Data as a Corporate Resource

Tracing the Path from Investments to Value

F.E.C. Voets

This page was intentionally left blank

Data as a Corporate Resource: Tracing the Path from Investment to Value

A study to determine the organisational capabilities required to successfully become data-driven

A thesis submitted in fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in Management of Technology

At Delft University of Technology

Faculty of Technology, Policy, and Management

In collaboration with KPMG the Netherlands

By

By F.E.C. Voets

July 2019

Graduation committee:

First supervisor: M.F.W.H.A. Janssen
Second supervisor: C. Werker
KPMG supervisors: J. Hendriks and R. Akça

This page was intentionally left blank

Executive Summary

Due to its increasing availability, data has become a popular topic among both scholars and industrialists; the expression "the oil of the digital era" only begins to describe the craze surrounding the term. Today, there is a growing realisation that data is not simply a by-product of organisations' primary processes, and that it is a valuable resource by itself. Data can improve decision-making and business performance, and is therefore seen as a low-hanging key to a successful company. However, despite these promises surrounding data, there does not seem to be a direct link between investments in data initiatives and improved company results. In this thesis, this discrepancy is termed the data value paradox.

The data value paradox displays many parallels with the IT productivity paradox that occupied scholar's minds several decades ago. For this reason, a literature review was performed from both a historical and a modern-day perspective. By building on these findings and two organisational frameworks, i.e. the resource-based and the capability-based theory of the firm, it was identified that the data value paradox is rooted in two main causes. First, it is often forgotten that the widely-available assets obtained through data investments do not provide a competitive advantage on their own. In order to do so, these need to be translated into unique, company-specific capabilities. Second, in both literature and industry, there is a strong focus on the technology required to properly employ data in organisations. Although this is an indispensable element for successfully extracting value from data, there is a network of underlying business and data management mechanisms which is equally important but often forgotten.

Based on these findings, a knowledge gap was identified: no explicit overview of the capabilities that allow an organisation to successfully become data-driven exists. As the process of organisational change is a complex one requiring as much guidance as possible, the aim of this thesis was to create this overview. This led to the main research question: *Which capabilities should organisations develop to facilitate overcoming the data value paradox?* Because of the strong technology focus that usually exists, the scope was confined to business and data-specific capabilities. It was chosen to combine several qualitative techniques to address the research question, as this allows a more holistic analysis to be made.

A first step towards answering the research question, was conceptualising the path an organisation follows in becoming more data-driven. This was done by reviewing existing publications and frameworks on the topic, and then integrating these into a model consisting of five data maturity stages. This model showed that an organisation becomes data-driven through both incremental and radical changes that build on each other over time; throughout the different stages, data management practices slowly become more and more centralised and advanced. Examining this path in more detail also revealed that it is crucial to develop the right capabilities at the right time. Developing capabilities too soon may cause organisations to be overwhelmed, whereas developing them too late will hinder improvement.

As qualitative techniques were used, replicability to increase rigor was an important factor throughout this thesis. Therefore three iterative rounds of research were conducted to determine the capabilities an organisation should develop. Each round was chosen to both validate and enrich previous findings.

In a first round, nine case studies were performed to establish a first overview of the capabilities to be developed. Cases were selected based on availability, recency, and content. Selected cases focused on organisations with the intention of becoming more data-driven. Reports included an assessment of their current state, their ambitions, and the recommendations to fulfil these. Case study findings were then validated through three interviews, which were simultaneously used to sketch more context and therewith enrich results.

Case studies and interviews were conducted in two different industries: manufacturing and financial services. This was done to determine whether data-driven capabilities were generally applicable throughout various industries, or whether the advice should be tailored to industry-specific needs. Furthermore, the fact that similar capabilities are present in different industries is seen as extra validation of previous findings. The reasons for choosing these specific industries were twofold. On the one hand, they could be attributed to two significantly differing sectors: goods and services. The more divergent the sectors, the more rigorous the claims about generalisability. On the other hand, earlier publications had shown that, within these sectors, manufacturing and financial services carried the most potential for capturing value from data. The findings of this second round of research showed that the majority of capabilities could be applied across different industries and sectors. This can be attributed to the fact that, overall organisations face similar challenges when becoming data-driven. Similar capabilities will therefore have to emerge to overcome these challenges. A handful of capabilities was found to be industry-specific, leading to the insight that unique capabilities can stem from the drivers behind an organisation's data ambitions.

In a final round of validation, the generally applicable practical findings were compared to capabilities derived from existing literature. Most findings were again confirmed, two capabilities were reformulated, and one was added. This resulted in a final list of twenty-one generally applicable capabilities, and three industry-specific capabilities that allow an organisation to become more data-mature.

To answer the research question of this thesis, the generally applicable capabilities were integrated with the data maturity stage model derived earlier. From this it became apparent that a lot of resources have to be invested in an organisation's data management practices at early stages. At this point the organisation is just starting to become more data-mature, and therefore little value in terms of increased business performance will be derived from these investments. The underwhelming return on investment may discourage company management from properly developing these capabilities. The defined capabilities, however, are strongly interrelated, and to successfully overcome the data value paradox, an aggregate of all of them is necessary. This is an important managerial implication, as company management is made aware of the fact that they should carefully develop the organisational foundation on which further data ambitions can be built, even if this means not having immediate pay-offs for investments made.

This research contributed to both research and practice by developing a framework of capabilities that go beyond technology and tooling. This framework explicitly defines the business and data-specific capabilities necessary to develop the organisational foundation on which more complex data initiatives can then be built. As the research into data-driven capabilities is still in a very young state, the findings from this thesis form a starting block for future research to expand on.

This page was intentionally left blank

Preface

This thesis was written in fulfilment of the MSc Management of Technology at Delft University of Technology. It contains the results of a study that aims to identify the capabilities that organisations wanting to become more data-mature should develop in order to successfully realise their ambitions. This research was conducted in collaboration with KPMG the Netherlands.

Herewith I would like to thank KPMG's Enterprise Data Management team for offering me the ability to conduct this research and for enabling me to use KPMG's resources. In particular I would like to thank my supervisors, Jai Hendriks and Ramazan Akça, who always managed to free up time from their busy schedules to provide me with day-to-day guidance. Finally, I would like to thank my faculty supervisors, Marijn Janssen and Claudia Werker, who have always enthusiastically supported this project.

Delft, July 12th 2019

F.E.C. Voets

MSc Student Management of Technology
Delft University of Technology
Faculty of Technology, Policy, and Management

This page was intentionally left blank

Table of Contents

List of Figures

List of Tables

This page was intentionally left blank

1. Introduction

With its remarkable growth rate, data is a concept that cannot be overlooked in this day and age. Studies have predicted that, by 2020, the digital universe will have expanded to a staggering 44 zettabytes (4.4×10^{13} gigabytes); a tenfold increase from its 4.4 zettabytes in 2013 (Dell EMC & IDC, 2014). Another interesting perspective is provided by Domo in their yearly “Data Never Sleeps” publication. They claim that, by 2020, 1.7 megabytes of data will be produced per person per second (Domo, 2018). Furthermore, Bernard Marr states that, over the past three years, 90 percent of all the world’s data has been created (Marr, 2018). The tremendous growth in the availability of data can be contributed to technological developments that allow the world to become more and more digitalised. Data is extracted from numerous sources, some of which are obvious and others are more intricate. Examples include smartphones, e-mails, online records and transactions, online services, sensors, and the Internet of Things (Marr, 2018; Sagiroglu & Sinanc, 2013). This mind-boggling amount of information is subsequently stored in databases, which are quickly expanding in size and hence are becoming more and more complex to manage (Sagiroglu & Sinanc, 2013).

The increasing availability of data has resulted into its value increasing. Extant literature has strongly emphasised the potential of data to transform numerous societal practices. Possible applications lie in the field of e-learning, drug discovery, governmental operations such as law enforcement and traffic control, and many others (Kim, Trimi, & Chung, 2014). This thesis will focus on data as a corporate resource; where it was once considered a by-product of certain business processes, it is now a valuable resource on its own (Martijn & Jonker, 2015). It has been described as “the next management revolution” (McAfee & Brynjolfsson, 2012, p. 4) and “the next frontier of innovation, competition, and productivity” (Manyika et al., 2011, p. 1). A study conducted by the Massachusetts Institute of Technology and IBM has shown a strong correlation between firm performance and data-driven management, significantly increasing measures such as productivity, return on equity, and market value (Brynjolfsson, Hitt, & Kim, 2011). Companies that integrated data into their business operations clearly outperformed their less adapted competitors, overcoming hurdles and seizing opportunities with more ease (LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011). According to these authors, the main rationale behind these claims is that, when using data, managers are less reliant on intuitive decision-making. Instead, firms that succeed in meaningfully collecting and analysing data can make more insightful decisions based on actual evidence. As McAfee and Brynjolfsson (2012, p. 4) put it: “Because of big data, managers can measure, and hence know, radically more about their businesses, and directly translate that knowledge into improved decision-making and performance.”

It should be noted that, in extant literature, the terms data and big data are used interchangeably and in a wide variety of contexts, leading to conceptual vagueness surrounding its meaning (De Mauro, Greco, & Grimaldi, 2015). Throughout the remainder of this thesis, no distinction will be made between these two terms. To avoid confusion, it was chosen to adopt the term “data” rather than “big data”. This thesis will adhere to the definition of data suggested by De Mauro et al. (2015, p. 103): “Information assets characterised by such a high volume, velocity and variety to require specific technology and analytical methods for its transformation into value.”

1.1 The Link between Data and Firm Performance

From the previous it has become clear that modern-day literature has been generating a considerable amount of buzz throughout different industries with respect to the use of data as a corporate resource, even going so far as to call it "the oil of the digital era" (The Economist, 2017). This section will shortly substantiate the claims about the value of data, and will then discuss why data has the potential to generate such a high amount of firm value in a more detailed manner.

Earlier it was stated that a correlation can be observed between data-driven management and a company's subsequent performance. Brynjolfsson et al. (2011) tested this hypothesis by linking the results of structured interviews with financial data from annual reports. The authors found that indeed, there was a positive correlation between companies that characterised themselves as being data-driven and high financial performance. This finding remained significant even after corrections for inter-company variables, such as IT investments and analytical talent. The same relationship was shown in a study by LaValle et al. (2011), who conducted extensive surveys which clearly showed that top-performing firms employed analytics five times more often than low-performers. More practically, Fortune 500's ranking of the 2018 most valuable firms (table 1) also seems to support this claim, as the top five is occupied by well-known data-driven giants.

Table 1. Most valuable companies in the 2018 Fortune 500 List. Adapted from Shen (2018).

Rank	Company	Market value
1	Apple	$921 billion
2	Amazon	$765 billion
3	Alphabet	$750 billion
4	Microsoft	$746 billion
5	Facebook	$531 billion

A broad spectrum of potential ways in which data can provide a firm with added value can be found in existing literature. Several authors seem to agree that the all-encompassing benefit is data-driven decision-making. LaValle et al. (2011) stress the importance of being aware of trends that are currently developing in the market; to some extent, companies should be able to predict what will happen next and make the right decisions based on these insights. The authors claim that, knowing what happened and why after the facts no longer suffices; in today's environment, timely and informed decisions should be made, and data can provide the guidance for doing so. McAfee and Brynjolfsson (2012) emphasize that this also applies to the firm itself; data allows management to gain accurate and numerically-underpinned insights into a company, allowing them to make the right decisions at the right moment. Manyika et al. (2011) add that data-driven decision-making does not necessarily imply that the process becomes fully automated; managerial scrutiny is still required. The authors also provide some examples of such enhanced decision-making: tailoring products and services to meet specific customer needs, increased productivity (e.g. cutting time to market, using fewer resources), increased product quality, and determining the root-causes for performance variability. Furthermore, Gobble (2013) states that data can open doors for innovation, both for new products and for business processes and strategies. The latter allows firms to develop new ways of thinking, which can help them in overcoming certain challenges and can thus be an important source of competitive advantage (Davenport, Barth, & Bean, 2012; Wamba, Akter, Edwards, Chopin, & Gnanzou, 2015).

1.2 Problem Statement

Despite data's promising uses as a corporate resource, companies are struggling to capture its business value (Wamba et al., 2015). As it turns out, the link between investments into turning over a more data-driven leaf and increased firm performance, is not necessarily a straightforward one (Gupta & George, 2016). The difficulty in doing so can be clarified by defining the data that is currently available to companies in terms of three Vs: volume, velocity, and variety (Russom, 2011). Respectively, these terms indicate the enormous amount of data available, the high frequency at which it is created, and its large variety of sources and formats (Wamba et al., 2015). Combined, these characteristics challenge the traditional ways in which data is being used by firms; not only does a company need to extract data from the bulk, often it is also of varying quality and interoperability is not always evident (Constantiou & Kallinikos, 2015; Janssen & Kuk, 2016; McAfee & Brynjolfsson, 2012).

Thus, before a company can fully benefit from becoming data-driven, several managerial challenges must be overcome (Manyika et al., 2011; McAfee & Brynjolfsson, 2012). These challenges are mostly concerned with combining strategic leadership and the organisational changes needed to implement data-initiatives (McAfee & Brynjolfsson, 2012). Organisational changes can, for example, range from acquiring new technical talent to processes as complex as rearranging the established business structure to accommodate data. At the same time, these organisational changes must be in line with a company strategy that is also changing through data. Some difficulties beyond company organisation can also be identified. For example, McAfee and Brynjolfsson identify company culture as a complicating factor, as most executives are used to making decisions based on experience. Being guided by data may feel counterintuitive to them, leading to resistance. Manyika et al. (2011) furthermore stress that the network of privacy, security, and liability issues becomes more intricate with a growing amount of data.

From what was discussed in this section, it has become clear that a discrepancy between on the one hand the promised, and on the other hand the extracted value of data exists. The nature of the data available to organisations seems to form the base of this phenomenon, and it is clear that various managerial challenges must be overcome in order to address it. Measures to do this, however, remains largely unexplored.

1.3 Research Objectives and Approach

Given the problem statement, this thesis aims to determine how organisations can successfully extract value from data. In doing so, the outcomes of this thesis will provide company management with tangible guidance in becoming more data-driven.

This goal will be achieved through a more in-depth exploration of the aforementioned discrepancy. Existing literature in this research domain has already established certain key elements and directions in the pursuit of a data-driven business model. By building on these findings, a literature review will trace the path from investments in data management practices to enhanced firm performance in a theoretical manner. Two organisational theories play an important role in doing so: the resource- and capability-based theories of the firm. Tracing this path allows the underlying mechanisms of the discrepancy to be pinpointed, which in turn allows the formulation of more focussed research questions. For this reason, a more definitive research formulation, i.e. the research nature, scope, and questions, will be constructed based on the findings of the literature review.

1.4 Thesis Outline

This chapter briefly introduced the concept of data as a corporate resource and why this topic has been gaining popularity in the past years. It then presented the research problem, which states that there seems to be no direct correlation between investments in becoming more data-driven and the subsequent pay-offs for organisations. Finally, it shortly touched on the research objectives and approach of this thesis.

The remainder of this thesis is structured along the research diagram depicted on the next page (figure 1). It can be seen that the research is divided into three main phases: research formulation, data collection and analysis, and integration and finalisation.

1. *Research formulation:*
   The research formulation includes this chapter, and then builds on it throughout the following two chapters. Chapter two provides an extensive literature review in which the root-cause of the discrepancy problem will be examined. This chapter will also result in a more specific and definitive research objective. These outcomes will serve as input for chapter three, which will formulate the actual research questions employed in this research. This chapter will also briefly elaborate on the research scope, nature, and methods.

2. *Data collection and analysis:*
   The sub-questions presented in chapter three were designed to build on each other, i.e. with answering each question, the answer to the main research question becomes more substantial. It was therefore chosen to structure chapters four through seven around these research questions. Each of these chapters addresses a sub-question and consists of the following sections: methods, results, and (sub-)conclusions. In this phase, a parallel with on the one hand chapter four and on the other hand chapters five through seven can be observed. This was introduced as the latter three chapters all focus on capabilities, whereas chapter four is concerned with placing these capabilities in the context of organisational changes that comes with becoming more data-driven.

3. *Integration and finalisation:*
   The final phase of this research consists of two chapters. Chapter integrates all previous findings and presents the main managerial implication by answering this thesis' main research question. The final chapter, chapter nine, will reflect on the methods and findings of this thesis. It will conclude with recommendations for future research.

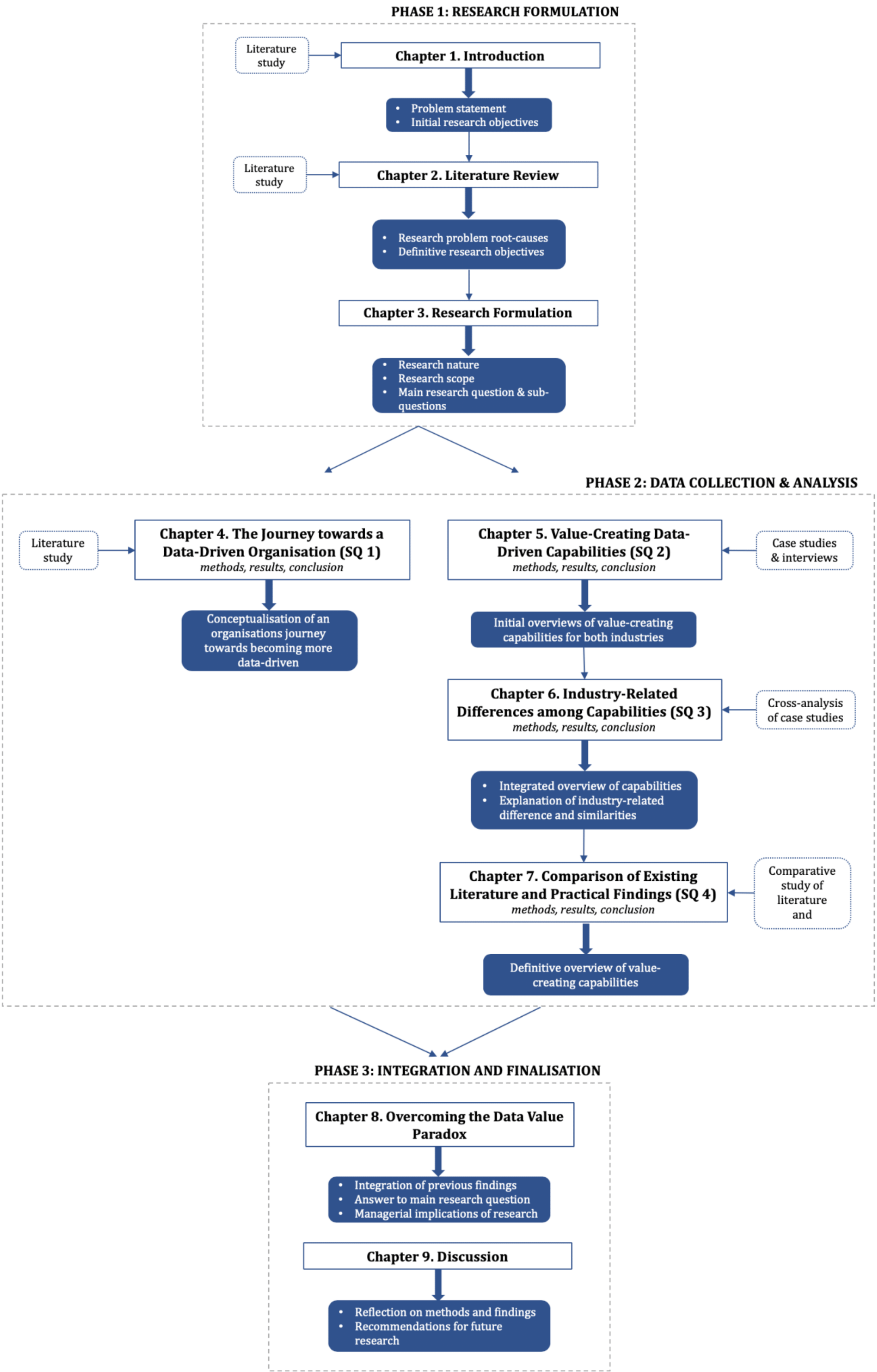

Figure 1. Research flow diagram.

