and Google Scholar. They could identity 26 relevant papers of which only six papers offered a clear definition of digital capabilities.

Besides the small number of definitions, they are also largely heterogenous. Westerman, Bonnet, and McAfee (2012), for instance, define digital capabilities broadly as skills that go beyond pure IT, and even move in the direction of BDA capabilities, speaking of analytics to retrieve value from big data. Also Lyytinen, Yoo, and Boland Jr. (2016) define digital capabilities as an ability with the aim of creating "new outputs, structures, and behaviours [. . .] without deliberate planning from the originator of the system" (Lyytinen et al., 2016, p. 53).

Even though there is no homogenous understanding of digital capabilities in research yet, most definitions point in a similar direction: Researchers expect a firm's functioning IT landscape to be a basic prerequisite and aggregate possible extensions under the umbrella of digital capabilities.

5.3. Comparison of the Different Types of Capabilities

IT capabilities have been found to explain how organizations can transform analogous work and processes into digital processes. Doing so, different systems as CRM and ERP systems can be implemented that offer not only more efficient solution for companies, but also provide useful extensions compared to analogous old processes (as connecting an ERP system with suppliers to automatically re-order etc.).

Digital Capabilities are different, as they do not primarily focus on the digital transformation in firms but focus on finding extensions that technology offers to their business. Thus, digital capabilities can be seen as a broad extension of IT capabilities, while BDA capabilities – as a very specific extension of digital capabilities – form a part of digital capabilities.

To also understand how the three technical capabilities – IT, digital, and BDA capabilities – relate to each other, a good indicator is to look at the chronological development (s. Fig. 6).

For most companies, IT became a topic in the beginning of the 1970s when first computers moved in corporations. Researchers and practitioners have deeply discussed the potential and possible pitfalls of bringing IT in an organization. One major discussion was the emergence of the *IT productivity paradox* (s. 2.4). Researchers could not show how higher IT spending would result in higher performance, but eventually this paradox could be resolved (Gupta & George, 2016).

Furthermore, researchers were already discussing IT Capabilities in the 1990s before the term big data with its popular 3 V's characteristic has been defined in 2001 (Laney, 2001). And another ten years later – after the successful implementation of today's widely used software framework of Hadoop in 2005 (Apache, 2021) – first papers about BDA capabilities were published (LaValle et al., 2011).

With the accelerating supply of cloud computing in the beginning of the 2010s, the processing of big data became easier and - most important - cheaper for organizations, which has most probably also given a boost to BDA usage (West, Battleson, Kim, & Ramesh, 2014; Winegar, 2021). This aspect was out of scope of this paper and thus not further discussed. Nevertheless, it is still relevant to name cloud computing when looking at the chronological development of BDA.

6. Discussion

6.1. Pitfalls for organizations after implementing Big Data Analytics

To answer the first research question, this thesis could clearly show and discuss proposed BDA capabilities frameworks. Nevertheless, a further discussion in conducted to also answer the second part of the first research question: "How do BDA capabilities influence the success of BDA?"

6.1.1. Tangible Resources

Researchers commonly describe distributed and real-time analytics as the next trends in BDA's infrastructure (Landset et al., 2015).

In terms of data itself, on possible pitfall for organizations is to concentrate on building big data resources and improving the BDA process, but not implementing quality checks on the data to "screen out redundant, inaccurate, and duplicated data" (Boldosova, 2019, p. 1142).

Apart from quality, quantity is a major game changer in the world of big data. Google's director of research, Peter Norvig, once stated: "We don't have better algorithms. We