2. Literature Review

To place this research in a meaningful context, a literature review was conducted. The aim of this review was to trace the path from data investments to value creation in order to understand the impediments in this process. This was done by taking on a historical perspective and extrapolating these findings to the present-day problems encountered in becoming more data-driven.

This historical perspective forms the main body of this literature review and is presented in the first section of this chapter by means of an integration of older information technology (IT) publications. Although publications about the incorporation of IT into organisations may seem outdated at present, a lot of parallels can be drawn between these processes and the current obstacles encountered in effectively integrating data into business processes. After all, IT systems form the basis of current data creation, collection, and utilisation systems. The common focus of these publications was addressing a set of problems collectively termed the "IT productivity paradox", a phenomenon which shares many similarities with the problem statement of this thesis. As the IT productivity paradox was successfully resolved in the end, an in-depth analysis of this concept provides valuable insight into the root-causes underlying the discrepancy between data investments and value extracted.

The second section of this chapter will then steer the literature review back to the present, by linking the historical findings back to data. Through the incorporation of modern-day literature and frameworks, it will be shown how the findings from the previous section remain relevant today. This section simultaneously works towards a the knowledge gap that will be addressed in this thesis, which is presented in the third and final section of this chapter.

2.1 Historical Perspective: IT Productivity Paradox

Based on what was discussed up till now, the incorporation of data into business processes almost seems to be a prerequisite if companies want to remain competitive in today's dynamic environment. However, as mentioned in the introduction, companies are struggling to capture data's business value, despite considerable investments. Thirty years ago, the same was happening for IT systems. In 1987, Roach first introduced the so-called "IT productivity paradox" and from there on, it occupied the mind of many scholars. The pace of innovation was higher than ever, and a lot of companies were investing heavily in their IT infrastructure. The pay-offs, however, were rather disappointing; contrary to expectations, there seemed to be little connection between firms' IT investments and their performances (Roach, 1987). Eventually, however, this discrepancy was resolved; today no one will deny the undisputable link between IT and productivity. Since a similar "value paradox" seems to present itself for data now, the path from investment to value will be traced in this section. Two organisational theories, the resource-based theory and its extension into the capability-based theory, lie at the basis of this paradox. This section will be structured along these theories: a resource-based view is introduced first, followed by a more in-depth analysis built on the capability-based theory.

2.1.1 A Resource-Based View of the Productivity Paradox

Throughout the 1990s, a lot of publications were devoted to solving the productivity paradox. In 1998, Brynjolfsson & Hitt reflected on this body of literature, allowing them to draw an important conclusion. The authors found that early research looked for a direct link between IT investments and productivity. As studies found little evidence for such a link, it became common thought that it did not exist. Later on in the decade, however, the realisation started growing that IT doesn't automatically lead to increased productivity. Instead, the underlying mechanisms linking IT to increased performance are more intricate and complex; IT contributes to a wider system of organisational changes that eventually enhance productivity.

The fact that there is no straightforward link between IT investments and increased firm performance, can be explained by means of Barney's resource-based theory (RBT) of the firm. Being one of the leading theories in studying organisations, the RBT allows a wide range of resource types to be linked to firm strategy and performance (Barney, 1991; Wade & Hulland, 2004). In the context of this thesis, the RBT can be used to determine what data-related resources are needed, and how they would have to be leveraged in order to obtain a competitive advantage and ultimately increased firm performance (Gupta & George, 2016).

Leading up to the RBT, two assumptions concerning a firm's sources for competitive advantage are made: resource heterogeneity and resource immobility. Respectively, these terms argue that different firms can possess different resources, and that these resources can be confined to a single firm. Building on these assumptions, Barney's theory states that a firm's sustained competitive advantage can be attributed to the unique combination of resources at its disposal. To supplement his theory, Barney created the VRIO framework. It refers to the attributes a resource must have in order to provide a firm with a sustained competitive advantage, i.e. valuable, rare, imperfectly imitable, and organised. In other words, resources must exploit opportunities and eliminate threats (valuable), can only be shared by a limited number of others (rare), must be hard to imitate (imperfectly imitable), and must be leveraged in such a way that allows a firm to fully capitalise on them (organised). As can be inferred from table 2, a resource only provides sustained competitive advantage when all requirements are met. (Barney, 1991)

Table 2. Competitive implications from the VRIO framework. Adapted from Barney & Hesterly (2008).

Valuable	Rare	Inimitable	Organised	Competitive implications
No				Competitive disadvantage
Yes	No			Competitive parity
Yes	Yes	No		Temporary competitive advantage
Yes	Yes	Yes	No	Unused competitive advantage
Yes	Yes	Yes	Yes	Sustained competitive advantage

It is important to carefully consider what exactly resources entail. Wade and Hulland (2004, p. 109) define resources as "assets and capabilities that are available and useful in detecting and responding to market opportunities or threats." This thesis will also adopt the distinction between assets and capabilities. Assets are further defined as basic units (tangible or intangible) that can be used by firms to implement their envisioned strategies (Barney & Arikan, 2001; Bharadwaj, 2000). Capabilities, on the other hand, are defined as competences that are developed within firms over time and that allow the creation of additional value (Bharadwaj, 2000; Wade & Hulland, 2004). As Makadok (1999, p. 398) formally puts it: "[capabilities are] an organisationally embedded, non-transferable firm-specific resource whose purpose is to improve the productivity of the other resources possessed by the firm."

2.1.2 A Capability-Based View of the Productivity Paradox

The diverging definitions of assets and resources explain why a distinction between these terms is important. Existing literature agrees that it is not simply obtaining a certain asset that gives a firm a competitive advantage. In fact, assets on their own are unlikely to provide an edge (Bharadwaj, 2000; Mikalef, Pappas, Krogstie, & Giannakos, 2017; Wade & Hulland, 2004). Assets can often be obtained easily by competing firms, and as the RBT shows, this will lead to a temporary competitive advantage at most. What truly provides a firm with a sustained competitive advantage, is inimitability and correct organisation of these assets into capabilities. Capabilities are built over time and through experience, and involve complex interactions between the assets available to a firm and their competences (Grant, 1996). They are thus hard to imitate, and with correct organisation they can provide a firm with a sustained competitive advantage. In the context of the IT productivity paradox, the competitive edge must thus stem from the unique ways in which different firms leverage their IT investments (Bharadwaj, 2000). Teece et al. (1997) add that capabilities stem from the assets a firm possesses. In other words, the capabilities a firm is able to develop are dependent on the assets it possesses. Along this line of reasoning, the RBT is thus extended to the so-called capability-based theory or CBT (Wang, 2014). As the term "resources" in the RBT may falsely lead to the assumption that assets alone may lead to a competitive advantage, the CBT is an important construct to include in this research.

It is essential to realise, however, that not all capabilities will lead to a sustained competitive advantage. Bhatt and Grover (2005) claim that capabilities should be divided into those that bring value and those that bring competitive advantage. The authors identify three types of capabilities: value capabilities, dynamic capabilities, and competitive capabilities. This thesis will focus on the latter two. Value capabilities refer to those capabilities that allow a firm to create value, but are at the same time at the disposition of competitors. Although they are a prerequisite for gaining competitive advantage, they do not suffice on their own. Dynamic capabilities relate to the fast-paced competitive environment of a firm. These capabilities allow a firm to respond quickly to both threats and opportunities. Dynamic capabilities have been extensively discussed in earlier literature and are commonly defined as "a firm's ability to integrate, build, and reconfigure internal and external competences to address rapidly changing environments" (Teece et al., 1997, p. 516). Finally, competitive capabilities predominantly involve strategic decisions and actions, hence aligning specific aspects of the business with the overall business strategy. (Bhatt & Grover, 2005; Fatemeh Ahmadi Zeleti & Ojo, 2017)

Zeleti (2018) elaborated on how the different capability types relate to both each other and to the competitive advantage of the firm. This is depicted in figure 2. A network of strong interrelations can be observed, which together form a value chain eventually leading to competitive advantage. The model can be understood by establishing a link with the VRIO framework presented earlier. At the very base of a successful firm lie assets and the value capabilities that stem from these assets. An example could be a firm's IT infrastructure and its deployment across the organisation. However, IT infrastructure is readily available (i.e. not rare) and therefore it will not differentiate a firm from its competitors. Value capabilities contribute to dynamic capabilities which, in turn, can guide an organisation in leveraging their resources in a valuable way. Dynamic capabilities are developed over time through experience and are therefore unique and hard to imitate, thus bringing a firm closer to competitive advantage. An example is an organisation's continuous improvement through organisational learning. The value chain ends with competitive capabilities, which align other capabilities with the overall business strategy. For example, arranging separate IT capabilities in such a way that strategic business needs are met. It relates to the organisational aspect of the VRIO framework and is thus the key to competitive advantage.

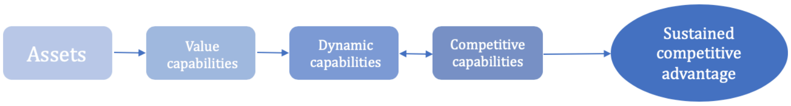

Figure 2. Value chain of capabilities. Adapted from Zeleti (2018, p. 19).

For IT, unique capabilities are achieved through the interaction of three assets: an appropriate IT infrastructure, competent IT staff, and a good relationship between IT and business management (Ross, Beath, & Goodhue, 1996). Bharadwaj (2000) later added to these findings by expanding the relationship asset to IT-enabled intangibles, a term which encompasses several extra qualities that should not be overlooked. IT infrastructure relates to physical resources which are on their own unlikely to create a competitive advantage. However, their integration into more complex systems which complement business needs in a cost-effective manner can create valuable synergy. To achieve this, a technology architecture with appropriate standards should be defined. Furthermore, human IT skills are needed to solve business problems and exploit opportunities. These skills consist of both technical and managerial skills, which usually evolve over time through training and experience, making them hard to imitate. Whereas technical skills refer to know-how about the technology used to gain insights from data, managerial skills are necessary to strategically employ these insights. According to Bharadwaj (2000), especially the managerial ability to coordinate all aspects of the IT structure is a key distinguisher in successful firms. This is not a simple task, however. As recognised by Reich and Benbasat (2000), achieving good alignment between IT and organisational objectives does not always come naturally. Taylor-Cummings (1998) attributes these difficulties to a culture gap that exists between people rooted in IT and business. These different backgrounds can lead to a lack of understanding of each other's needs. This can, in turn, lead to friction in accommodating these needs, for example management being unwilling to make the investments deemed necessary by IT. In the end management has an executive role whereas IT operates on a performing level. The final element suggested by Ross' et al. (1996), relationships within a firm, can contribute to successful multidisciplinary collaboration to overcome this culture gap. These are established and strengthened through frequent communication and the trust that is subsequently built up. These relationships, in turn lead to a much wider spectrum of assets, termed intangibles (Bharadwaj, 2000). Intangibles are tacit and their development is thus highly dependent on communication between staff. They stem from accumulated experience and entail amongst others organisational learning, know-how, corporate culture, etc. It is important to note that all these assets are strongly interrelated and mutually reinforcing.

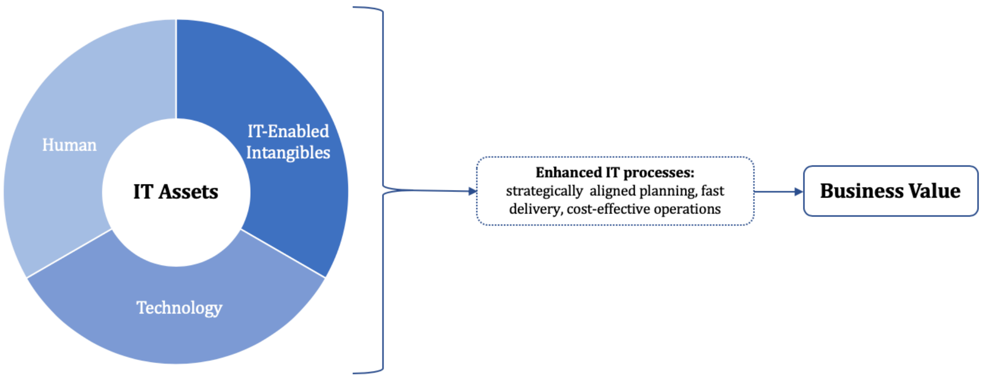

Figure 3. Delivering business value from IT assets. Adapted from Ross et al. (1996, p. 37).

How IT capabilities lead to business value, is summarised in figure 3 (Ross et al., 1996). When IT is aligned with the right assets, capabilities result. As previously described, these capabilities are firm-specific and therefore form a source of competitive advantage. These capabilities will then interact with and subsequently improve established business processes (e.g. making operations more cost-effective), leading to increased business value, i.e. increased firm performance. (Melville, Kraemer, & Gurbaxani, 2004; Ross et al., 1996)

2.2 Back to the Present: Data Value Paradox

Davenport, Harris, De Long, & Jacobson (2001) were among the first to address the path from data to knowledge to results. The authors recognised that companies started to gain more and more access to transactional data, but that it was rarely transformed into valuable knowledge. Just like the IT literature discussed earlier, Davenport et al. (2001, p.117) agree that, in order to address this problem, "companies have to develop capabilities to aggregate, analyse, and use data to make informed decisions". The authors also recognised that a strong emphasis is placed on the technological aspects of integrating data, while managerial aspects are largely neglected, thus preventing these capabilities from being developed. The importance of contextual factors is stressed in the holistic framework in figure 4, which shows three levels: context, transformation, and outcomes. Respectively, these build on each other and it is thus clear that one must be in order before the next can follow. In other words, the organisational foundation of a company cannot be overlooked and must be in place before a company can fully benefit from being data-driven.

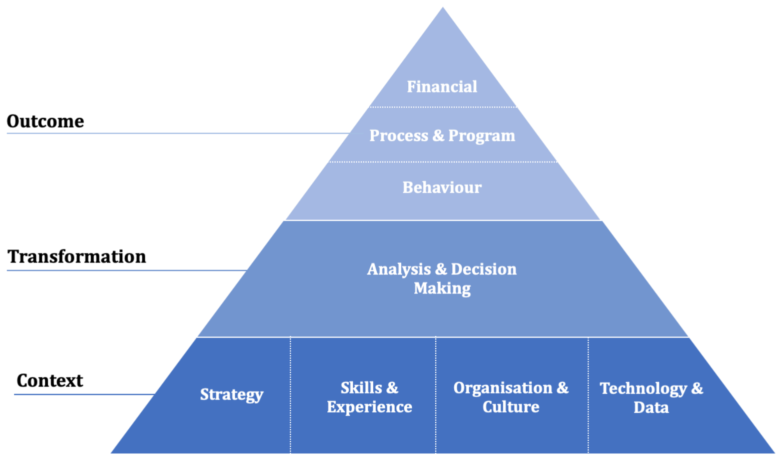

Figure 4. Early framework for building data-driven capabilities. Adapted from Davenport et al. (2001, p. 121).

The claims by Davenport et al. (2001) are still relevant today; companies continue to struggle with capturing value from data. What makes these problems even more compelling, is that more data of different sorts than ever is available to more and more competitors. This makes embedding data in business processes a must for staying ahead of competitors, while at the same time making it more difficult to do so in a distinguishing manner (Kiron, Prentice, & Ferguson, 2014). By now it has become clear that the right organisation of assets can leverage them in a way that leads to competitive advantage. In other words, the way in which data is managed can be the key differentiator relative to competitors. A deeper look into data management, however, reveals that this is a complicated matter consisting of many aspects. Data management is a collective term that entails all processes involved in the data lifecycle, from creation or acquisition to disposal (Henderson, Earley, & Data Administration Management Association, 2017). Figure 5 depicts the DAMA Data Management Body of Knowledge (DAMA-DMBOK) framework, which provides a detailed overview of these processes. At the heart of this framework lies data management surrounded by ten other management domains. A detailed description of these domains is given in appendix A. The multiple dimensions in this framework illustrate the complexity that surrounds data management. Such complexity clearly contributes to the aforementioned organisation of data assets, thus hindering the creation of value from data. This, in turn, makes it hard for organisations to ultimately extract financial benefits from their investments intro data initiatives.

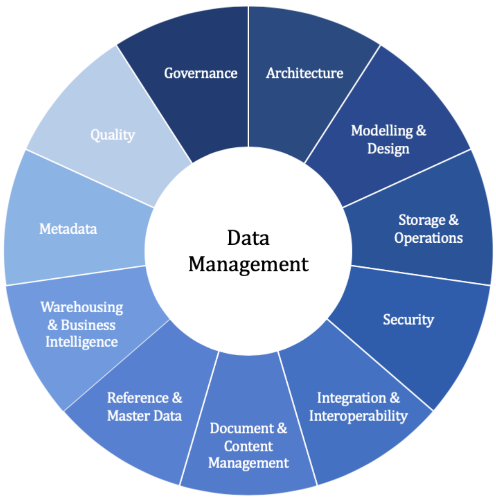

Figure 5. DAMA-DMBOK. Adapted from Henderson et al. (2017, p. 67).

Despite the scientific community's growing enthusiasm for data, research about its financial impacts remains in an embryonic state (Gupta & George, 2016). Mikalef et al. (2017) state that a strong emphasis remains on the technological aspects behind data, while organisational aspects necessary to turn data into actionable insights are largely ignored. The lessons learned from the IT productivity paradox, however, show that both technology and organisation should come together to successfully leverage data's value. Modern literature also recognises the importance of being aware of the nuances that can create value from data (Constantiou & Kallinikos, 2015; Mikalef et al., 2017; Schryen, 2013). It is thus clear that, for data too, firms must build up the right capabilities. For data, these capabilities are defined as a firm's ability to deploy data to gain unique insights leading to competitive advantage (Akter, Wamba, Gunasekaran, Dubey, & Childe, 2016; Gupta & George, 2016; Kiron et al., 2014). As stated by the CBT, capabilities are built up by leveraging a unique blend of company assets. In figure 6, Gupta and George (2016) narrowed these assets down to seven prerequisites, which companies should then combine in such a way that it catalyses and improves current business processes. The author's classification of these assets into tangible, human, and intangible clearly stems from the IT assets defined Bharadwaj, the only difference being the broadening of "IT infrastructure" into "tangibles".

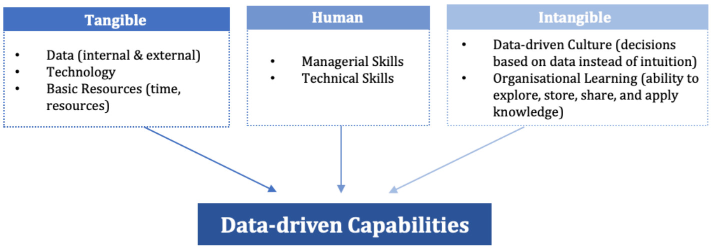

Figure 6. Assets required for data-driven capabilities. Adapted from Gupta & George (2016, p. 1051).

The assets presented in figure 6 are argued to form the foundation of the capabilities necessary to successfully leverage data. It is still necessary, however, to explore towards which capabilities these assets should be steered (Mikalef & Pateli, 2016). Akter et al. (2016) made a first attempt in doing so and identified three primary building blocks: management, talent, and technology capabilities. They further divided these into sub-capabilities, which are briefly elaborated on in table 3.

Table 3. Framework elaborating on the building blocks of data capabilities. Adapted from Akter et al. (2016).

	Primary building blocks	**Sub-capabilities**	**Explanation**
Data-driven capabilities	Management capabilities, i.e. linking data to firm strategy and logistics.	Planning	Ability to identify business opportunities and determine how data can contribute.
		Investment	Ability to generate greater revenues by strategically investing in data initiatives.
		Coordination	Ability to coordinate cross-functional activities.
		Controlling	Ability to strategically assign resources to data initiatives.
	Technological capabilities, i.e. control over IT infrastructure and data.	Connectivity	Ability to connect different business departments.
		Compatibility	Ability to streamline data in order to facilitate processes such as clean-ups and merging.
		Modularity	Ability to quickly change IT infrastructure in response to changes in the market.
	Talent capabilities, i.e. ability of employees to execute certain tasks.	Technology management	Ability to align technology and business goals.
		Technical	Ability to support IT infrastructure and data tooling.
		Business	Ability to understand both the internal and external environment and how to react to changes.
		Relational	Ability to communicate with inter-disciplinary environments.

2.3 Key Takeaways and Knowledge Gap

In the introduction of this thesis it was shown that, these days, data is a critical asset for firms. Integrating data into the day-to-day business enables data-driven decision-making, which can improve firm performance by several means. It was then stated, however, that firms experience difficulty in actually extracting value from their data investments. This problem was explored in more detail in this literature review. A parallel was established with a similar phenomenon in old IT literature, termed the "IT productivity paradox". This led to the adoption of the term "data value paradox" for the disparity between data investments and the benefits reaped from them.

The review resulted in a two-fold root-cause for the value paradox. First and foremost, it was established that there is an important distinction to be made between on the one hand assets and on the other hand capabilities. Investments directly lead to the acquisition of certain assets, but it is the leveraging of these assets into unique, organisation-specific capabilities that gives the firm its competitive advantage. The key to fruitfully becoming more data-driven is thus developing data-driven capabilities. Secondly, it became clear that these capabilities stem from a diverse blend of assets, consisting of both technological and organisational elements. Literature agrees, however, that there is a strong technology-focus, causing the organisational elements to be overlooked.

At present, little research into these organisational data-driven capabilities exist. The literature identified two frameworks that provide a general direction for future research to build on. However, a knowledge gap remains: no explicit overview of organisational data-driven capabilities to be developed exists, while organisational change is a complex endeavour for which managerial guidance is desirable. Earlier, the aim of this thesis was defined as determining how organisations can overcome the data value paradox. The theoretical foundation established in this literature review combined with the knowledge gap allows this aim to be shaped more specifically. This will be done in the next chapter.

3. Research Formulation

This section transforms the findings from both the introduction and literature review into tangible research. The first section elaborates on the general nature of the research as well as the scope. Next, the research questions used to address the knowledge gap are defined. The third section shortly touches on the methods that will be used to address these research questions. Finally, the relevance of this research in the context of the master programme Management of Technology is stated.

3.1 Research Nature and Scope

Based on the literature review conducted in the previous chapter, the aim of this research can be narrowed down to determining which capabilities a firm needs to develop to successfully overcome the value paradox. Existing literature in this research domain has already established certain key elements and directions in the pursuit of a data-driven business model. By building on these findings, this thesis aims to explore and explain these constructs in a more detailed manner in order to support company management with a guiding framework for improving their businesses. This implies that the research performed here will be of a predominantly exploratory nature. New insights will be gained by carefully collecting and analysing existing data, making the approach of this research inductive. With these characteristics and the aim of this thesis in mind, a qualitative research approach was adopted. Existing literature has stressed that, in order to answer the research question presented here, it is essential to look at the process of incorporating data into the business in a holistic manner. For this reason, a qualitative approach is suitable as it provides a lot of flexibility, allowing the incorporation of more complexity and nuance, while still providing insightful generalities (Mason, 2002).

The scope of this research will be confined to the organisational factors necessary for companies to become data-driven. These factors will be considered from both a business and a data perspective, as these go hand in hand when implementing organisational changes. Furthermore, it is important to create an understanding between organisational and technological aspects of the business in order to successfully implement data initiatives. In the literature review, this context was presented through the framework created by Davenport et al. (2001).

3.2 Research Questions

The literature review conducted in the previous section has led to a gap between, on the one hand, what is known theoretically about employing data as a corporate resource, and on the other hand, about how a data-driven structure should be implemented practically. It has become clear that, in order to successfully extract value from data, firms should develop certain capabilities. The question remains, however, what these capabilities are and when they should be implemented. This thesis will therefore address the following research question:

Which capabilities should organisations develop to facilitate overcoming the data value paradox?

The main research question will be answered by means of several sub-questions. These questions were specifically tailored to both enrich and validate findings with each step. First and foremost, however, it must become clear how an organisation progresses towards a data-mature state. Conceptualising this journey will provide insight into which capabilities have to be developed at what moment in time. To do so, the following sub-question is used:

1. *How does an organisation progress towards a more data-mature state?*

Secondly, the capabilities to be developed need to be identified. While data-specific capabilities are leading, it is crucial to also focus on overarching business capabilities. The latter are often forgotten, despite being the enablers of larger changes on enterprise level. As discussed before, both factors contribute to a contextual foundation that must be in place for companies to successfully employ data. Furthermore, the culture gap between business and IT stakeholders remains relevant to this day and should therefore be taken into account. These factors lead to the second sub-question:

2. *From both an organisational and a data-specific perspective, which capabilities do companies develop in practice to extract value from data?*

A nuance that should be examined, are possible differences between industries. The development of diverging capabilities between industries should be taken into consideration, as it will influence the applicability of the findings resulting from this thesis. The decision was made to compare material from two industries: manufacturing and financial services. The choice for these specific industries is twofold. First, they can be placed in two significantly different overarching sectors; goods and services respectively. Based on Grönroos (2000), these differences can be roughly summarised as follows. The goods industry produces tangible products, the ownership of which is transferred when purchased. Services, on the other hand, are intangible and do not involve transfer of ownership. Furthermore, the value of goods is created at production, whereas the value of services stems from interactions. These characteristics also imply differences in both the data organisations employ and the way in which this is managed, emphasizing why a comparison is relevant. Secondly, a study by the McKinsey Global Institute claims that some industries can obtain larger gains from using data. Among the goods and services sectors, manufacturing and financial services show the most potential for capturing value from data (Manyika et al., 2011). With the choice for the industries elaborated, the third sub-question is presented:

3. *Do the capabilities developed in the manufacturing and financial services industries differ?*

Finally, findings from both existing literature and practice will be compared. Such a comparison will further validate this research' findings. Furthermore, it will enrich the knowledge derived from existing literature, and may even reveal (additional) discrepancies between what is thought to be best practice and what are actually best practices in industry examples. This leads to the final sub-question:

4. *How do capabilities derived from existing literature contribute to the practical findings of this thesis?*

3.3 Methods

The methods used for answering the research questions formulated above are summarised in table 4. They will be addressed in more detail throughout the following sections, which are structured according to the different research questions. In these sections, a detailed description of a question's accompanying methods is given. Furthermore, the reasons for choosing them, as well as their specific requirements and possible limitations are elaborated upon.

Table 4. Methods for answering the research questions.

Research question	Methods
How does an organisation progress towards a more data-mature state?	• Literature review • Combination of stage and maturity model
From both an organisational and a data-specific perspective, which capabilities do companies develop in practice to extract value from data?	• Multiple case studies • Interviews
Do the capabilities developed in the manufacturing and financial services industries differ?	• Cross-analysis of case studies
How do findings from existing literature compare to practical findings?	• Comparison of literature and case studies
Which capabilities should organisations develop to facilitate overcoming the data value paradox?	• Integration of previous findings

3.4 Link to Master Programme

Analysing the path of value creation from technology investments is a particularly suitable research topic for the master Management of Technology (MoT) at Delft University of Technology. The program aims to train students to understand how technology can function as a corporate resource, so the link with this thesis is self-explanatory. The integration and leveraging of data into current business practices in order to improve processes and results is currently a very relevant topic that fits well into modern business cases. The field, however, is still young and requires more theoretically backed research (Wamba et al., 2015). The analysis envisioned for this thesis will be performed by building on several MoT-related concepts and theories, such as the resource-based theory of the firm, capabilities, and stage/maturity models.

Furthermore, from what has been discussed so far, it has become clear that, these days, becoming data-driven is more of a prerequisite than an extra, both for established and younger companies. As has become clear throughout this thesis, managing data with the goal of extracting value is very complex. This complexity can be attributed to both the grand and diverging nature of data, and to the many stakeholders involved. Addressing such a problem demonstrates the need for managers with more technological knowledge, which is the main rationale of the MoT programme.

4. The Journey towards a Data-Driven Organisation

This chapter will answer the first sub-question of this thesis, i.e. *"How does an organisation progress towards a more data-mature state?"*. The focus will thus be on conceptualising an organisation's journey towards becoming more data-driven. Becoming more data-driven is a complex process that requires significant organisational changes. Sketching a context for this process and breaking it up into distinct stages will facilitate the comparison of different organisations and will later allow capabilities to be assigned to certain stages along this progression. Conceptualising the progression towards a more data-mature state will therefore pave the way for the further research of thesis. This chapter will present the methods for doing so, as well as the subsequent findings. It will be concluded by summarising the main findings.

4.1 Methods and Limitations

The research question addressed in this section serves to position the research in a broader context. This topic has been researched before, and since there is no use in re-inventing the wheel, a literature study will be performed to answer this question. Information will be gathered from two main sources. On the one hand, online databases such as Scopus and the TU Delft repository were consulted. Search queries used are for example data maturity, maturity assessment, data management, information management, et cetera. On the other hand, internal KPMG documents were consulted. These include data frameworks created by KPMG, as well as the more elaborate frameworks on which these were built.

The outcome of this literature review will then be integrated into a combined stage and maturity model in order to provide a clear overview of the steps a company needs to make in becoming fully data-driven. Nolan (1979) argues that stage models can be used to help organisations in transitioning from one stage to the next, which is a feature that fits exceptionally well with the aim of this thesis. Klievink and Janssen (2009, p. 276), further elaborate on this model: "Identification of stages needs to be based on the concept of discontinuity, while dynamic capabilities theory can be used to provide guidance for the transition from one stage to another." This rationale again goes hand in hand with the aim of this thesis, as it integrates the maturity stages in which a company can find itself and the capabilities necessary to improve on this scale.

Even though using extant literature is an effective way to gather large amounts of data, it cannot be guaranteed that the study conducted here provides a complete integration of the available literature. Due to the sheer amount of literature available, publications which have been frequently cited are more likely to be incorporated than others. This way of filtering may have caused relevant papers to be left out, thus posing as a limitation in this research.

4.2 Results

The journey towards a data-mature organisation is not an overnight process. It evolves over time and involves intricate social interactions and organisational learning. The process becomes even harder to grasp when one takes into account the fact that, these days, data extends throughout the entire organisation, thus growing both in scope and complexity. Zachman (1987) was early to note that well-defined constructs are necessary for maintaining understanding and control over such complex systems. In order to make the progression towards a data-driven organisation less abstract, a combination of stage and maturity models will be used here.

Stage models were first introduced by Nolan (1979). Such models break down an organisation's journey into different sequential stages in which they progressively become more mature in a certain area; specifically data for this thesis. An important concept within stage models is discontinuity, as depicted in figure 7 (Klievink & Janssen, 2009). As can be seen, within stages an organisation improves incrementally. For transitioning to the next stage, however, more significant changes have to be made. These transitions are thus characterised by discontinuity, as organisations will have to develop new capabilities to realise the next stage. The authors furthermore stress the importance of keeping the development of these capabilities dynamic, stating that "the evolution of capabilities is influenced by the pacing of experience (Klievink & Janssen, 2009, p. 277)". In other words, capabilities must be developed at the right time. If they are developed too early, a company's lack of experience may cause them to not be able to leverage these capabilities properly. If they are developed too late, however, the organisation will lack the knowledge necessary for successfully transitioning to the next stage. Stage models can make this transition go over more smoothly, as they provide insight into which capabilities must be developed over time. This makes this type of model particularly suitable for the research conducted in this thesis.

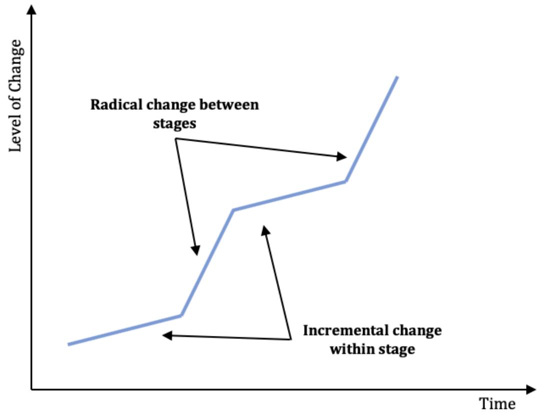

Figure 7. Transitioning between stages. (Klievink & Janssen, 2009, p. 277).

From the above, it is clear that companies have to develop different capabilities depending on their level of data maturity. Therefore, in order to improve and progress to a more data mature organisation, companies must be able to pinpoint where they are currently positioned. This, in turn, requires more insight into the stages of incremental improvement, for which maturity models are useful instruments (De Bruin, Freeze, Kaulkarni, & Rosemann, 2005). Similarly to stage models, maturity models consist of different sequential levels, each accompanied by a set of characteristics (Becker, Knackstedt, & Pöppelbuß, 2009). If a company displays these characteristics, it can be said to have achieved the corresponding level. Moving from one stage to the next again requires the development of certain capabilities.

Maturity models that describe a company's progression in becoming more mature with respect to data specifically, will be termed data maturity models from now on. From existing literature, multiple data maturity models are available. Three diverse frameworks (DataFlux, COBIT, and CMMI DMM) will be discussed here to eventually come to an integrated pathway to data maturity.

The first framework presented, was created by DataFlux in 2007. The model aims to provide a roadmap for improving data quality, and takes on a holistic approach to do so. It therefore considers the different levels of data maturity along three major axes: people, policies, and technology. The resulting four levels (undisciplined, reactive, proactive, and governed) are presented in table 5. (DataFlux Corporation, n.d.)

Table 5. DataFlux maturity model. Adapted from DataFlux Corporation (n.d.).

	Level	Explanation
1	Undisciplined	*"Think locally, act locally"* – Little executive insight in the value of data and data management practices. No defined rules and polices regarding data management, no standardised roles and responsibilities. A lot of redundant and outdated data is present, no procedures defined to correct this. Results in poor data quality, risking lost opportunities and incorrect decisions.
2	Reactive	*"Think globally, act locally"* – Standardised procedures, rules, and roles and responsibilities are established in individual departments. However, still little executive insight in the value of data, so no organisation-wide standards. Emphasis remains on fixing problems after they occur.
3	Proactive	*"Think globally, act collectively"* – Value of data is understood on an executive level, data culture emerges across departments. Standardised procedures, rules, and roles and responsibilities are established organisation-wide. Data governance is embedded into processes. The level of automation increases, allowing more consistent data quality monitoring. Focus shifts from resolving problems to preventing problems.
4	Governed	*"Think globally, act globally"* – The value of data is understood throughout the entire organisation. A single data strategy is in place. Data processes and data quality are highly automated and embedded in standardised processes throughout the organisation. Data quality is monitored continuously and issues are resolved immediately.

A second framework, is the COBIT® framework. COBIT, an acronym for Control Objectives for Information and related Technologies, was developed in 2012 to optimally leverage IT investments by aligning IT processes with the business. It is argued that, in order to achieve this, a guiding framework needs to be in place. COBIT consists of six maturity levels: non-existent, initial, repeatable, defined process, managed and measurable, and optimised. These levels are explained in table 6. (IT Governance Institute, n.d.)

Table 6. COBIT® maturity model. Adapted from IT Governance Institute (n.d.).

	Level	Explanation
0	Non-existent	No data process in place, no sense of an issue to be addressed.
1	Initial	Realisation of an issue to be addressed, however no standardised or organised procedures to do so. Instead, processes are ad hoc.
2	Repeatable	Different stakeholders executing the same task to so according to similar procedures. However, no formal guidelines or roles and responsibilities have been established.
3	Defined process	Procedures are standardised and documented, and this is communicated to relevant stakeholders. However, no way to check whether executed correctly. Standard processes are formalised existing practices.
4	Managed and measurable	Adherence to standard procedures are monitored. Processes are kept up-to-date and reflect best practices. Automation remains limited.
5	Optimised	Processes have been refined based on continuous improvement and results from the industry. High level of automation, improving quality and effectiveness.

The final maturity model presented, the data management maturity (DMM) model, was developed in 2014 by the Capability Maturity Model Integration (CMMI®) institute. This model traces a path to maturity which integrates the earlier discussed DAMA-DMBOK's most crucial elements: strategy, data governance, data quality, data operations, and data architecture. This has led to a maturity model consisting of five levels: performed, managed, defined, measured, and optimised. These levels are described in more detail in table 7 below. (CMMI Institute, 2014)

Table 7. CMMI® DMM model. Adapted from CMMI Institute (2014).

	Level	Explanation
1	Performed	Data is seen as a requirement for other projects. Processes are performed ad hoc at project level. Not applied across different business areas. Focus is on repair rather than prevention.
2	Managed	There is increasing awareness about the importance of data. Skilled stakeholders can produce controlled outcomes. Within different departments, standardisation is starting to develop per process. Adherence to processes is monitored.
3	Defined	Data is managed centrally. Standard processes are documented, employed, and followed organisation-wide.
4	Measured	Data processes are continuously monitored on the basis of pre-defined metrics and KPIs. Monitoring results are used to improve processes periodically.
5	Optimised	In addition to level 4, analyses are used to pro-actively steer the effects on the overall strategy. Results and best practices are shared with peers and industry.

A comparison of these models shows that there is clear consensus over a common thread running through the different levels. All models roughly start with ad hoc data processes and then slowly start spreading standardised procedures throughout the entire organisation. There are, however, also clear differences between the three models. To overcome these differences, levels of all models were integrated into a stage model which is best-suited for this research, as depicted in figure 8. The resulting five levels are discussed below.

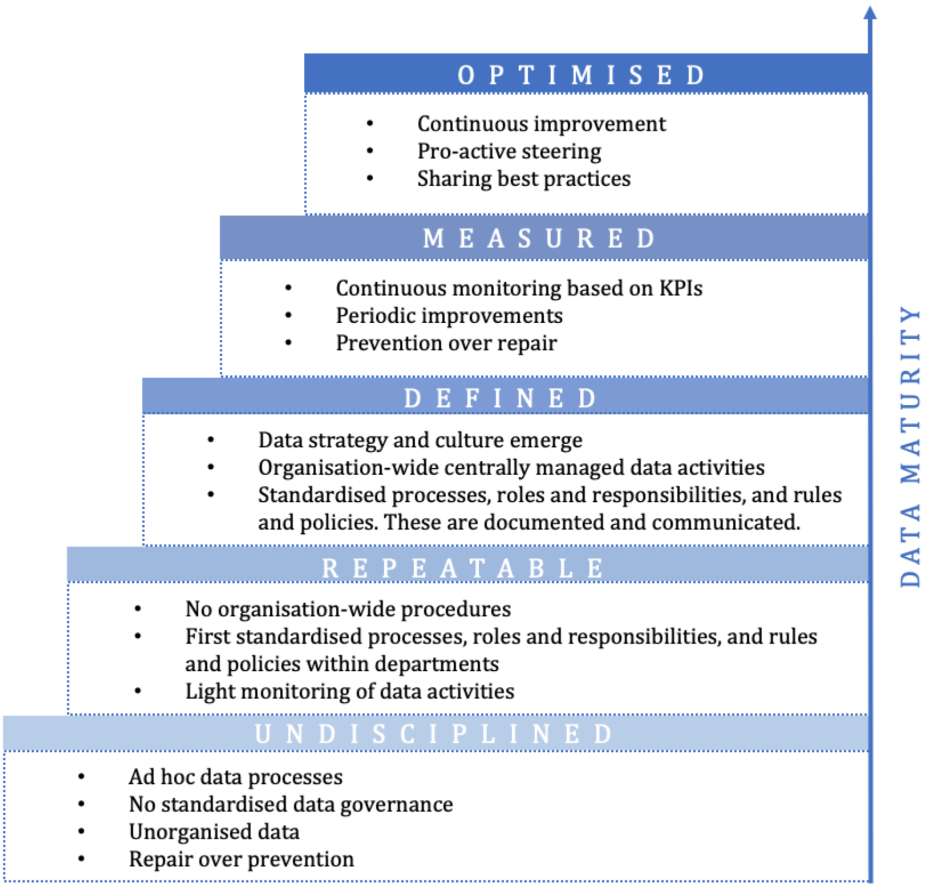

Figure 8. Conceptualised journey towards a more data-driven organisation.