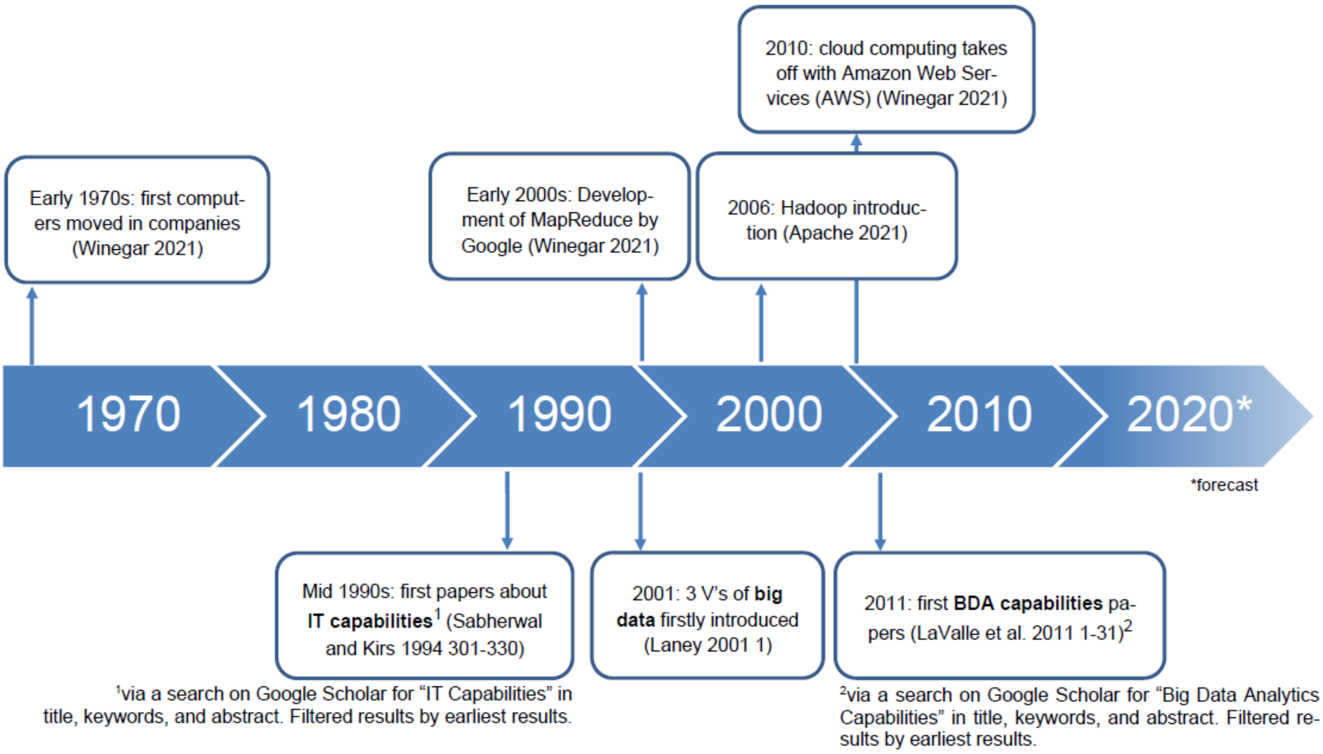

Figure 6: Timeline of technological developments

just have more data." (McAfee et al., 2012, p. 5). In the same direction pins a report from the Executive Office of the President of the United States suggesting that: "technical capabilities of big data have reached a level of sophistication" (Podesta, Pritzker, Moniz, Holdren, & Zients, 2014, p. 5).

It seems that in terms of BDA, data itself can be the differentiator for companies. Ultimately, reaching a competitive advantage in BDA could come down to the simple size of data available to an organization.

6.1.2. Human Resources

A key factor in terms of BDA human resources is the availability of skilled staff. In 2011, McKinsey has predicted a shortage in the US of 140,000 to 190,000 big data skilled individuals by 2018. However, a study by Bain from 2018 also said that they expect that the "global supply of advanced analytics talent will double" (Brahm, Sheth, Sinha, & Dai, 2019, p. 1) from half a million to one million within the next two years.

Another reason why finding skilled employees might not be the key difficulty within the execution of BDA capabilities is offered by a further study from McKinsey in 2020 on AI. In this article it is questioned whether an organization still needs to hire that many data scientist to build machine learning models (Hürtgen, Kerkhoff, Lubatschowski, & Möller, 2020). In this case (based on AI, not BDA), the authors "don't foresee demand for substantial, functional data-science expertise going away anytime soon" (Hürtgen et al., 2020, p. 4). Nevertheless, one can expect that also the field of BDA will become more accessible as tools and techniques can get easier to handle also for not specialised employees.