It was chosen to start the stage model at a point where the organisation is aware of data. Therefore, the "non-existent" level from the COBIT model was disregarded. The model starts with level one, termed *undisciplined*. At this level, data is seen as a requirement to execute certain projects, which results in data processes being performed on an ad-hoc basis. There are no standardised procedures, rules and policies, or roles and responsibilities. Data is often unorganised, leading to outdated and redundant data cluttering databases. The focus is in repairing issues rather than preventing them. At the second level, *repeated*, there is still little executive insight, so no organisation-wide procedures are adopted. Within departments, however, stakeholders start executing tasks in similar ways, leading to standardised procedures, rules and policies, and roles and responsibilities within departments. Within the departments, adherence to these practices can be lightly monitored. At the third level, *defined* the value of data has become clear on an executive level, leading to a data strategy and the emergence of a data-driven culture throughout the organisation. This, in turn, leads to organisation-wide centrally-managed data management activities. Existing procedures are standardised, documented and communicated to relevant stakeholders. Furthermore, formal rules and policies, and roles and responsibilities are defined. Central in level four, *measured*, is the process of automation. The level of automation of data processes increases, allowing their continuous measuring and monitoring based on pre-defined metrics and KPIs. These can then be used to periodically improve processes, leading to best practices. Prevention of incidents is now prioritised over repair. In the final level, called *optimised*, continuous improvement and pro-actively steering practices in the desired direction is central. Improvements are based on both internal and external knowledge. Furthermore, best practices are shared with peers and industry.

4.3 Conclusion

This chapter was aimed at answering the sub-question *"How does an organisation progress towards a more data-mature state?".* The integration of stage- and maturity models has shown that this is a complex process which cannot be achieved overnight. Instead, both incremental and radical changes that allow an organisation to be more data-mature build on each other over time. Here, this process was divided into five stages ranging from undisciplined to optimised. Throughout these stages, data management practices become more and more centralised and advanced. The findings of this chapter will be used as a foundation to build this thesis' further analyses on.

5. Value-Creating Data-Driven Capabilities

It has been established that a network of data-driven capabilities needs to be in place to serve as a foundation for more elaborate data initiatives. To determine what these capabilities entail, this chapter will go into the sub-question *"From both an organisational and a data-specific perspective, which capabilities do companies develop in practice to extract value from data?"* It is composed of three sections: methods, results, and conclusions.

5.1 Methods and Limitations

To answer this question, two qualitative research tools were used: multiple case studies and semi-structured interviews. This thesis was written in collaboration with the Enterprise Data Management department of KPMG the Netherlands. Their experience in the field of organising data and their extensive network of clients, has greatly supported data gathering and processing activities. Cases were selected based on a database of historical projects in which KPMG helped companies to specifically improve their data management practices. Interviews, in turn, were conducted with representatives of the respective firms. The required contacts were also acquired through KPMG. Note that the data used to answer this question is confidential and was therefore anonymised.

5.1.1 Case Studies

Case studies are often surrounded by a fair amount of scepticism. This scepticism is rooted in several reasons, the most pressing of which is the lack of rigor in case study research (Hodkinson & Hodkinson, 2001; Siggelkow, 2007; Yin, 2014). Data obtained through cases is often hard to quantify. It may therefore be complicated to justify the validity of conclusions derived. A second point of criticism towards case studies, is the fact that there is little basis on which to generalise conclusions, as there are usually few samples which are not randomly selected (Hodkinson & Hodkinson, 2001; Siggelkow, 2007; Yin, 2014). A final concern is the overwhelming amount of data case studies can generate, which can in turn lead to a loss of focus (Hodkinson & Hodkinson, 2001; Yin, 2014).

These cons, however, are for this research outweighed by the fact that case studies allow mechanisms to be studied within their context without experimental interference from a researcher. This results in rich data which can lead to a deeper understanding of complex interdependencies, thus providing new and interesting insights (Hodkinson & Hodkinson, 2001; Siggelkow, 2007). These nuanced insights are especially important in this research, as its focus is on the appropriate context that has to be created for data initiatives to succeed. New insights may also lead to the discovery of discrepancies, which allows for the improvement of existing theories (Yin, 2014). The latter is very suitable for new areas of research such as the one researched here.

The criticism towards case studies can, to some extent, be overcome by setting up a clear research approach. This was done in accordance with Eisenhardt's 1989 roadmap for building theories from case studies. The following steps were included: case selection, conducting studies and internal analyses, conducting cross-analyses, and drawing conclusions. The steps presented by Eisenhardt as well as a summary of how they were implementation in this thesis are pictured in figure 9. For clarity, the steps are addressed bullet wise below.

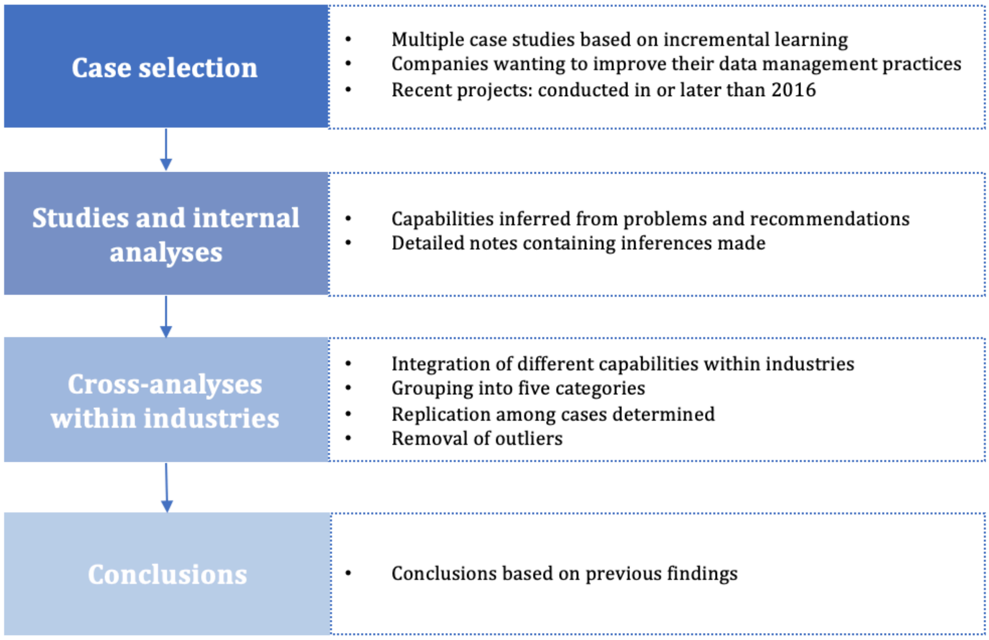

Figure 9. Case study research approach.

1. After determining the research question, cases were selected. It was chosen to perform multiple case studies per industry, as the replication of results contributes to the study's rigor. This is especially important as this question attempts to derive common trends. The number of cases was not set in stone beforehand, but was decided upon by monitoring the level of incremental learning with each case. After four to five cases per industry, incremental learning was minimal. Cases were selected according to several criteria. They had to cover companies wanting to improve their data management practices by giving an assessment of a company's current state, its ambitions, and the recommended ways to reach these ambitions. Moreover, to ensure relevance in an ever-changing technological environment only recent cases (projects performed in or after 2016) were included. Based on these criteria five cases were selected for the manufacturing industry, while four cases were included for financial services. This asymmetry can be attributed to availability of relevant cases. An anonymised description of these companies can be found in table 8. The letter M indicates manufacturing companies, while F denotes companies active in the financial services industry.

Table 8. Anonymised description of studied companies.

Company	Anonymised description
M1	Manufacturer of chemicals
M2	Manufacturer of integrated transport systems
M3	Manufacturer of maritime solutions
M4	Manufacturer of professional cleaning equipment
M5	Manufacturer of renewable packaging solutions
F1	Enterprising bank
F2	Investment fund
F3	Insurance broker
F4	Insurance concern

2. The second step involved carrying out the studies and simultaneously making internal analyses. During this process, detailed notes were kept about the findings. For reasons of confidentiality, these notes were not included in this report. Careful analysis of observed problems and given recommendations resulted in a list of inferred capabilities for each separate case, which will be presented in the following section.

3. After analysing all cases separately, cross-analyses were performed within each industry group. Per industry, all findings were listed and integrated to form a collective overview of all capabilities derived. For this purpose, some capabilities were reformulated and combined to convey more complete and meaningful information.

   Furthermore, capabilities were grouped into one of five categories: business, data foundations, data governance, data quality, and data processes. These categories can be traced back to Davenport's contextual factors summarised in figure 4. Strategy and organisation & culture were integrated into the business category. Capabilities pertaining to skills & experience were summarised as governance. Finally, to make Davenport's data building block more complete it was split into three separate categories: foundation, processes, and quality.

   It was then determined in how many of the treated cases the derived capabilities were present. Elements present in two or less cases were considered outliers and were discussed with a KPMG colleague involved in the project to decide on whether or not to incorporate them. This was decided by considering whether these outliers were due to company-specific singularities or not. If so, they were left out for the sake of usability of the eventual conclusions.

4. In a final step, conclusions were drawn based on the analysis described above.

Throughout the process of conducting case studies, several measures were taken to improve the study's validity. Yin (2014) identifies four quality checks: construct validity, internal validity, external validity, and reliability. The author also presents the tactics to execute these checks, as well as the phase of the research in which they should be performed. In table 9 below, the measures taken to improve and maintain the validity in this research are presented.

1. A first check that was incorporated, is the one for construct validity. This ensures that the methods used are appropriate for the conducted research. In case studies, it can be difficult to operationalise the process in an objective manner. Here, two measures were taken to sufficiently do so. First of all, two sources of information were used: documentation and interviews. This allows triangulation, which makes findings more convincing. Second, a chain of evidence was kept so that external observers can follow the line of reasoning for certain conclusions in case studies.

2. The second check to be incorporated covers internal validity. For exploratory case studies, this mainly entails that the inferences made from and within cases are trustworthy. This was be done by substantiating findings with solid explanations.

3. To address external validity and to thus make sure findings are as generalisable as possible, replication logic will be used. Multiple similar cases were selected in order to find similarities between outcomes, providing rigor.

4. Finally, the reliability of the research should be appropriate to an extent that an external party should be able to reproduce the research. To accomplish this, a detailed research design and case study reports were developed.

Table 9. Validity checks. Adapted from Yin (2014).

Check	Tactic	Phase of research
Construct validity *Fitting method?*	Use multiple sources of evidence	Data collection
	Establish chain of evidence	Data collection
Internal validity *Trustworthy findings?*	Use explanation building	Data collection
	Address rival explanations	Data collection
External validity *Generalisable findings?*	Replication logic	Research design
Reliability *Reproducible research?*	Use case study protocol	Data collection
	Develop case study reports	Data collection

5.1.2 Interviews

To both complement and validate the findings from the case studies, semi-structured interviews were be conducted with representatives of selected firms. The interviews were chosen to be semi-structured, because this gives the interviewees the room to tell their story in more detail, thus providing more context while still maintaining a general direction. Interviews with professionals may lead to extra insights since they may provide certain nuances that are left out in the case studies. As the interviews are meant to provide extra insights on top of the case studies, they were performed after the case studies.

The starting point for selecting interviews were the cases treated previously. Based on time and availability, it was chosen to conduct interviews at two companies per industry, with one individual at each company. The most promising interview candidates were determined in accordance with a KPMG contact of the respective cases. Candidates were then contacted through the same colleague. Three of the four invitations were replied positively; for this reason only one interview could be conducted in the financial services industry. The content of the interviews can roughly be split into two parts. The first section focussed on sketching context to gain deeper insights. Interviewees were asked about the situation and challenges of their organisation before involving an external party, about the implemented organisational changes and the current state of their organisation. On top of this, interviewees were asked about their reflection on the whole process. The second part served to explicitly validate case study findings. To this end, interviewees were presented with a list of capabilities derived from their specific case. For each capability, they were then asked whether or not they deemed these capabilities relevant, and if they would classify them as core, supporting, or specialised. The latter was asked with an eye on the priorities to be assigned later on. An interview protocol with the questions used to guide the interview can be found in appendix B. It should be noted that, as the interviews were semi-structured, these questions were not strictly adhered to.

Limitations for the interviews conducted in this thesis, are tied to time and availability. Interviews were not performed at all companies studied, affecting the generalisability of the obtained results. Furthermore, only a single stakeholder per company was questioned. The interviews will therefore only convey the opinion of a single individual within the firm, which may lead to biased answers. Due to these factors, results should be interpreted with scrutiny. For the purpose the interviews serve in this research, i.e. sketching context and validating results obtained from case studies, the results remain a valuable addition.

5.2 Results

In this section, the results from the case studies and interviews described above are presented. The results from both methods will be integrated in the subsequent concluding section.

5.2.1 Case Studies

As mentioned earlier, case studies were performed on the final reports of ten KPMG data management projects, which contain information such as key data-related problems encountered by a firm, as well as the recommendations made to overcome these. Subsequently, capabilities were derived from these original observations, the final selection of which is presented here for both the manufacturing and the financial services industry.

Table 10 summarises the results of the manufacturing industry case studies. For this industry, seventeen capabilities have been derived, which have been categorised according to the five categories defined earlier. Table 10 also shows the cases in which the capabilities were encountered. As mentioned before, capabilities only present in two cases or less were subjected to additional analysis with KPMG colleagues involved in the projects. In this way, it was determined whether these findings stemmed from company-specific singularities or whether they remained relevant across the industry. Singularities were removed and therefore all capabilities presented here are considered relevant.

Table 10. Capabilities derived from manufacturing case studies.

	Category	Capability	Present in cases
1	Business	The ability to organise the company in a suitable way	M1, M2, M3, M4, M5
2	Business	The ability to promote the business rationale of high quality data to increase awareness and acquire sponsorship for change initiatives	M1, M2, M3, M4, M5
3	Business	The ability to align data initiatives with business objectives	M1, M2, M3, M4
4	Business	The ability to actively involve data in new projects	M1, M3, M4
5	Business	The ability to create a data-friendly company culture	M1, M2, M5
6	Business	The ability to facilitate the cross-pollination of good practices by supporting cross-functional collaboration	M2, M3, M5
7	Business	The ability to ensure effective decision-making for change initiatives	M4
8	Foundation	The ability to develop and maintain a data strategy, to document and communicate it throughout the organisation	M1, M2, M3, M4, M5
9	Foundation	The ability to create a shared data language throughout the organisation by developing and communicating definitions, standards, policies, and rules	M1, M2, M3, M4, M5
10	Foundation	The ability to monitor and subsequently improve data management activities by measuring internalised pre-determined KPIs	M1, M2, M3, M4, M5
11	Foundation	The ability to determine to optimal way in which data should be organised within the business	M2, M4
12	Governance	The ability to appoint the right central data governance bodies	M1, M2, M3, M4, M5
13	Governance	The ability to communicate governance bodies' roles throughout the organisation	M2, M3, M4, M5
14	Governance	The ability to recruit the right talent	M1, M2, M3, M5
15	Processes	The ability to centralise and standardise data processes	M1, M2, M3, M4, M5

16	Processes	The ability to embed controls in data processes	M3, M4
17	Quality	The ability to monitor data quality based on pre-defined metrics and targets, to report on these findings in a standardised manner, and to continuously improve data	M1, M2, M3, M4, M5

The financial services cases studied resulted in twenty capabilities, which are displayed in table 11. Similarly to table 10, capabilities have been grouped in five categories and any outliers have been removed.

Table 11. Capabilities derived from financial services case studies.

	Category	Capability	Present in cases
1	Business	The ability to promote the business rationale of high quality data to increase awareness and acquire sponsorship for change initiatives	F1, F2, F3, F4
2	Business	The ability to align data initiatives with business objectives	F1, F2
3	Business	The ability to create a data-friendly company culture	F1, F2, F3
4	Business	The ability to facilitate the cross-pollination of good practices by supporting cross-functional collaboration	F1, F2
5	Business	The ability to organise the company in a suitable way	F5
6	Business	The ability to define a roadmap with prioritised short- and long-term projects, depending on the organisation's ambitions	F3
7	Foundation	The ability to develop and maintain a data strategy, to document and communicate it throughout the organisation	F1, F2, F3, F4
8	Foundation	The ability to create a shared data language throughout the organisation by developing and communicating definitions, standards, policies, and rules	F1, F2, F4
9	Foundation	The ability to determine to optimal way in which data should be organised within the business	F1, F2
10	Foundation	The ability to centralise data management activities	F1,F2
11	Foundation	The ability to monitor and subsequently improve data management activities by measuring internalised pre-determined KPIs	F1, F2, F3
12	Governance	The ability to appoint the right central data governance bodies	F1, F2, F3, F4
13	Governance	The ability to communicate governance bodies' roles throughout the organisation	F1, F2, F3, F4
14	Governance	The ability to recruit the right talent	F2, F3
15	Processes	The ability to centralise and standardise data processes	F1, F2, F3, F4
16	Processes	The ability to develop and communicate frameworks for guiding data processes	F1, F2, F3
17	Processes	The ability to stablish a central security policy aligned with regulatory requirements	F1, F2, F3
18	Processes	The ability to establish a central data life cycle management process	F1, F2
19	Processes	The ability to establish organisation-wide document management standards, regularly review these	F1, F2
20	Quality	The ability to monitor data quality based on pre-defined metrics and targets, to report on these findings in a standardised manner, and to continuously improve data	F1, F2, F3, F4

5.2.2 Interviews

This section elaborates briefly on the interviews and presents the most relevant results per interview conducted. The contextual half of the interview and the inferences made from this are summarised bullet-wise, whereas the validations are summarised in accompanying tables. Observations made from were included in blue italics. Detailed transcriptions of the interviews can be found in appendix C. Before including any results in this thesis, outcomes were explicitly approved by the interviewees.

COMPANY M1

M1 is a global chemical manufacturer headquartered in the Netherlands. This company's data management practices were deeply decentralised, causing them to not be able to leverage the full potential of data. When the firm switched to a global business service

(GBS) model, data management was included in this. Furthermore, a process house was created in which governance for end-to-end processes was built. Interviewee has two functions: head of the global data management service in GBS, and global process delivery lead. This makes the interviewee responsible for both implementation on a higher level and overseeing the team responsible for implementation. KPMG was involved in designing the GBS for data management. Initially, four areas of attention were determined: governance (not in place), organisation (no central organisation or harmonisation), data quality (lacking), and tooling (sufficient). Tooling was later disregarded and governance was delegated, leaving two focus areas: organisation and data quality. Below, the most noteworthy findings are presented.