6.1.3. Intangible Resources

After implementing the first BDA pipelines (big data via analytics to results), it is easy to miss closing the cycle to full value creation. One possible scenario are managers that now have access to novel data insights, but still keep to a decision-making process based on intuition (Mikalef et al., 2018, p. 571). Referring to the BDA capabilities framework discussed in the literature review, this occurrence is mostly connected to a bad execution of an intuition-based culture to a data-driven culture. Especially managers are visible as role models and thus must take the results of BDA to decisions and value creation (Davenport et al., 2001, p. 136). Also, McAfee et al. (2012) could show that data-driven firms are six percent more profitable and five percent more productive than their rivals.

Another pitfall regarding the organizational aspect is to implement BDA only in specific departments as marketing (Mikalef et al., 2018, p. 570). This also hinders an organization wide change to a data-driven culture.

Furthermore, one aspect that only came up recently and was not considered in the papers from Gupta and George (2016) and Mikalef et al. (2018) is *governance*. When working with data, firms might see themselves confronted with a dilemma of collecting as much data as possible on the one side and obeying governmental privacy rules (and ultimately limiting the value of BDA) on the other side (Llave et al., 2018). Moreover, data regulations are not settled yet, and changes in the next few years can be expected.

6.2. BDA: Temporary trend versus mega game changer

Concerning the second research question, BDA capabilities were compared to IT and digital capabilities in the second part of this thesis. Here, papers were investigated whether BDA can be seen as a new phenomenon or BDA rather is "old wine in new bottles".

Looking back, the aspiration of making sense of data is nothing new – the statistical field dates to the 18th century. Still, a clear enhancement in analytics has occurred over the past years. Now, data not only moves online at a rapid pace, but can also be analysed in (near to) real-time (Agarwal & Dhar, 2014, p. 444).

This development has highly increased possibilities that can be reached with BDA. Today, the old – more theoretical - idea of making sense of data can eventually lead to big impact in two ways. First, as described in the introduction of this thesis, several papers have already shown that BDA has a positive influence on firm performance (s. 1). Secondly, the new tools and techniques offer new (business) possibilities and insights that could have not been detected before (Wiener, Saunders, & Marabelli, 2020). In this context, literature stresses the importance for incumbent organizations to "rethink their existing business models and how these may be affected by big data" (Günther et al., 2017, p. 197).

A difficult aspect to judge is its chance of substitution – simply put: Will BDA still be important in the next years, while new technical possibilities as AI are emerging, too? To evaluate this question, it is useful to consider the industry an organization is working in.

Regarding adaptability of BDA in an organization, some managers may shrink when thinking about the implementation of BDA and be unsure about the final payoffs for moving from an institutional to an data-driven organization (Gupta & George, 2016). This thesis could show that BDA capabilities are indeed different to IT capabilities and with that the implementation is not straight forward as one might expect. Managers cannot simply rely on their strong IT power of their firm to also conduct analytics. Certainly, literature does suggest similar fields a company has to pay attention to when working with IT or BDA (e.g., both do incorporate human skills), but the needed abilities and adoptions for BDA are far more specific compared to an IT departments' abilities. Here, Barton and Court (2012, p. 81) also suggest that especially managing unstructured data remains beyond IT capabilities. Not to forget, for BDA (in difference to IT) a cultural change among all department is needed. One common misadaptation of BDA is to only use it in specific operational departments as marketing or finance (Mikalef et al., 2018, p. 570), which hinders BDA to spread its full potential.

Concluding, the conceptual idea of analytics has been around for a long time, but for only some years it is feasible to reach high impact results (with new tools and techniques that can handle large and diverse amounts of data quickly). A good IT infrastructure is helpful, but to conduct BDA a firm must invest in the more specific BDA capabilities to reach satisfying results (Barton & Court, 2012). Thus, BDA from its structural architecture is similar to IT capabilities but looking at the specificity of BDA capabilities suggest that they are more than simply "old wine in new bottles".

7. Implications, Limitations, and Outlook

7.1. Implications

This thesis enriches literature on BDA capabilities by presenting researchers' most recent views on the topic of BDA capabilities and delimiting its boards to similar areas of IT and digital capabilities.

First, a clear distinction of important terms regarding big data was proposed, which has not been separated by literature that clearly, yet. With it, also theory frameworks are better accessible, as this paper could provide a clear differentiation between the resource-picking activities that entail the creation of accessible big data, and its capability-building activities that rotate around implementing BDA (s. 2.2).

In the structured literature review different frameworks were shown and discussed and an overview of current's literature status quo was provided. Moreover, a deep dive beyond the core framework of proposed capabilities was conducted, describing in detail what characteristics each capability entails (Gupta & George, 2016; Mikalef et al., 2018).

In the second part of this thesis, BDA was compared to IT and digital capabilities. This differentiation is a novel approach and enriches current literature to offer a clear distinction between the three technical capabilities. Moreover, this thesis could show how BDA has emerged in the beginning of the 2000s and developed since then.

7.2. Limitations

Also, this thesis has limitations. The main limitation is due to the methods used. First, the research was narrowed on papers that matched specific search words ("big data analytics", "big data analytics capabilities") in their title, keywords or abstract. Doing so, studies were filtered that have not used the exact wording but could have added valuable insights. Another problem of literature reviews in general is the chance to miss complementary insights from other fields. To partly prevent this, the research in the second part of this paper was helpful, where near-field capabilities (IT and digital capabilities) were considered and compared to BDA capabilities.

7.3. Outlook

This thesis could also show how BDA has emerged in the early 2000s (s. 5.3). Now, we already see formations of new applications in various directions. A considerable example form Gupta and George (2016) and Mikalef et al. (2018) with their highly cited paper on BDA capabilities. This year, Mikalef (who has written the extension on Gupta's BDA paper) and Gupta have published a paper on Artificial Intelligence (AI) capabilities together (Mikalef & Gupta, 2021). Gupta has described this paper himself to be an extension to his initial BDA paper.

In this thesis, capabilities from other technological areas were identified and discussed, still, both (IT and digital capabilities) have emerged before BDA. Thus, it is interesting to investigate which new technologies have emerged after BDA. Moreover, no clear distinction is made, yet that weighs parameters to help organizations to decide which technologies should be implemented next or could be potentially substitute by a later technology (or, e.g., could a company skip the implementation of BDA and directly implement AI extensions?).

8. Conclusion

With all the focus of this thesis on needed capabilities for firms to extract the maximum value for firms, one should not forget that big data is not only helpful for organizations. Big data is also helpful for people.

In healthcare, for instance, big data tools are used to support the early detection of diseases or during the COVID-19 crisis to offer an overview about a country's status in the pandemic (Ajah & Nweke, 2019, p. 23).

Also in our daily life, we get in touch with BDA even though we might not notice it at first. When watching Netflix, we value the highly ranked recommendation system that suggests new movies and series to watch – this is possible because of BDA (Lycett, 2013). When driving with our car, we value to have real-time information on traffic – this is possible because of BDA (Ajah & Nweke, 2019, p. 25). When shopping online, we value that stores can predict what we want – before we even know it (Ajah & Nweke, 2019, p. 25).

This thesis has focused on implications for companies when implementing BDA in their work. But one should not forget – big data is not just some company advancement project; it is impacting peoples' lives.

References

Agarwal, R., & Dhar, V. (2014). Big data, data science, and analytics: The opportunity and challenge for IS research. *INFORMS*.

Agrawal, S., & Sureka, A. (2013). *Detecting copyright infringement on YouTube videos using YouTube metadata* (Unpublished doctoral dissertation).

Ajah, I. A., & Nweke, H. F. (2019). Big data and business analytics: Trends, platforms, success factors and applications. *Big Data and Cognitive Computing*, *3*(2), 32.

Akhtar, P., Frynas, J. G., Mellahi, K., & Ullah, S. (2019). Big data-savvy teams' skills, big data-driven actions and business performance. *British Journal of Management*, *30*(2), 252–271.

Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? *International Journal of Production Economics*, *182*, 113–131.

Apache, H. (2021). *Hadoop Releases - an Apache Webpage.* https://archive.apache.org/dist/hadoop/common/. (Accessed: 2021-06-25)

Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1), 99–120.

Barney, J. B., Ketchen Jr., D. J., & Wright, M. (2011). The Future of Resource-Based Theory: Revitalization or Decline? *Journal of Management*, *37*(5), 1299–1315.

Barton, D., & Court, D. (2012). Making advanced analytics work for you. *Harvard Business Review*, *90*(10), 78–83.

Bharadwaj, A. S. (2000). A resource-based perspective on information technology capability and firm performance: an empirical investigation. *MIS quarterly*, 169–196.