Concrete organisational changes implemented	• Organisation: Data management was set up in GBS by transferring existing decentralised practices. Centre of expertise and an operations unit were created. Led to labour arbitrage, so business case. ▪ *Confirms the importance of organising the company in a suitable way, adds the need to centralise data management activities* • A global dashboard was created to give a single view of data quality for the entire business. ▪ *Adds the need for centralised data management activities, confirms the importance of monitoring data quality and speaking the same data language* • Interviewee noted that, in all of this it is important to gain sponsorship. It can be hard to link data management with the value it delivers, as this is a very time and effort intensive project. ▪ *Confirms the importance of gaining sponsorship* • It is better to address problems at the source. At M1 this is done by determining the business' pain points, priorities, and ambitions. It is then determined what pain points are caused by lacking data management, and these are then solved based on the priorities that were established. ▪ *Confirms the importance of aligning data initiatives with business objectives*
Future ambitions	• Extend the scope of current data management; not all data has been centralised. ▪ *Confirms the importance of speaking the same data language, and the importance of having the right talent in house, adds the need for centralised data management activities* • Automating and optimising processes to reduce costs. M1 is able to pursue these new projects because time is freed up for relevant stakeholders and there is more visibility on data quality and performance monitoring. ▪ *Confirms the importance of centralising and standardising data processes, adds that automation is needed for optimisation*
Most important changes in hindsight	• Governance is key, if this had not been set up properly, M1 would not be where it is today. It is essential to have a good foundation in place. ▪ *Confirms the importance of central data governance, stresses that this is the most important capability for M1* • Advantage of governance is that people are forced to take responsibility over data objects, so if something goes wrong you know where to go. ▪ *Stresses the importance of responsibility for data objects*

Difficulties collaborating with business and IT	• M1 tries not to think of data management as IT. It is a complete service of which technology and tooling is only a small part. Many people take this shortcut too easily. ▪ *Confirms the importance of good collaboration between different departments*

Table 12. Validation of capabilities found in M1.

Capability	Relevant?	Priority?
The ability to organise the company in a suitable way	Yes	Core
The ability to promote the business rationale of high quality data to increase awareness and acquire sponsorship for change initiatives	Yes	Supporting
The ability to actively involve data in new projects	Yes	Core
The ability to create a data-friendly company culture	Yes	Supporting
The ability to develop and maintain a data strategy, to document and communicate it throughout the organisation	Yes	Core
The ability to create a shared data language throughout the organisation by developing and communicating definitions, standards, policies, and rules	Yes	Core
The ability to appoint the right central data governance bodies	Yes	Core
The ability to recruit the right talent	Yes	Core
The ability to monitor data quality based on pre-defined metrics and targets, to report on these findings in a standardised manner, and to continuously improve data	Yes	Core
The ability to centralise and standardise data processes	Yes	Core

COMPANY M3

Dutch manufacturer of maritime solutions. In the past, M3 consisted of several smaller companies, which merged over time. Now these once separate companies are being integrated in a uniform matter, i.e. processes are being centralised. This is happening over four axes, one of which is data. Before, data management was primarily contained to creating corporate data objects. There are however, much more elements to data management. The interviewee noted that the reasons for involving KPMG were threefold: professionalising multiple domains of data management, increasing data awareness, and gaining momentum for transforming data practices within the organisation by employing an expensive third party. Interviewee is team lead within M3's data department. Below, the most noteworthy findings are presented.

Concrete organisational changes implemented	• Interviewee stressed that the process is one of incremental growth. Several steps, however, form the foundations on which further data management practiced can be built. • Awareness for data management was created through a service catalogue, which introduces the team and the services they can provide. ▪ *Confirms the importance of creating awareness* • Getting the organisation to speak the same language by developing a data dictionary. This contains e.g. the data objects present in certain processes, the responsible stakeholders, etc. The complicated factor here is not creating it, but translating it to the business, this requires the right business channels. ▪ *Confirms the importance of speaking the same data language* • Interviewee noted that working with processes instead of business units allows one to transcend business units, creating uniformity. ▪ *Confirms the importance of cross-organisational collaboration*
Future ambitions	• Data is a broad concept, so you can keep professionalising in all the different domains. ▪ *Adds that it data management is a broad concept and for optimisation, all domains should be considered*

	• Up till (post-KPMG) now the focus has been on government and data quality, which form the base. For custom goods there is a limited amount of data, making quality very important. For quality to be high, data governance needs to be in order. ▪ *Stresses the importance of high quality data and governance*
Most important changes in hindsight	• Implementing a common language and awareness throughout the organisation (service catalogue and data dictionary). • Implementing data quality checks. ▪ *Confirms the importance of monitoring data quality and embedding controls for this*
Difficulties collaborating with business and IT	• Not really an issue as both groups work in the same office (i.e. close personal contact) and SCRUM teams are used (share visions and come to an agreement). ▪ *Confirms the importance of good communication between stakeholders from different departments* • Interviewee noted that business stakeholders often think projects come off the ground to slowly; they don't know the back-office work it takes to generate solutions. • Overall top management is involved, but it can be difficult to get their attention as data investments are usually small. ▪ *Confirms the importance of gaining sponsorship*

Table 13. Validation of capabilities found in M3.

Capability	Relevant?	Priority?
The ability to align data initiatives with business objectives	Yes	Supporting
The ability to actively involve data in new projects	Yes	Core
The ability to create a shared data language throughout the organisation by developing and communicating definitions, standards, policies, and rules	Yes	Core
The ability to monitor and subsequently improve data management activities by measuring internalised pre-determined KPIs	Yes	Support
The ability to ensure good communication and collaboration between data, business, and IT stakeholders	Yes	Supporting
The ability to appoint the right central data governance bodies	Yes	Core
The ability to communicate governance bodies' roles throughout the organisation	Yes	Supporting
The ability to recruit the right talent	Yes	Core
The ability to monitor data quality based on pre-defined metrics and targets, to report on these findings in a standardised manner, and to continuously improve data	Yes	Core
The ability to centralise and standardise data processes	Yes	Core
The ability to embed controls in data processes	Yes	Core

COMPANY F3

F3 is an insurance broker headquartered in the United Kingdom, with offices globally. At the time of KPMG's involvement, F3 was about to acquire another company. This forced them to rethink their architecture on different levels: the company itself, processes, the systems supporting these processes, and eventually the data underlying these systems. Before the acquisition, data was given no real thought. Now, F3 was realised their data management practices should be assessed and improved. Interviewee's function is manager data management within F3. Below, the most noteworthy findings are presented.

Concrete organisational changes implemented	• Ad hoc activities were transformed into standard processes, executed in the same way with the same tooling. This drastically increased efficiency. ▪ *Confirms the importance of standardising and centralising processes* • Ownership of data objects was established, forcing people to take responsibility and subsequently better execution of processes.

	Interviewee notes that data management should be guiding, not executing. If you remove the burden of cleaning data from the operational layer, they will be less motivated to execute projects correctly. ▪ *Confirms the importance of governance*
Future ambitions	• Extending data management team to include enough FTEs to extend scope. ▪ *Confirms the importance of recruiting the right talent and of centralising data management activities* • Increasing levels of automation. Interviewee stated that due to the large amounts of data, processes cannot be executed manually. It is however, crucial to incorporate human checks. ▪ *Confirms the importance of embedding checks in processes, adds that automation is necessary for optimisation*
Most important changes in hindsight	• Doing maturity scan explicitly showed the low levels of data maturity. Interviewee stated that it is not always easy getting sponsorship from top management, as data management is not the organisation's primary process. Concrete examples, such as a business case or the scan helps this cause. ▪ *Confirms the importance of gaining sponsorship and creating awareness* • Governance is the most vital implementation. ▪ *Stresses that governance is the most important aspect* • You need all elements of the data management framework, but these two form a crucial foundation. ▪ *Stresses that data management is a broad concept, and all domains should be treated*
Difficulties collaborating with business and IT	• Data management sits on the border between business and IT: it is the responsibility of the business, but IT is deeply involved in the automated solutions. • Business does not always have the technological insights to smoothly operate with IT, and vice versa. • Within F3 it is heavily stressed that data management should not be considered as solely an IT activity. ▪ *Confirms the importance of collaboration between different stakeholders*
Outside of interview protocol	• The interviewee shortly elaborated on impact versus effort for certain changes. Low hanging fruit is for example data governance and standardising processes. An example of high effort/high impact is implementing CRM systems.

Table 14. Validation of capabilities found in F3.

Capability	Relevant?	Priority?
The ability to promote the business rationale of high quality data to increase awareness and acquire sponsorship for change initiatives	Yes	Core
The ability to create a data-friendly company culture	Yes	Core
The ability to define a roadmap with prioritised short- and long-term projects, depending on the organisation's ambitions	Yes	Supporting
The ability to develop and maintain a data strategy, to document and communicate it throughout the organisation	Yes	Supporting
The ability to determine to optimal way in which data should be organised within the business	Yes	Supporting
The ability to monitor and subsequently improve data management activities by measuring internalised pre-determined KPIs	Yes	Core
The ability to monitor data quality based on pre-defined metrics and targets, to report on these findings in a standardised manner, and to continuously improve data	Yes	Core
The ability to appoint the right central data governance bodies	Yes	Core
The ability to communicate governance bodies' roles throughout the organisation	Yes	Supporting

The ability to recruit the right talent	Yes	Core
The ability to centralise and standardise data processes	Yes	Core
The ability to embed controls in data processes	Yes	Supporting
The ability to stablish a central security policy aligned with regulatory requirements	Yes	Core

5.3 Conclusion

Integrating results from case studies and interviews shows that, overall both methods yield similar results. Based on the replication principle, these similarities are considered to be a validation of the capabilities presented earlier. As interviews were evaluating and not leading, capabilities that were not confirmed in company interviews were not rejected. Lacking confirmations were found to be due to these capabilities stemming from cases of companies that were not interviewed. Arguably, it can be expected that more capabilities could have been confirmed if more interviews had been conducted. An overview of which capabilities have been validated in which interviews, is given in appendix D.

The conducted interviews not only validated results, they also enriched the original findings. Both industries underlined the importance of automating data processes, as the sheer amount of data that is worked with is too large for accurate manual processing. Furthermore, while the need for centralised data management activities did not come forward in the manufacturing cases, the importance of doing so was stressed during interviews. These enrichments were incorporated in the original findings, resulting in two changes.

- Automation was added to the centralisation and standardisation of processes for both industries.
- The manufacturing industry gained an extra capability, i.e. centralising data management activities.

By integrating the findings from both research methods, it was thus determined which capabilities companies in both industries need to develop to extract value from data. Moreover, from the different categories it has become clear that, indeed, both business- and data-specific capabilities should be developed. A more detailed explanation of these capabilities will be provided in the next chapter, in which it will be determined whether or not the capabilities are applicable for both industries. The integrated and final results for both industries are presented below, in table 15 and 16.

Table 15. Capabilities derived from manufacturing cases, adjusted to interview results.

	Category	Capability
1	Business	The ability to organise the company in a suitable way
2	Business	The ability to promote the business rationale of high quality data to increase awareness and acquire sponsorship for change initiatives
3	Business	The ability to create a data-friendly company culture
4	Business	The ability to align data initiatives with business objectives
5	Business	The ability to facilitate the cross-pollination of good practices by supporting cross-functional collaboration
6	Business	The ability to actively involve data in new projects
7	Business	The ability to ensure effective decision-making for change initiatives
8	Foundation	The ability to develop and maintain a data strategy, to document and communicate it throughout the organisation
9	Foundation	The ability to create a shared data language throughout the organisation by developing and communicating definitions, standards, policies, and rules
10	Foundation	The ability to centralise data management activities
11	Foundation	The ability to monitor and subsequently improve data management activities by measuring internalised pre-determined KPIs

12	Foundation	The ability to determine to optimal way in which data should be organised within the business
13	Governance	The ability to appoint the right central data governance bodies
14	Governance	The ability to communicate governance bodies' roles throughout the organisation
15	Governance	The ability to recruit the right talent
16	Quality	The ability to monitor data quality based on pre-defined metrics and targets, to report on these findings in a standardised manner, and to continuously improve data
17	Processes	The ability to centralise, standardise, and automate data processes
18	Processes	The ability to embed controls in data processes

Table 16. Capabilities derived from financial services cases, adjusted to interview results.

	Category	Capability
1	Business	The ability to organise the company in a suitable way
2	Business	The ability to promote the business rationale of high quality data to increase awareness and acquire sponsorship for change initiatives
3	Business	The ability to create a data-friendly company culture
4	Business	The ability to align data initiatives with business objectives
5	Business	The ability to facilitate the cross-pollination of good practices by supporting cross-functional collaboration
6	Business	The ability to define a roadmap with prioritised short- and long-term projects, depending on the organisation's ambitions
7	Foundation	The ability to develop and maintain a data strategy, to document and communicate it throughout the organisation
8	Foundation	The ability to create a shared data language throughout the organisation by developing and communicating definitions, standards, policies, and rules
9	Foundation	The ability to centralise data management activities
10	Foundation	The ability to monitor and subsequently improve data management activities by measuring internalised pre-determined KPIs
11	Foundation	The ability to determine to optimal way in which data should be organised within the business
12	Governance	The ability to appoint the right central data governance bodies
13	Governance	The ability to communicate governance bodies' roles throughout the organisation
14	Governance	The ability to recruit the right talent
15	Quality	The ability to monitor data quality based on pre-defined metrics and targets, to report on these findings in a standardised manner, and to continuously improve data
16	Processes	The ability to centralise, standardise, and automate data processes
17	Processes	The ability to develop and communicate frameworks for guiding data processes
18	Processes	The ability to stablish a central security policy aligned with regulatory requirements
19	Processes	The ability to establish a central data life cycle management process
20	Processes	The ability to establish organisation-wide document management standards, regularly review these

6. Industry-Related Differences among Capabilities

The previous chapter determined the organisational capabilities that both the manufacturing and the financial services industries should develop in order to establish a foundation on which to build further data initiatives. To add to these findings, this chapter will test whether or not there are significant differences between the findings in both industries. By researching this, it can be determined whether similar recommendations can be made for different industries, or whether different industries should be treated in isolation. First the methods will shortly be elaborated upon. Afterwards, the results will be discussed, followed by a conclusion.

6.1 Methods and Limitations

This question will be answered by means of a cross-analysis between the practical findings of both industries as presented in the previous chapter. From this analysis, similarities and differences will be defined. Based on these results, inferences will be made concerning the generalisability of the findings. For capabilities that are similar throughout both industries, the assumption is made that they can be applied over several different industries. For significantly differing capabilities fitting explanations will be sought.

As the cross-analyses are based on the case studies and interviews conducted earlier, these are subject to the same limitations. For this research question in particular, generalisability is the most pressing limitation. Conclusions about the generalisability of the findings will be drawn from nine cases in two industries. For more substantiated claims, more cases in more industries should have been treated. Broadening the scope of the industries to include, for example, retailing and governmental sectors, was not feasible in this thesis for reasons of time and availability.

Another limitation which is not related to prior findings, is the fact that studies for both industries were conducted by the same researcher. This may have led to a bias for recognising similar capabilities within different industries. To counteract this bias, findings were extensively discussed with KPMG colleagues that are specialised in either the manufacturing or the financial services industry.

6.2 Results

The conducted cross-analysis resulted in three classes of capabilities: capabilities shared between the two industries, capabilities that deviated but remain applicable in both industries, and capabilities that seem to be industry-specific. These results are summarised in table 17.

Table 17. Results of cross-analysis between industries.

	Shared capabilities
Business	The ability to organise the company in a suitable way
Business	The ability to promote the business rationale of high quality data to increase awareness and acquire sponsorship for change initiatives
Business	The ability to create a data-friendly company culture
Business	The ability to align data initiatives with business objectives
Business	The ability to facilitate the cross-pollination of good practices by supporting cross-functional collaboration
Foundation	The ability to develop and maintain a data strategy, to document and communicate it throughout the organisation

Foundation	The ability to create a shared data language throughout the organisation by developing and communicating definitions, standards, policies, and rules
Foundation	The ability to centralise data management activities
Foundation	The ability to monitor and subsequently improve data management activities by measuring internalised pre-determined KPIs
Foundation	The ability to determine to optimal way in which data should be organised within the business
Governance	The ability to appoint the right central data governance bodies
Governance	The ability to communicate governance bodies' roles throughout the organisation
Governance	The ability to recruit the right talent
Processes	The ability to centralise, standardise, and automate data processes
Quality	The ability to monitor data quality based on pre-defined metrics and targets, to report on these findings in a standardised manner, and to continuously improve data
	Generally applicable deviations
Business	The ability to actively involve data in new projects
Business	The ability to ensure effective decision-making for change initiatives
Business	The ability to define a roadmap with prioritised short- and long-term projects, depending on the organisation's ambitions
Processes	The ability to embed controls in data processes
Processes	The ability to develop and communicate frameworks for guiding data processes
	Industry-specific deviations
Processes	The ability to stablish a central security policy aligned with regulatory requirements
Processes	The ability to establish a central data life cycle management process
Processes	The ability to establish organisation-wide document management standards, regularly review these

From these results it is clear that the undisputed majority of capabilities is shared between both industries, and is thus generally-applicable across various industries. In turn, five capabilities were defined as being deviating, yet generally-applicable. These were only found in the results of a single industry, but are deemed relevant across industries based on the fact that they are trivial in progressing towards a more data-mature organisation. In contrast, three capabilities were defined as being industry-specific to financial services. These capabilities are all focussed on particular processes, i.e. document management, data lifecycle management, and data security. An explanation of the pattern observed here, i.e. only a few industry-specific capabilities among a bulk of generally-applicable ones, is given in the following sections.

The predominance of similar capabilities can be explained by the fact that, overall, an organisation's journey towards data-maturity is similar across different industries. The stages an organisation progresses through in this undertaking were discussed in detail in chapter four. Even though the exact way in which an organisation progresses through these stages may differ, they all share the underlying data-related challenges which have to be addressed in similar ways. The capabilities to overcome these communal challenges were summarised into five categories and are discussed as such below.

1. *Business* capabilities relate to the overarching organisation and the changes to its established practices that are necessary to accommodate the integration of data into the business. The capabilities in this category have a strong strategic focus, relating to organisation and culture. As data management is a complex undertaking, a business should be organised in a way that facilitates it. This will often involve establishing a separate managerial body for strategic decisions relating to data. This body should have the right executive mandate to effectively make these decisions; finding consensus among multiple executive layers will hinder the decision-making process. Furthermore, in order to both align data initiatives with business objectives, and involve data in business projects, a roadmap specifying short- and long-term ambitions should be created. Finally, a

data-driven culture must emerge from the business. In other words, data should be considered a vital and reliable resource throughout all layers of the organisation. Clearly communicating the business rationale of data through, for example success stories, will facilitate the emergence of such a mindset. A data-driven culture also entails that the cross-pollination of good practices is encouraged throughout the organisation, thus catalysing organisational learning.

2. Capabilities in the data *foundations* category are the basic building blocks of further data practices; the development of these capabilities ensures that an organisation is sufficiently equipped for the data initiatives it intends to pursue. These foundations start with, on the one hand a data strategy which provides direction for future initiatives, and on the other hand a set of data standards, definitions, rules, and policies to create a shared language throughout the organisation. This in turn allows the centralisation of data management practices and the optimal integration of data into the business. At a later stage, data practices should be optimised by monitoring and improving data management practices.

3. *Governance* capabilities refer to the human aspect of data-driven capabilities. For any role within an organisation, it is self-evident that the right talent should be recruited and trained. For data in particular, however, the right central governance bodies should be appointed and communicated throughout the organisation. From the interviews it became clear that data ownership is an indispensable concept for successfully becoming data-driven. By appointing ownership for data objects, a sense of responsibility is created, which will in turn improve data operations.