Bharadwaj, A. S., Sambamurthy, V., & Zmud, R. W. (1999). IT Capabilites: Theoretical Perspectives And Empirical Operationalization. *ICIS*, 378–385.

Boldosova, V. (2019). Deliberate storytelling in big data analytics adoption. *Information Systems Journal*, *29*(6), 1126–1152.

Brahm, C., Sheth, A., Sinha, V., & Dai, J. (2019). *Solving the New Equation for Advanced Analytics Talent* (Tech. Rep.).

Brown, S. L., & Eisenhardt, K. M. (1997). The art of continuous change: Linking complexity theory and time-paced evolution in relentlessly shifting organizations. *Administrative Science Quarterly*, 1–34.

Brynjolfsson, E. (1993). The productivity paradox of information technology. *Communications of the ACM*, *36*(12), 66–77.

Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data Science Journal*, *14*.

Chae, H.-C., Koh, C. E., & Prybutok, V. R. (2014). Information technology capability and firm performance: contradictory findings and their possible causes. *MIS quarterly*, *38*(1), 305–326.

Choi, T.-M., Wallace, S. W., & Wang, Y. (2018). Big data analytics in operations management. *Production and Operations Management*, *27*(10), 1868–1883.

Court, D. (2015). Getting big impact from big data. *McKinsey Quarterly*, 1–8.

Davenport, T. H., Harris, J. G., De Long, D. W., & Jacobson, A. L. (2001). Data to knowledge to results: building an analytic capability. *California Management Review*, *43*(2), 117–138.

Davenport, T. H., Harris, J. G., & Morison, R. (2010). *Analytics at work: Smarter decisions, better results*. Harvard Business Press.

Dixon, J. (2010). *Pentaho, Hadoop, and Data Lakes.* https://jamesdixon.wordpress.com/2010/10/14/pentaho-hadoop-and-data-lakes/. (Accessed: 2021-06-16)

Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic capabilities: what are they? *Strategic Management Journal*, *21*(10-11), 1105–1121.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144.

Günther, M., Wendy Mehrizi, Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *Journal of Strategic Information Systems*, *26*, 191–209.

Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Elsevier: Information & Management*, *53*, 1049–1064.

Hao, S., Zhang, H., & Song, M. (2019). Big Data, Big Data Analytics Capability, and Sustainable Innovation Performance. *Sustainability*, *11*, 1–15.

Hu, W., Xie, N., Li, L., Zeng, X., & Maybank, S. (2011). A Survey on Visual Content-Based Video Indexing and Retrieval. *IEEE Transactions on Systems*, *41*(6), 797.

Hürtgen, H., Kerkhoff, S., Lubatschowski, J., & Möller, M. (2020). *Rethinking AI talent strategy as automated machine learning comes of age* (Tech. Rep.).

Jaworski, B. J., & Kohli, A. K. (1993). Market orientation: antecedents and consequences. *Journal of Marketing*, *57*(3), 53–70.

Junior, J. C. d. S. F., Maçada, A. C. G., Brinkhues, R., & Montesdioca, G. P. Z. (2016). *Digital Capabilities as Driver to Digital Business Performance*. AMCIS.

Keller, W. (2009). Using Capabilities in Enterprise Architecture Management [White paper].

Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management*, *34*(3), 387–394.

Lamba, H. S., & Dubey, S. K. (2015). Analysis of requirements for big data adoption to maximize IT business value. *IEEE - 4th International Conference on Reliability, Infocom Technologies and Optimization*, 1–6.

Landset, S., Khoshgoftaar, T. M., Richter, A. N., & Hasanin, T. (2015). A survey of open source tools for machine learning with big data in the Hadoop ecosystem. *Journal of Big Data*, *2*(1), 1–36.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*, *6*(70), 1.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, S. (2011). Big Data, Analytics and the Path From Insights to Value. *MITSloan Management Review*, 1–31.

Llave, M. R., Hustad, E., & Olsen, D. H. (2018). *Creating value from business intelligence and analytics in SMEs: Insights from experts*. AMCIS.

Lycett, M. (2013). "Datafication': making sense of (big) data in a complex world. *European Journal of Information Systems*, *22*, 381–386.

Lyytinen, K., Yoo, Y., & Boland Jr., R. J. (2016). Digital product innovation within four classes of innovation networks. *Information Systems Journal*, *26*(1), 47–75.