4. The focus of data *processes* is twofold. On the one hand, it focuses on designing business processes in such a way that they are well-suited for the incorporation of data. On the other hand, it focusses on optimally structuring data-specific processes. Such data processes should be centralised and standardised, and clear work instructions should be available to guide them. At a later maturity stage, controls can be incorporated into these processes to both monitor and improve them.

5. The final category, *data quality*, refers to continuously maintaining and improving data quality by building on the results of pre-defined metrics. The benefits and opportunities data can bring about in an organisation all rely on high-quality data; in order for decisions made on the basis of data to be reliable, accurate and complete data is required. An organisation should thus be able to guarantee high-quality data.

The explanation for the presence of industry-specific capabilities, lies in organisation's drivers behind becoming data-driven. For the manufacturing industry, the main goal of becoming more data-driven is the business case that results from optimising and innovating their processes. Here, data management is used to proactively improve and reshape the primary processes of the organisation. Providers of financial services, on the other hand, are tightly intertwined with governmental bodies. Because of this, they are subject to strict legislative requirements, making compliance their number one priority.

Data management in financial institutions is therefore reactive and risk-averse, focussing mainly on justifying their every move to supervisory bodies. Furthermore, it is already largely shaped by regulatory demands, based on what is necessary for the reports you need to hand in. Such a regulatory focus of data within the company therefore influences the way in which data management is organised, directing a lot of the attention to developing and finetuning specific processes that will facilitate reacting to changes in the legislative environment.

6.3 Conclusion

It can be concluded that, overall, the organisational capabilities that companies should develop to become more data-driven are generally-applicable, meaning that the same recommendations can be made across different industries. This can be attributed to the fact that, underneath the surface companies all face similar challenges in their pursuit of becoming a data-mature organisation. As these challenges are similar, the mechanism to address them is as well, i.e. creating a solid foundation of data management practices on which more advanced initiatives can be built. Creating such a foundation is a complex undertaking which relies on several capabilities from five different categories, i.e. business, data foundation, data governance, data processes, and data quality.

However, caution should still be exercised when generalising capabilities. This chapter has shown that, depending on the drivers behind an organisation's data ambitions, industry-specific capabilities can emerge. These capabilities serve to better execute tasks that require more focus in one industry than another. Here, these were only observed for the financial services industry and they were shown to be linked with the strong compliance requirements these companies need to satisfy.

7. Comparison of Existing Literature and Practical Findings

This comparison serves both as a final validation and an enrichment of previous findings. It will do so by comparing the practical results from this thesis with capabilities derived from existing academic literature. Again, first the methods will be elaborated upon and subsequently the results will be presented. The chapter is then wrapped up through a discussion.

7.1 Methods and Limitations

To answer this question, existing literature on data-driven capabilities was collected, integrated, and then compared to the findings of the practical studies conducted in chapter five. From extant literature an overview of capabilities was developed, which was then used as a base on which the practical findings were projected. Similarities and differences identified in this manner were then used to enrich the findings of the research conducted in this thesis. Similarities were interpreted as providing extra strength to the practical findings derived in this study. Differences were analysed and either discarded or incorporated to make the findings from earlier chapters more complete. It should be noted that capabilities related to tooling and technology were removed for the sake of the scope of this thesis.

While looking for the material needed to answer this research question, it soon became clear that very limited academic research explicitly covers data-driven capabilities. While a lack of academic literature stresses the added value of this thesis' findings, it also hinders the answering of this research question. Only a single paper describing such capabilities was found. This paper by Akter et al. (2016) was already briefly touched upon in the literature review of this thesis. The authors developed a framework containing a handful of capabilities needed to become successfully data-driven (table 3). These capabilities, however, have a strong managerial focus and do not dive into data-specific capabilities that should be acquired.

To overcome the lack of literature, two academic maturity models were used to further derive capabilities from. These models (Comuzzi & Patel, 2016; Spruit & Pietzka, 2015) were both based on extant literature and therefore provide a relevant integration of previous academic models. Based on the earlier description of maturity levels, capabilities were derived from levels three and up. When an organisation is positioned at level three, it is considered to use data in a central and thus efficient manner. At higher levels, data management practices mostly are being optimised through automation and benchmarking. These capabilities were then integrated along the same five axes that were used to categorise the earlier findings: business, foundation, governance, processes, and quality.

A limiting factor of these maturity models, is that their primary purpose is not establishing capabilities, but positioning a company on a data maturity scale; results should thus be interpreted with scrutiny. They are still relevant, however, as they provide detailed descriptions of the functioning of mature data organisations. As the capabilities defined by Akter et al. (2016) provide a very managerial perspective, those derived from the maturity model will shed more light on data-specific capabilities.

Several smaller limitations should also be mentioned. While an explicit distinction between industries was made in this thesis, this is not the case for most, if not all, extant literature. To overcome this, the three industry-specific capabilities distinguished in the previous chapter, were not included in the comparisons made. Furthermore, as also discussed in chapter four, more frequently cited publications are more likely to be incorporated. This may have caused additional research on data-driven capabilities to not have been included, therefore limiting this research.

7.2 Results

Table 18 shows the results of the comparison between the capabilities derived in this thesis and those defined by Akter et al. (2016). From these results, it is clear that this thesis defines a higher number of capabilities than the framework by Akter et al., which makes it difficult to map capabilities from both projects one-on-one. This can be attributed to the fact that the capabilities of this thesis go into more detail. What also stands out is the fact that only capabilities from the categories business and data foundations are represented in Akter et al. This is explained by the author's strong focus on managerial and talent capabilities; they do not go into much detail concerning data-specific capabilities. However, while not all capabilities were stated explicitly, some were included indirectly. The capabilities related to governance, for example, implicitly follow from several capabilities that require stakeholders to execute certain tasks.

As there are no contradictory capabilities in this comparison, it was decided not to remove any of the capabilities defined in chapter five. A single capability defined by Akter et al. was not present in the results of this thesis: the so-called business capability, which refers to making the right decisions based on changes in the internal and external environment. This, however, is a basic requirement for running a successful organisation and therefore it is assumed to be in-house independently of any data initiatives and therefore this capability was not added to the earlier findings. It was, however, chosen to reformulate two capabilities based on the work of Akter et al. to make them more accurate:

- "The ability to align data initiatives with business objectives" becomes "the ability to align data initiatives with business objectives *and resources*";
- "The ability to facilitate the cross-pollination of good practices by supporting cross-functional collaboration" becomes "The ability to facilitate the cross-pollination of good practices by supporting cross-functional *and inter-disciplinary* collaboration".

Table 18. Comparison of results with Akter et al. (2016).

Akter et al. (2016)	Thesis	Category
Technology management: Ability to align technology and business goals.	The ability to align data initiatives with business objectives	Business
	The ability to develop and maintain a data strategy, to document and communicate it throughout the organisation	Foundation
Planning: Ability to identify business opportunities and determine how data can contribute.	The ability to promote the business rationale of high quality data to increase awareness and acquire sponsorship for change initiatives	Business
	The ability to actively involve data in new projects	Business
	The ability to define a roadmap with prioritised short- and long-term projects, depending on the organisation's ambitions	Business
	The ability to align data initiatives with business objectives	Business
	The ability to actively involve data in new projects	Business

Investment: Ability to generate greater revenues by strategically investing in data initiatives.	The ability to define a roadmap with prioritised short- and long-term projects, depending on the organisation's ambitions	Business
	The ability to promote the business rationale of high quality data to increase awareness and acquire sponsorship for change initiatives	Business
	The ability to develop and maintain a data strategy, to document and communicate it throughout the organisation	Foundation
Controlling: Ability to strategically assign resources to data initiatives.	The ability to align data initiatives with business objectives	Business
Coordination: Ability to coordinate cross-functional activities.	The ability to organise the company in a suitable way	Business
	The ability to facilitate the cross-pollination of good practices by supporting cross-functional collaboration	Business
	The ability to create a data-friendly company culture	Business
Business: Ability to understand both the internal and external environment and how to react to changes.		
Relational: Ability to communicate with inter-disciplinary environments.	The ability to facilitate the cross-pollination of good practices by supporting cross-functional collaboration	Business
Connectivity: Ability to connect different business departments.	The ability to organise the company in a suitable way	Business
	The ability to centralise data management activities	Foundation
	The ability to create a shared data language throughout the organisation by developing and communicating definitions, standards, policies, and rules	Foundation
Compatibility: Ability to streamline data in order to facilitate processes such as clean-ups and merging.	The ability to create a shared data language throughout the organisation by developing and communicating definitions, standards, policies, and rules	Foundation
	The ability to centralise, standardise, and automate data processes	Foundation
	The ability to determine to optimal way in which data should be organised within the business	Foundation

In turn, table 19 presents the comparison between the capabilities derived in this thesis and those derived from the maturity models. From these results it can be seen that, as predicted in the methods, the capabilities derived from the maturity models confirmed a lot more of the data-specific capabilities defined in this research. This data focus also explains the fact that several business capabilities that were determined in chapter five, were not found in the maturity models. From the data-specific capabilities, only two were not confirmed: recruiting the right talent and controls in processes. These are, however, implied by stating the need for central governance and monitored processes respectively.

Again, no contradictory capabilities were defined, and therefore none were removed. Based on these findings, no capabilities were reformulated. However, the capabilities derived from the maturity models provide valuable elaborations which will be incorporated in the discussion of this thesis. Finally, one capability of the maturity models were not present in those derived here. This will be added to the findings as followed:

- The ability to create a centrally-available overview of all data sources and the systems by which they are used.

Table 19. Comparison of results with extant maturity models.

Author	Maturity models	Thesis
Business		
Comuzzi	There is awareness about the value of data among employees of all levels. Data initiatives are sponsored by top management.	The ability to promote the business rationale of high quality data to increase awareness and acquire sponsorship for change initiatives
Comuzzi	Positive and proactive data-attitude; it is regarded as an important asset for the company. Operations and	The ability to create a data-friendly company culture

	decision-making based on data are embedded in culture and leadership style.	
Comuzzi	Success stories involving data are shared throughout the organisation. This creates a positive attitude which stresses the importance of data.	The ability to create a data-friendly company culture
Comuzzi	Staff feel able to experiment with the possibilities of data.	The ability to create a data-friendly company culture
Comuzzi	Data is incorporated into company-wide strategic and decision-making processes.	The ability to align data initiatives with business objectives
Foundation		
Spruit Comuzzi	There are official definitions, standards, and policies of data for the organisation, takes into account any organisation-specific circumstances. Definitions are known by relevant stakeholders and tangible guides are accessible.	The ability to create a shared data language throughout the organisation by developing and communicating definitions, standards, policies, and rules
Spruit	There is an overview of all data sources and the systems that use them.	
Spruit	There is an organisation-wide data model, which is maintained regularly through clear work instructions.	The ability to determine to optimal way in which data should be organised within the business
Spruit Comuzzi	There are standard formats and interfaces for different departments that frequently exchange data. Data objects can easily be shared across departments and functions.	The ability to centralise data management activities
Comuzzi	Corporate strategy includes data vision and strategy.	The ability to develop and maintain a data strategy, to document and communicate it throughout the organisation
Comuzzi	All data is centrally stored and available.	The ability to centralise data management activities
Comuzzi	Data insights are used to improve strategic alignment objectively.	The ability to monitor and subsequently improve data management activities by measuring internalised pre-determined KPIs
Comuzzi	Data objects are periodically reviewed to assess their usefulness.	The ability to monitor and subsequently improve data management activities by measuring internalised pre-determined KPIs
Quality		
Spruit	The organisation understands how poor data quality influences the business. This applies to both monetary and non-monetary ways.	The ability to promote the business rationale of high quality data to increase awareness and acquire sponsorship for change initiatives
Spruit	Data quality is measured continuously. Measures are in place to improve data quality by pro-actively following up on reported problems.	The ability to monitor data quality based on pre-defined metrics and targets, to report on these findings in a standardised manner, and to continuously improve data
Processes		
Spruit	The entering, updating, and deleting of data is automatically logged by systems to decrease documentation effort and facilitate auditing.	The ability to centralise, standardise, and automate data processes
Spruit	Data processes are accompanied by clear guidelines.	The ability to develop and communicate frameworks for guiding data processes
Comuzzi	KPIs are used to monitor processes across different departments.	The ability to monitor and subsequently improve data management activities by measuring internalised pre-determined KPIs
Comuzzi	Data best practices are communicated across different departments.	The ability to facilitate the cross-pollination of good practices by supporting cross-functional collaboration
Governance		
Comuzzi	Ownership of data is defined, relevant stakeholders engage pro-actively with this role.	The ability to appoint the right central data governance bodies
Spruit	Data stewardship is promoted and embedded in job descriptions.	The ability to appoint the right central data governance bodies The ability to communicate governance bodies' roles throughout the organisation
Spruit	Employees are aware of what data is (not) available to them and where to find it.	The ability to determine to optimal way in which data should be organised within the business

7.3 Conclusion

The aim of this chapter was to both validate and add to the capabilities derived in previous chapters by comparing these to extant literature. As a result, many of the capabilities have been confirmed again, thus providing this study with extra rigor. Furthermore, the results of this chapter have to some extent enriched the original findings; two capabilities have been reformulated to be more complete, and one capability has been added.

After integrating the results from case studies, interviews, industry cross-analyses, and a comparison with existing literature, twenty-one generally applicable capabilities were determined. Furthermore, three industry-specific deviations were identified. Both categories can be found in table 20, which presents the final overview of capabilities that should be developed.

Table 20. Final overview of recommended capabilities.

		Generally applicable capabilities
Business	1	The ability to organise the company in a suitable way
Business	2	The ability to promote the business rationale of high quality data to increase awareness and acquire sponsorship for change initiatives
Business	3	The ability to create a data-friendly company culture
Business	4	The ability to align data initiatives with business objectives and resources
Business	5	The ability to facilitate the cross-pollination of good practices by supporting cross-and inter-disciplinary collaboration
Business	6	The ability to actively involve data in new projects
Business	7	The ability to ensure effective decision-making for change initiatives
Business	8	The ability to define a roadmap with prioritised short- and long-term projects, depending on the organisation's ambitions
Foundation	9	The ability to develop and maintain a data strategy, to document and communicate it throughout the organisation
Foundation	10	The ability to create a shared data language throughout the organisation by developing and communicating definitions, standards, policies, and rules
Foundation	11	The ability to centralise data management activities
Foundation	12	The ability to monitor and subsequently improve data management activities by measuring internalised pre-determined KPIs
Foundation	13	The ability to determine the optimal way in which data should be organised within the business
Foundation	14	The ability to create a centrally-available overview of all data sources and the systems by which they are used.
Governance	15	The ability to appoint the right central data governance bodies
Governance	16	The ability to communicate governance bodies' roles throughout the organisation
Governance	17	The ability to recruit the right talent
Processes	18	The ability to centralise, standardise, and automate data processes
Processes	19	The ability to embed controls in data processes
Processes	20	The ability to develop and communicate frameworks for guiding data processes
Quality	21	The ability to monitor data quality based on pre-defined metrics and targets, to report on these findings in a standardised manner, and to continuously improve data
		Industry-specific deviations
Processes	1	The ability to stablish a central security policy aligned with regulatory requirements
Processes	2	The ability to establish a central data life cycle management process
Processes	3	The ability to establish organisation-wide document management standards, regularly review these

8. Overcoming the Data Value Paradox

Throughout the previous chapters, an overview of both organisational and data-specific capabilities required for transitioning towards a more data-driven organisation was created. Furthermore, the transition itself was explored in more detail, resulting in a model with five maturity stages. By integrating findings from both aspects, this chapter will answer this thesis' final research question: *"Which capabilities should organisations develop to facilitate overcoming the data value paradox?"* In doing so, the overview of capabilities presented in table 20 was cast into a more workable framework for company management, which was the initial research objective of this thesis. The conclusion to this thesis is twofold: previous research will first be synthesised into a workable model, after which the managerial implications for the value paradox are discussed.

8.1 Integration of Previous Findings

Earlier in this thesis, it was shown that strategically developing the right capabilities plays a key role in extracting value from data investments. At the same time, however, it was also stated that only little research into what these capabilities entail was available. Through several qualitative methods, the research in this thesis has generated a list of twenty-one general capabilities on which organisations should focus. In chapter four it was demonstrated that these capabilities cannot be developed at any given moment; their development should, to some extent, be coordinated with an organisation's progression through specific stages. Establishing an order in these capabilities, however, is not a straightforward task. This can be attributed to the fact that the process of becoming more data-driven is not a sequential, step-by-step process. Data management consists of many elements, which all require attention. For this reason, the capabilities presented in table 20 form a network of interdependencies, which influence both each other and the organisation's progression towards a more data-driven business model. To still link them to an organisation's journey, three clusters were derived: establishing, expanding, and enhancing. These clusters reflect which capabilities should be developed over time with decreasing priority. Figure 10 provides a visual representation of these clusters and the capabilities they contain. The figure is addressed in more detail below.

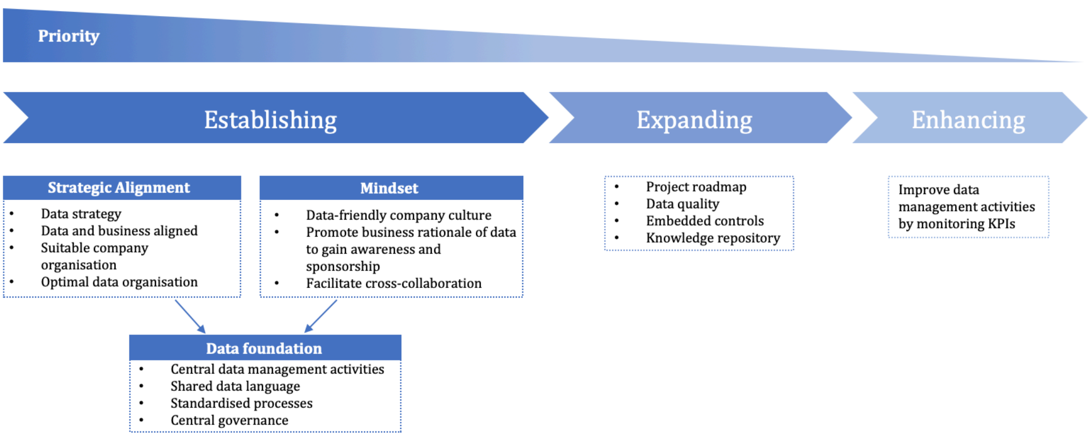

Figure 10. Clustered data-driven capabilities.

The capabilities in the first cluster, *establishing*, form the base on which all future data capabilities and initiatives can be built and are therefore assigned the highest priority. It consists of three sub-clusters: strategic alignment, mindset, and data foundation.

The changes brought about by implementing data initiatives are significant and demand a lot from the organisation. For this reason they should be backed by clear *strategic alignment*. Data activities are often not the primary business process of an organisation, so they should be aligned with business objectives and resources. Establishing an organisation-wide data strategy helps in doing so. It will also contribute to determining the optimal organisation of data within the company. Furthermore, the way a company is organised at top-level and the way decisions are made should be able to accommodate the envisioned changes.

Having a clear view on the company's data ambitions will help in establishing the correct *mindset* throughout the company. Capabilities contributing to this are creating a data-friendly company culture in which different departments and disciplines collaborate to achieve optimal results. Furthermore, the business rationale of data should be communicated throughout the organisation in a convincing manner, both to create awareness and to gain sponsorship from management.

When the data strategy and mindset are sufficiently present in an organisation, the foundations for data management capabilities can be set up. The four capabilities assigned to the *data foundation* sub-cluster are the most basic data-specific capabilities an organisation should develop. Data management activities should be centralised, meaning that they are directed and controlled from a central function in the organisation. Three main implementations help in achieving this. Organisation-wide data standards, rules, definitions, and policies should be implemented, in order to create a shared language for data. Furthermore, data objects should be centrally governed. In this step it is crucial to assign ownership to data objects. This creates a sense of responsibility, thus improving the quality of data objects and therefore improving the execution of processes. These processes, in turn, should be standardised and guidelines for executing them should be clearly communicated throughout the organisation.

In the second cluster, *expanding*, more particular data management advances can be made. To implement these in a strategic and well-ordered manner, a roadmap for short- and long-term projects should be created. This roadmap concerns both data projects, as well as involving data in seemingly unrelated projects. An important point on the strategic agenda should be the pro-active monitoring of data quality by means of pre-determined KPIs. By following up on these results, data quality can be continuously improved. Furthermore, by embedding controls in data processes, unnecessary rework rooted in faulty data can be eliminated. This will significantly enhance these process. Finally, at this stage data is becoming more and more incorporated into day-to-day practices. To maintain an overview of the available data which is accessible to all relevant stakeholders, a knowledge repository should therefore be created. At the time of creating the roadmap, industry-related differences should also be taken into account. Earlier it became clear that an organisation's drive behind becoming more data-driven can lead to the need for specific capabilities to be developed. These ambitions, which are often industry-related, should therefore always be kept in mind when creating a roadmap for the implementation of data initiatives.

In the final cluster, called *enhancing*, the focus is on optimising data management practices. This cluster only defines a single capability which applies to all data management domains. This is the capability to improve data management practices by continuously measuring and monitoring pre-defined KPIs. The effectiveness of different data domains is measured and reported on. By following up on any incidents or bottlenecks defined in these reports, data management, and with it the value derived from data initiatives, can be improved.

8.2 Managerial Implications for the Value Paradox

From figure 10 it is clear that the establishing cluster contains the majority of capabilities, whereas the expanding and enhancing ones contain significantly fewer. When linked to the stage model defined in chapter four, this observation holds a key implication for the data value paradox.

In the first two data maturity stages, undisciplined and repeatable, data management practices have not yet been standardised and adopted throughout the entire organisation. Data is predominantly used in a reactive manner, and therefore pay-offs are minimal. Pay-offs from data initiatives only start to emerge when an organisation reaches the third and fourth data maturity stages: repeatable and measurable. At this point, data management practices are being centralised and employed on an organisation-wide scale. Furthermore, these practices are starting to being improved by tracking certain pre-determined KPIs. These features allow an organisation to use its data in a pro-active and steered manner. With this, value is generated as a natural consequence. When an organisation reaches the final data maturity stage, continuous improvement of data management practices is central. This, in turns, allows an organisation to extract increasing amounts of value with more ease. To summarise: the more an organisation progresses through the different data maturity stages, the more value it can extract from its data investments.

In order to progress through these maturity stages, certain capabilities have to be developed. In the previous section, figure 10 has provided both an extensive overview of these capabilities as well as a clustered order in which these should be developed. The first cluster, establishing, can be coupled to an organisation which finds itself around the first two data maturity stages. Before an organisation can successfully move to the next stage, the capabilities listed in this cluster should be in place. This implies that the majority of capabilities needs to be developed at a point where the organisation will not yet extract considerable value from its data practices. In other words, the return on both the time and money invested in developing these capabilities, will be minimal at this point. For this reason, investing in establishing capabilities will often not be of direct interest to company management. The capabilities categorised as expanding (maturity levels three and four) and enhancing (maturity level five), on the other hand, will have a more direct on the business and are therefore often considered more attractive investments.

The previous paragraph has made clear that, while theoretically certain capabilities are deemed crucial in progressing towards a data-mature organisation, the investments necessary for developing them are not always considered as lucrative from a managerial perspective. However, the fact that the clear majority of capabilities defined in this thesis is grouped in the establishing cluster, whereas significantly fewer are present in the other two, stresses the importance of creating a good foundation that paves the way for realising further data ambitions. Establishing this foundation requires many elements to be addressed, not all of which will have direct effects on business performance. All elements, however, build on each other and contribute to an organisation's data-maturity.

The key managerial takeaway is that, it is the aggregate of multiple elements that lies at the source of a company's ability to create value from data, and thus of its ability to overcome the data value paradox. A successful data-driven organisation stems from many interwoven capabilities that were developed at the right time for them to build on each other. Even though not all capabilities presented here will result in quick wins, they should nevertheless be developed with care. As both data and data management are complex concepts, neglecting establishing capabilities would be at the detriment of any future data-ambition an organisation may hold. A fitting metaphor to illustrate this conclusion, would be building a house. The different levels build on each other and require a good foundation to eventually get to the top. Foundations with elements missing, might ultimately collapse.

9. Discussion

This thesis took a step-by-step approach to create an overview of the capabilities organisations should develop in the process of becoming more data-driven. The final chapter presented here will reflect back on these findings, by discussing them in light of the literature review conducted earlier. Next, a reflection on the methods employed is given, followed by recommendations for future research. The latter section will conclude this chapter.

9.1 Reflection on findings

The starting point for the research conducted in this thesis was the value paradox surrounding data; despite data's promising benefits, businesses struggle to capture its value. Through a literature review, this paradox was shown to be rooted in the fact that, before they can create value, assets acquired through data investments need to be transformed into firm-specific capabilities. In turn, these capabilities allow companies to transform data into knowledge, which can then lead to value by means of, among others, data-driven decision-making, optimising processes, and improved compliance. While extant literature has made valuable contributions to this topic, it was argued to be limited, as an overview of what capabilities should be developed did not exist. Furthermore, the main focus of existing data literature lies with tooling and technology, while organisational aspects remain largely neglected. The research conducted in this thesis has addressed this shortcoming by generating a list business and data-specific capabilities that organisations are advised to develop.

Capabilities were defined in five categories: business, data foundations, data governance, data processes, and data quality. Together these categories encompass the contextual factors Davenport et al. (2001) deemed necessary for turning data into knowledge and value (i.e. strategy, skills & experience, organisation & culture, and data). Leaving technology out of scope allowed a deeper focus on the organisational aspects of both business and multiple data dimensions. On the one hand, the capabilities resulting from the analyses were directed at elements such as creating the right company organisation and culture, and aligning data initiatives with overarching business objectives. On the other hand, several vital data elements were outlined. These include, among others, centralising data management practices, striving for high data quality, appointing the right governance bodies, designing suitable data processes, et cetera. A parallel can be established between these results and the general directions provided by earlier authors, such as Ross et al. (1996), Gupta and George (2016) and Akter et al. (2016). The axes around which these authors evolved their findings can be also found in the findings of this research. These axes were discussed in detail throughout the literature review of this thesis. They involve tangible, human, and intangible assets, as well as capabilities that allow for optimal planning, coordination and alignment, and the management of business and relational aspects.

A unique extension of these insights was provided by the comparison of two different industries. From cross-analyses between manufacturing and financial services, it became apparent that the majority of capabilities to be developed is applicable across industries. This was shown to be rooted in the fact that, overall, organisations go through the same journey when becoming more data-driven, roughly moving from ad hoc to highly centralised, standardised, and optimised data practices. Besides these general capabilities, several capabilities were also labelled as being industry-specific. These were shown to result from the unique drivers behind becoming more data-driven. For manufacturing these drivers were argued to lie in optimising and innovating production processes, while financial services were argued to incorporate data mainly for compliance purposes. This makes sense when relating these findings to Grönroos (2000). The value of manufacturing products, as parts of the goods industry, is created in the production phase. Therefore, improving the production process will ultimately increase the value of products and subsequently business value through several mechanisms, such as lower production costs, higher quality products, more customer satisfaction, increased brand reputation, et cetera. Contrarily, the value of services revolves around interactions. As these customer interactions are prone to sensitivity due to reasons such as privacy, a logical link can be made by increasing intrinsic company value through good compliance.

The final major finding of this thesis is concerned with developing the right capabilities at the right time, which was shown to be an important element of complex change undertakings in chapter four. The level of complexity for becoming more data-driven can be argued to be particularly high due to two reasons. First, the nature of data was touched upon in the introduction. The high volume, velocity, and variety of today's data challenges the traditional ways in which organisations use data. Second, it was argued that, because data is becoming more widely available among organisations, it is simultaneously becoming more important and more difficult to employ it in a value-creating way. To do so, data should thus be managed skilfully. Through the DAMA-DMBOK framework, however, it was shown that data management is an intricate endeavour, consisting of eleven dimensions that should all be sufficiently addressed. Because of this additional complexity, the capabilities defined in this thesis are a valuable contribution to navigate company management towards a data-mature organisation. From assigning a sequential order to the defined capabilities, it became clear that they form a network of interdependencies which should not be treated in isolation. Instead, three clusters were identified, which could roughly be matched with the data maturity stage model defined in this thesis. From these clusters it became apparent that many investments have to be made at an early stage in which the return will still be low. Investing in these capabilities may therefore not be a top priority of company management. The findings of this research, however, strengthen the claims about the importance of a solid foundation of organisational capabilities, thereby providing an extra trigger for management to carefully build their data management practices and to not skip steps in this process.

9.2 Reflection on Methods and Recommendations for Future Research

Methodologically, the research conducted in this thesis adds to existing literature by employing a spectrum of methods that were specifically chosen to iteratively refine and enrich results. A literature review first shed light on the steps an organisation takes in becoming data-mature. Capabilities were then derived from use cases and were then validated and enriched by means of interviews with the respective firms. Their generalisability was then researched by comparing two industries, and finally a comparison with the limited literature available was made for further validation and enrichment.

Taking on a qualitative approach allowed an in-depth exploration of a topic on which little research has been performed. By studying use cases of various companies that were transitioning towards a more data-driven organisation, the phenomenon of data-driven capabilities was studied in a real-life setting. Although the available literature was a valuable guide to build on, the practical approach taken in this research provided the context and insights necessary to deduct specific capabilities. The interviews that were subsequently performed sketched important additional context. These insights, combined with the findings from the literature review on data maturity, formed an important contribution to the managerial implications presented in chapter eight. Although the final review of existing literature contributed to the findings by validating them, the enrichments stemming from it are minimal. Furthermore, as academic literature on the topic was limited to begin with, including this method may have not have had the desired effectiveness.

Despite the chosen methods being suited for this research, the findings are also subject to several limitations. For the most part, the limitations can be translated into opportunities for future research. These opportunities, as well as the limitations they stem from, will be discussed below.

A first constraining factor pertains to the scope of this research. The scope is limited over two axes: The data sample contained limited observations (confined to ten cases and three interviews) and these were taken from a selection of two industries. Future research can benefit from extending this scope along both axes. Increasing the sample size for cases and interviews will increase the reliability of findings. Including more industries would allow more robust inferences to be made concerning the generalisability of the recommended capabilities. Industries of different categories, such as the public industry, can be included first, after which comparisons can be extended to include more specific sectors within the same group.

Secondly, although qualitative methods seem to be the right choice for this research, taking on a quantitative approach in the future would provide additional insights. Besides providing statistical validation, quantitative analyses allow explicit links between capabilities and firm performance to be established. Seeing that the literature review of this thesis argues that incorporating data into the business can lead to a competitive advantage, it would be a valuable contribution to research such links, for example based on pre-determined KPIs that can be obtained through organisations' public reports.

A third factor that may influence this research, is the close collaboration with a third party. Although this affiliation made this research possible, it may have simultaneously led to a bias. Cases were based on projects in which solutions were provided to firms that encountered problems in their data management practices. As consulting firms have established methods of making recommendations, the capabilities defined may have been influenced by this. To avoid possible biases like these in future research, data could for example be collected through interviewing individuals at data-driven companies which are not involved with a specific third party.

Finally, a recommendation which is not based on methodological constraints can be made. Due to the early stages in which research into the capabilities leading to better data management finds itself, this research was mostly exploratory. In the future, it would be of interest to incorporate these capabilities into a more holistic model of the firm , which elaborates on, for example the specific interdependencies between capabilities and external influences.

References

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. https://doi.org/10.1016/j.ijpe.2016.08.018

Barney. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management*, *17*(1), 99–120.

Barney, J. B., & Arikan, A. M. (2001). The resource-based view: Origins and implications. *Handbook of Strategic Management*, *124188*.

Barney, J. B., & Hesterly, W. S. (2008). *Strategic management and competitive advantage: Concepts* (2. ed). Upper Saddle River, NJ: Pearson Prentice Hall.

Becker, J., Knackstedt, R., & Pöppelbuß, J. (2009). Developing Maturity Models for IT Management: A Procedure Model and its Application. *Business & Information Systems Engineering*, *1*(3), 213–222. https://doi.org/10.1007/s12599-009-0044-5

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

Bhatt, G. D., & Grover, V. (2005). Types of information technology capabilities and their role in competitive advantage: An empirical study. *Journal of Management Information Systems*, *22*(2), 253–277.

Brynjolfsson, E., & Hitt, L. M. (1998). Beyond the productivity paradox: Computers are the catalyst for bigger changes. *Communications of the ACM*, *41*(8), 49.

Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in numbers: How does data-driven decisionmaking affect firm performance? *Available at SSRN 1819486*.

CMMI Institute. (2014). *Data Management Maturity (DMM) Model*. CMMI Institute.

Comuzzi, M., & Patel, A. (2016). How organisations leverage Big Data: A maturity model. *Industrial Management & Data Systems*, *116*(8), 1468–1492. https://doi.org/10.1108/IMDS-12-2015-0495

Constantiou, I. D., & Kallinikos, J. (2015). New games, new rules: Big data and the changing context of strategy. *Journal of Information Technology*, *30*(1), 44–57. https://doi.org/10.1057/jit.2014.17

DataFlux Corporation. (n.d.). *The Data Governance Maturity Model: Establishing the People, Policies and Technology that Manage Enterprise Data*. Retrieved from http://www.fstech.co.uk/fst/whitepapers/The_Data_Governance_Maturity_Model.pdf

Davenport, T. H., Barth, P., & Bean, R. (2012). How "big data" is different. *MIT Sloan Management Review*, *54*(1), 22–24.

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.

De Bruin, T., Freeze, R., Kaulkarni, U., & Rosemann, M. (2005). *Understanding the main phases of developing a maturity assessment model*.

De Mauro, A., Greco, M., & Grimaldi, M. (2015). *What is big data? A consensual definition and a review of key research topics*. 97–104. https://doi.org/10.1063/1.4907823

Dell EMC, & IDC. (2014). *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*. Retrieved from https://www.emc.com/collateral/analyst-reports/idc-digital-universe-2014.pdf

Domo. (2018). Data Never Sleeps. Retrieved March 18, 2019, from https://www.domo.com/blog/data-never-sleeps-6/

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, *14*(4), 532–550.

Gartner. (2019). Gartner IT Glossary. Retrieved June 19, 2019, from https://www.gartner.com/it-glossary/

Gobble, M. M. (2013). Big Data: The Next Big Thing in Innovation. *Research-Technology Management*, *56*(1), 64–67. https://doi.org/10.5437/08956308X5601005

Grant, R. M. (1996). Prospering in Dynamically-Competitive Environments: Organizational Capability as Knowledge Integration. *Organization Science*, *7*(4), 375–387. https://doi.org/10.1287/orsc.7.4.375

Grönroos, C. (2000). *Service management and marketing: A customer relationship management approach* (2nd ed). Chichester ; New York: Wiley.

Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management*, *53*(8), 1049–1064. https://doi.org/10.1016/j.im.2016.07.004

Henderson, D., Earley, S., & Data Administration Management Association (Eds.). (2017). *DAMA-DMBOK: Data management body of knowledge* (Second edition). Basking Ridge, New Jersey: Technics Publications.

Hodkinson, P., & Hodkinson, H. (2001, December). *The strengths and limitations of case study research*. *1*. Cambridge.

IT Governance Institute. (n.d.). *COBIT 4.1 Executive Summary*. Retrieved from https://www.isaca.org/Knowledge-Center/cobit/Documents/COBIT4.pdf

Janssen, M., & Kuk, G. (2016). Big and Open Linked Data (BOLD) in research, policy, and practice. *Journal of Organizational Computing and Electronic Commerce*, *26*(1–2), 3–13. https://doi.org/10.1080/10919392.2015.1124005

Kim, G.-H., Trimi, S., & Chung, J.-H. (2014). Big-data applications in the government sector. *Communications of the ACM*, *57*(3), 78–85. https://doi.org/10.1145/2500873

Kiron, D., Prentice, P., & Ferguson, R. (2014). *The Analytics Mandate*. MIT Sloan Management Review.

Klievink, B., & Janssen, M. (2009). Realizing joined-up government — Dynamic capabilities and stage models for transformation. *Government Information Quarterly*, *26*(2), 275–284. https://doi.org/10.1016/j.giq.2008.12.007

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review*, *52*(2), 21–31.

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

Marr, B. (2018, May 21). How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read. Retrieved March 18, 2019, from Forbes website: https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#64f2d22060ba

Martijn, N. L., & Jonker, R. A. (2015). The Effects of Data Governance in Theory and Practice. *Compact*, *42*(1), 25–31.

Mason, J. (2002). *Qualitative researching* (2nd ed). London ; Thousand Oaks, Calif: Sage Publications.

McAfee, A., Westerman, G., & Bonnet, D. (2014). The Nine Elements of Digital Transformation. Retrieved June 3, 2019, from MIT Sloan Management Review website: https://sloanreview.mit.edu/article/the-nine-elements-of-digital-transformation/

McAfee, M., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard Business Review*, (October), 61–68.

Melville, N., Kraemer, K., & Gurbaxani, V. (2004). Information technology and organizational performance: An integrative model of IT business value. *MIS Quarterly*, *28*(2), 283–322.

Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2017). Big data analytics capabilities: A systematic literature review and research agenda. *Information Systems and E-Business Management*. https://doi.org/10.1007/s10257-017-0362-y

Mikalef, P., & Pateli, A. G. (2016). Developing and Validating a Measurement Instrument of IT-Enabled Dynamic Capabilities. *ECIS*, ResearchPaper39.

Nolan, R. L. (1979). Managing Crises of Data Processing. *Harvard Business Review*, *3*(4).

Reich, B. H., & Benbasat, I. (2000). Factors that influence the social dimension of alignment between business and information technology objectives. *MIS Quarterly*, 81–113.

Roach, S. S. (1987). *America's technology dilemma: A profile of the information economy*. Morgan Stanley.

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* (pp. 1–34) [TDWI best practices report]. The Data Warehousing Institute.

Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. *2013 International Conference on Collaboration Technologies and Systems (CTS)*, 42–47. https://doi.org/10.1109/CTS.2013.6567202

Schryen, G. (2013). Revisiting IS business value research: What we already know, what we still need to know, and how we can get there. *European Journal of Information Systems*, *22*(2), 139–169. https://doi.org/10.1057/ejis.2012.45

Siggelkow, N. (2007). Persuasion with case studies. *Academy of Management Journal*, *50*(1), 20–24.

Taylor-Cummings, A. (1998). Bridging the user-IS gap: a study of major information systems projects. *Journal of Information Technology*, *13*(1), 29–54.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, *18*(7), 509–533.

The Economist. (2017, May 6). The world's most valuable resource is no longer oil, but data. *The Economist*. Retrieved from https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

Wade, M., & Hulland, J. (2004). Review: the resource-based view and information systems research: review, extension, and suggestions for future research. *MIS Q*, *28*(1), 107–142.

Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, *165*, 234–246. https://doi.org/10.1016/j.ijpe.2014.12.031

Wang, H.-L. (2014). *Theories for competitive advantage*.

Yin, R. K. (2014). *Case study research: Design and methods* (Fifth edition). Los Angeles: SAGE.

Zachman, J. A. (1987). A framework for information systems architecture. *IBM Systems Journal*, *26*(3), 276–292.