Makadok, R. (2001). Toward a synthesis of the resource-based and dynamic-capability views of rent creation. *Strategic Management Journal*, *22*(5), 387–401.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard Business Review*, *90*(10), 60–68.

Mikalef, P., Framnes, V. A., Danielsen, F., & Krogstie, J. (2018). *Big Data Analytics Capability: Antecedents and Business Value* (Vol. 16). Springer-Verlag GmbH Germany.

Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Elsevier: Information & Management*, *58*, 1–20.

Miloslavskaya, N., & Tolstoy, A. (2016). Big data, fast data and data lake concepts. *Procedia Computer Science*, *88*, 300–305.

Podesta, J., Pritzker, P., Moniz, E., Holdren, J., & Zients, J. (2014). Big Data: Seizing Opportunities, Preserving Values. *Executive Office of the President*.

Ransbotham, S., Kiron, D., & Prentice, P. K. (2015). *Minding the analytics gap* (Vol. 56). MIT Sloan Management Review.

Rosenthal, D. S. H., Rosenthal, D. C., Miller, E. L., Adams, I. F., Storer, M. W., & Zadok, E. (2012). The Economics of Long -Term Digital Storage. *The Memory of the World in the Digital Age: Digitization and Preservation*, 1–17.

Ross, J. W., Beath, C. M., & Goodhue, D. L. (1996). Develop long-term competitiveness through IT assets. *Sloan Management Review*, *38*(1), 31–42.

Russom, P. (2011). Big Data Analytics. *The Data Warehousing Institute*.

Sabherwal, R., & Kirs, P. (1994). The alignment between organizational critical success factors and information technology capability in aca-

demic institutions. *Decision Sciences*, *25*(2), 301–330.
Santhanam, R., & Hartono, E. (2003). Issues in linking information technology capability to firm performance. *MIS quarterly*, 125–153.
Seddon, P. B., Constantinidis, D., Tamm, T., & Dod, H. (2017). How does business analytics contribute to business value? *Information Systems Journal*, *27*(3), 237–269.
Soltanpoor, R., & Sellis, T. (2016). *Prescriptive Analytics for Big Data*. Springer International Publishing.
Soto Setzke, D., Opderbeck, L., Böhm, M., & Krcmar, H. (2020). Pathways to Successful Business Model Innovation in the Context of Digital Transformation. *24th Pacific Asia Conference on Information Systems*.
Tallon, P. P., Ramirez, R. V., & Short, J. E. (2013). The information artifact in IT governance: toward a theory of information governance. *Journal of Management Information Systems*, *30*(3), 141–178.
Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, *18*(7), 509–533.
Tiguint, B., & Hossari, H. (2020). Big data analytics and artificial intelligence: a meta-dynamic capability perspective. *changes*.
Vidgen, R., Shaw, S., & Grant, D. B. (2017). Management challenges in creating value from business analytics. *European Journal of Operational Research*, *261*(2), 626–639.
Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J.-F., & Dubey, R. (2017). Big data analytics and firm performance: effects of dynamic capabilities. *Journal of Business Research*, *70*, 356–365.
Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Elsevier: Technological Forecasting & Social Change*, *126*, 3–13.
Webster, J., & Watson, R. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, *26*(2), 13–13.
West, B. C., Battleson, D. A., Kim, J., & Ramesh, B. (2014). Achieving dynamic capabilities with cloud computing. *IT Professional*, *16*(6), 18–24.
Westerman, G., Bonnet, D., & McAfee, A. (2012). The digital capabilities your company needs. *MIT Sloan Management Review*, 1–5.
White, T. (2012). *Hadoop: The definitive guide*. O'Reilly Media, Inc.
Wiener, M., Saunders, C., & Marabelli, M. (2020). Big-data business models: A critical literature review and multiperspective research framework. *Journal of Information Technology*, *35*(1), 66–91.
Winegar, J. (2021). *What is "Big Data" – Understanding the History*. https://towardsdatascience.com/what-is-big-data-understanding-the-history-32078f3b53ce. (Accessed: 2021-06-15)
Yasmina, M., Tatoglua, E., Kilicb, H. S., Zaimc, S., & Delen, D. (2020). Big data analytics capabilities and firm performance: An integrated MCDM approach. *Elsevier: Journal of Business Research*, *114*, 1–15.