Zeleti, F.A. (2018). *Capability Architecture for Open Data*. National University of Ireland, Galway, Galway.

Zeleti, Fatemeh Ahmadi, & Ojo, A. (2017). Open data value capability architecture. *Information Systems Frontiers*, *19*(2), 337–360.

Appendix A – DAMA DMBOK

The DAMA DMBOK framework identifies eleven data management pillars. These are elaborated upon below. All information was retrieved from the official DAMA-DMBOK framework (Henderson et al., 2017).

1. **Data governance**
   Concerns shared decision making, roles and responsibilities, policies and guidelines, in other words the overall structure needed. Specifies for example how decisions are made or what actions people are supposed to follow. Formalising these elements allow an organisation to extract more value from data.

2. **Data architecture**
   Models, policies, rules, and standards that govern which data is collected and how it is stored, arranged, and put to use in a database system and/or in an organization. It bridges the gap between business and IT by mapping business requirements to technological ones. Data initiatives can then be aligned with the overall view on data and its storage.

3. **Data modelling and design**
   Concerns a holistic view of the data landscape, represented through a model. Used to discover, analyse, and scope data requirements, and then representing and communicating these in a data model. Enables an organization to better understand its data assets, shows how they fit together.

4. **Data storage and operations**
   Concerns managing the entire lifecycle from creation/acquisition until disposal. Includes implementing policies and procedures and executing these for acquisition, migration, transformation, retention, expiration and disposition of data.

5. **Data security**
   Concerns the planning, development, and execution of security policies and procedures to provide proper authentication, authorization, access, and auditing of data and information assets. Differs between industries and countries.

6. **Data integration and interoperability**
   Concerns managing the movement and consolidation of data between applications. Allows multiple systems to communicate and data to stay consistent throughout the process.

7. **Document and content management**
   Concerns storing, accessing, and using the organisation's electronic documents. Properly managing this increases efficiency by for example reducing the time spent searching for documents.

8. **Reference and master data**
   This is the key to efficient business operations, it ensures that decision-making is based on reliable and complete data. Activities include among others reducing redundancy, ensuring high data quality, et cetera.

9. **Data warehousing and business intelligence**
   Warehousing is concerned with the integration of data from multiple sources into a well-functioning database. Business intelligence is concerned with developing mechanisms to extract the exact data necessary for conducting specific analyses.

10. **Metadata**
    Concerns information that provides meaning or context to data. If not properly managed, the organization risks losing valuable info that it actually possesses.

11. **Data quality**
    Concerns all activities that are meant to increase the quality of data.

Appendix B – Interview protocol

Below the interview protocol defining the questions is given.

Part 1 – Introduction
Room for both parties to shortly introduce themselves.

Part 2 – Situation company pre-KPMG
This part will elaborate on the company's situation (data related) before the arrival of KPMG.

1. In the end, why did you choose to involve an external party?
2. Do you know of any incidents related to this decentralised manner of MDM that had negative impact on the overall business?

Part 3 – Situation post-KPMG
By now it has *x years* since KPMG has been involved. This part will focus on any additional changes that have been implemented and/or additional ambitions the company may have for the future.

1. Have any changes that were not advised by KPMG been implemented? If so, can you shortly elaborate on these?
2. Does your company have any new ambitions concerning data management? These ambitions can be located anywhere in the pipeline: ideas, projects that are being designed, or project that have already been employed.

Part 4 – Reflection
This part will shortly cover the interviewee's experience of the changes brought about by KPMG.

1. In hindsight, what changes would you classify as being the most important?
2. Has the improvement of data management practices contributed to you company's competitive differentiation in any way? If yes, how?
3. Have you ever experienced difficulties in the collaboration between business and IT stakeholders?

Part 5 – Capabilities
After the more open part of the interview, this part is more focused. In the attachment, I have included a list of capabilities that I defined based on studying the company's report. For each of these capabilities, could you:

1. Define whether or not you agree? In other words, do you deem this capability important or not so much?
2. Classify whether you consider this a core/supporting/specialised capability?

Part 6 – Wrap up

Appendix C – Interview transcripts

Interview company M1

1. Introduction

In this section, both parties introduce themselves.

Interviewer elaborates on the following:

- Thesis topic and goals;
- Importance of interviews;
- Anonymisation of data;
- Whether or not the interviewee agrees to recording the interview. In this case, the interview was recorded.

Interviewee introduces himself and company M1, he has two main roles:

1. Head in the global MDM organisation, which is of the organisation's GBS. M1 moved to a GBS model three years ago now, KPMG helped to set up MDM here. The rationale for doing so, is that it is difficult to find the right host for MDM. Usually other functions would host it, but in the GBS it has become a standalone functions next to e.g. finance, marketing, et cetera. Interviewee is thus responsible for the MDM team in this GBS. It was set up only recently (last year), so they are still transitioning and the team is growing.

   Created a dept based on existing decentralised team., but scope is not complete. They want to centralise more services. Not only data that are straightforward from eg ERP, but also HR, IM (internal IT function), legal and compliance, . Before ERP was the focus of MD but now extended it

2. Furthermore, interviewee is MDM GPDL. A process house with governance for end-to-end (E2E) processes was built. The project with KPMG helped to raise the need for this with top management. Within this process house, there are towers for multiple processes, e.g. innovation, sales, et cetera. Within each tower, there is an executive committee (highest management), a GPO (global process owner - for MDM, this is the CIO), and a GPDL (global process delivery lead). The GPO is responsible for setting a strategy for the roadmap, whereas the GPDL is responsible for implementation. Interviewee notes that the GPDL is also involved in strategic thinking, but he is really responsible .

The two roles fit very well together. He is responsible for implementation on one hand, but also responsible for the team that executes the implementation.

2. Reasons for involving KPMG and organisational changes

In this section, the reasons for involving KPMG were discussed. It also elaborates on the concrete changes implemented to realise M1's ambitions.

What eventually led to the involvement of KPMG?
The timing was just right, several events coincided:

- M1 embarked on the GBS journey;
- New CIO came on board;
- IM was transferred to another party;

All elements required processes that worked smoothly, for which MD is key. At the time, interviewee was responsible for the MDM department of IM, so looking after the application and the support of the applications. He was the only person with a global role, and was therefore asked to run an MDM project; determine what is going well, what is needed for improvement, et cetera. This all occurred around the same time that M1 met with KPMG, so it was easy to carry on jointly.

What concrete steps have been taken to accomplish your ambitions?
With KPMG a proposal was made, which led to four areas of attention:
1. Governance – M1 had no governance system;
2. Organisation – There was no central organisation, everything occurred on a global level with different processes, no harmonisation;
3. Data quality and KPIs – reporting was lacking, barely any mechanisms in place to track DQ;
4. Tooling – M1 was at a good level for this aspect.

This proposal led to a follow-up project. In this, tooling was disregarded as this was already at a good level. The process governance mandate was delegated to the executive committee, as this was important to the entire organisation, not just MDM. This left the proposal with two elements:
1. Organisation – Set up MDM in GBS by transferring the existing decentralised practices. Create a CoE and an operations unit. The latter is where the business case was, as labour arbitrage could justify the investment.
2. Reporting – A global dashboard was to give a single view about DQ for the whole business. KPMG's initial design changed a lot. M1 needed a more global system that worked across systems.

In all this, it is important to come up with investment requests. Interviewee agrees that is hard to link MDM with the value it can deliver. There are ways to do this, but the problem is that the investment of doing so is high. To offset this, M1 made a business case/proof of value by focussing on part of a single process. They researched the impact of MDM on reworking processes. Ideal is a process that works smoothly from A to Z, without reworking anything; nothing misses and nothing is mismatched.

If we move a couple of steps back, you were talking about the value MDM can offer. Did the challenges related to MDM before have negative financial impact on the company?
Interviewee states that there is a known impact on the company. However, resources are not spend in order to prove this, these analyses are very work intensive. Instead focus on tackling problems at the source. GPOs and GPDLs are asked for the pain points in processes, and for their priorities and ambitions. What the MDM team then intends to do is determining which of these pain points are caused by MDM. Based on this and the priorities, you can improve and solve problems (by e.g. centralising MDM maintenance, by defining standards, et cetera.)

3. Reflection
This section reflects on the changes that have been implemented.

Have you reached your goals? Do you have any further ambitions on the horizon?
- One ambition is to further extend the scope of our current project, as not all MD has been centralised, controlled, and governed. Get full coverage on MD from e.g. customer, vendor, material, financial departments.
- There are also new projects in the pipeline, mainly concerning the automatization of processes, simplifying processes (some are over-engineered). Optimising processes for efficiency would save costs by e.g. reducing team sizes, defaulting certain data in certain systems. E.g. overall there are still a lot of manual checks. M1 is able to pursue these new projects cause of several reasons: time is freed up for relevant stakeholders and there is

more visibility concerning DQ and performance management (you can see the bottlenecks in the process with more ease).

In hindsight, what have been the most important changes?
Governance is key. Interviewee stresses that M1 would not be where it is today if the governance had not been set up properly. From governance you can drive everything else. E.g. M1 sets targets for their GPOs – they have to come up with a number of new standards within a certain amount of time, this in turn allows data cleansing, improving reporting quality, etc.

- *Would that confirm the claim I make in my thesis, about the foundation needing to be in place before good MDM can be set up?*
  Yes. An advantage of governance is also that you put people in charge of their processes, making them responsible for certain data. Last year M1 implemented data ownership (this is not sole ownership). In this way, people cannot "throw the hot potato at each other", i.e. when something goes wrong with finished good, you know where to go. The owners then need to pull any other stakeholders into the discussion. In conclusion, having governance and ownership of MD agreed and signed off on forces people to take responsibility for their data.

Did the changes you implemented contribute to differentiating yourself from competition?
Interviewee says this is difficult to say.

- *Was it more of an internal drive that led to your ambitions?*
  Yes, it was mostly about moving towards the GBS. Of course, indirectly this makes the company run in a smoother and harmonised way, and it gives a better view of what is going on and thus where improvement is needed. At the moment, however, it is still too early to benchmark, for this M1 needs to be a bit more mature – currently they are still evolving too much.
- Do you have the idea that you are setting a trend, i.e. centralising MD, for your competitors? Are you aware what they are doing?
  Interviewee states that he doesn't have much reference on this; there are little references in the same market. A large part of M1 operates in a B2C way, while most chemical manufacturers operate in a B2B way.

Have you ever experienced difficulties in the collaboration between IT and business stakeholders?
We try not to be IT. It is important not to look at MD as a technology/IT concept only. You should think of it as a complete service of which tooling is a (small) part. Many people take this shortcut too easily.

4. Capabilities

In this section, the interviewee's opinion on the capabilities defined in the case study was asked. Asked whether or not he found the capability relevant and if he would label it as core/supporting/specialised. These findings are specified in an Excel file. Some additional remarks are given below:

- "Communicate and promote the value of data"
  It is about more than communicating, also educating.
- "Align with and involve data in new projects and changes"
  It is a struggle to make sure that new projects involve the right stakeholders, often they prefer to execute projects in isolation. This is not the way to do it.
- "Classify the type of data: utility, modular, differentiating"
  Used to do this, but it became less relevant. All data can be looked at globally, and after you can determine the service it can help provide.

- "Ensure availability of right skills and expertise"
  Difficult, cause seniority tends to lack due to turnover. In your CoE you are left to work with what you have when the senior layer leaves and the junior layer comes along. M1 tries to solve this buy recruiting at senior levels. It is not the fact that people are not good, they are just junior and you cannot expect them to speak to the top manager in the right way.

Interview company M3

1. Introduction

In this section, both parties introduce themselves.

Interviewer elaborates on the following:

- Thesis topic and goals;
- Importance of interviews;
- Anonymisation of data;
- Whether or not the interviewee agrees to recording the interview. In this case, the interview was recorded.

Interviewee introduces himself and company M3:

- Four supporting pillar can be distinguished in the organisation: HR, processes, data, IT. Interviewee is team lead of the data department. Within the data team, there are two classes of people: data administrators (creating and maintaining data objects) and data analysts (use data for analysis purposes, e.g. distinguish trends).
- In the past, company M3 consisted of several separate companies which fused over time. Now, these "separate" companies are being integrated in a uniform manner (i.e. uniform processes, systems, data ,et cetera) over these four axes. This uniformization is collectively termed "one M3".
- Data management at M3 started in an operational manner: creating data objects for customers and suppliers. At this level, the collaboration with KPMG was started.

2. Reasons for involving KPMG

In this section, the reasons for involving KPMG were discussed.

What were the main data-related challenges M3 faced leading to the involvement of KPMG?
Up to KPMG, data management was confined to creating data objects for customers and suppliers. Interviewee refers to DAMA-DMBOK to stress the many aspects of data management. Data was used for administrative purposes rather than for pro-actively creating value. Although it was recognised that data had a lot of potential, this was not an active topic throughout the organisation; there was no team working to extract this extra potential.

The reasons for employing KPMG were roughly threefold:

- The KPMG team had more specialised knowledge of data management. With their help, the multiple aspects of data management became more professionalised. Interviewee mentions that, with the help of KPMG, the team realised they were actively pursuing more elements of data management than initially thought. These needed more professionalisation to be implemented efficiently.
- Hiring a costly external party increased the momentum of implementing changes; company is more willing to dedicate time in this manner.
- Within M3, there was little awareness about the data management team. KPMG helped to create this awareness.

3. Organisational changes

This section elaborates on the concrete changes implemented to realised M3's ambitions.

What concrete steps have been taken to accomplish your ambitions?
Interviewee stresses that the process is one of incremental growth. Several steps, however, have formed the foundations on which they can further build their data management practices.

- KPMG helped to put the data management department on the map throughout the entire company. With the help of KPMG, a service catalogue was created. This catalogue introduced the team and the services they can provide for the company.
- KPMG helped to create a data dictionary which helps the company in "speaking the same language" for data. Interviewee says that the hardest part is not creating this dictionary, but communicating and implementing it throughout the organisation in a gripping manner. It is complicated to link these data terms with actual work processes.

  So how has M3 addressed this challenge?
  After KPMG, M3 has continued to work on implementing the data dictionary; on concretely translating it to the rest of the business. The dictionary was integrated with M3's intranet, so that it is available to relevant stakeholders. Here you can either find how to create/change/… a certain master data object yourself, or you can find the data steward responsible for doing so.

  For instructing employees about involved data objects and stakeholders, a portal (workbase) has been developed. This functions as an internal Wikipedia, so to speak.
  - It starts with an overview of the entire organisation;
  - You can then selected a department;
  - This department is broken down in certain processes. Using processes allows you to transcend business units, which creates uniformity. Interviewee notes that business units can be very established in certain ways, and implementing changes can therefore have a large impact.
  - Per process, the portal shows all data objects involved, as well as the responsible stakeholders per data object (note that data administrator has a central position, whereas stewards are located in the business units).
  - Per data object, there is also a link to other processes with uses this object.

4. Reflection

This section reflects on the changes that have been implemented.

Have you reached your goals? Do you have any further ambitions on the horizon?
Data is a very broad concept, so you can keep working on new projects. M3 is currently updating BI practices (moving it to the cloud, implementing new tooling). They are also creating products that use data, which leads to a lot of security aspects. Basically you have all the different elements of data management (DMBOK) in which you keep professionalising. Up till now, the focus was very much on governance, this really forms the base on which other aspects can build.

Why is data governance so important, in your opinion?
As you have said, data is an assets from which you want to extract value. If you have a lot of data, e.g. from FMC goods, you can afford some errors as the amount of data filters them out to some extent. However, the solutions we provide are more complicated products, so only one evert six months is produced. In this situation, data quality becomes really important. For data quality to be in order, data governance needs to be in order in turn. It is a time-consuming and boring process to implement it, but it is crucial to realise good governance before doing anything else.

In hindsight, what have been the most important changes?

- Implementing a common language throughout the industry (by means of the data dictionary and service catalogue);
- Implementing data quality checks. KPMG delivered a first dashboard, and M3 has continued to improve this and build on this ever since. Every month some new checks are created.

Did the changes you implemented contribute to differentiating yourself from competition? What are the benefits the whole process led to?
Differentiating was never the goal. However, recently a new department (digital) was created. These people are concerned with developing products based on data. So in that sense, data is more of an enabler for creating new smart products, which in turn make you more competitive.

So there have not been direct fruits of your labour, such as lower production costs?
Through reporting some processes have become more efficient, you can for example make a spend analysis. Overall, it's more the collection of little things that will have a large effect in the end. It would be nicer if you had big wins, e.g. "we implemented this and this has gained use €2 million." – that would quickly open the eyes of other stakeholders. However in reality it is more a matter of creating insights into processes, which in turn can have positive effects on the business. It is really difficult to quantify the consequences of data implementations.

Based on your "big wins" remark: is C-level involved? Do they give you the room to develop?
Yes, the process of making M3 more uniform (oneM3), is a C-level programme. They are all sponsors of implementing data throughout the business, so the CEO is willing to collaborate. However, for data projects it can sometimes be difficult to attract top management's attention; to create awareness. Investments for these projects are quite low, and the CEO only gets involved when there is a certain amount of money involved. So on the base of funding, he is not involved a lot.

Have you ever experienced difficulties in the collaboration between IT and business stakeholders?
Not really. The reason is twofold:

- We work in the same small office, so were are in close personal contact with each other.
- All our change projects are SCRUM, so both of these people are in the same team. Of course people think in different ways, but by having one team, you force them to think together and come to solutions.

A single point of critique the interviewee points out, is the fact that business stakeholders often think new projects get off the ground too slowly. When they think of something that may work, they want it implemented the next day. However, IT stakeholders realise there is much more playing in the background, which makes implementation difficult.

Where would you put yourself on the DMM maturity scale before and now (explanation of this scale was given)?
Quite difficult to do so, due to the different elements of data management. I would say we move between 2 and 4. It is important to realise that some processes are already quite data-driven, e.g. finance. So for these processes it will cost less energy to transfer from level 4 to 5 for example. For other processes that are less data-driven, this will be more difficult.

Governance and quality are the most important elements to us, which have to be in place before anything else. In these aspects we have really made good progress over time.

5. Capabilities

In this section, the interviewee's opinion on the capabilities defined in the case study was asked. Asked whether or not he found the capability relevant and if he would label it as core/supporting/specialised. These findings are specified in an Excel file. Some additional remarks are given below:

- "Carefully plan data projects"
  We use few actual data projects, it's more on a use case base. When you encounter something during the job (this can be at the office, at sea, et cetera), you are going to research how you can avoid this or make it more efficient. This seems like an unintelligent approach, but there are so many aspects to data that it often becomes impossible to define where changes are needed. So it are these use cases that will steer towards the projects that require our focus. Eventually, these use cases then become projects, making it more generally applicable.
- "Ensure expert skills and knowledge are available"
  For a small company this is not an easy task. Large corporations can have large teams with different specialisations. We have four analysts, however. So ideally you'd want them to be able to do everything, to be very multidisciplinary. It's not always easy creating a capable team.
- "Report the results of DQ monitoring"
  They want to add relevant DQ checks to the data objects in the workbase portal. Also, they want to specify the role of the stakeholder responsible for this data check so that it is easily known where to go when quality is lacking.
  Furthermore reports contain a lot of information now, the goal is to reduce this based on the stakeholder it is meant for. So to include only relevant data.
- "Define problem areas based on reports and follow-up"
  A goal is to eventually work towards monthly meetings functional area in which the following will be discussed: (i) make sure the service catalogue is up to date, (ii) define new data quality checks, (iii) define new KPIs.

Case study company F3

1. Introduction

In this section, both parties introduce themselves.

Interviewer elaborates on the following:

- Thesis topic and goals;
- Importance of interviews;
- Anonymisation of data;
- Whether or not the interviewee agrees to recording the interview. In this case, the interview was recorded.

Interviewee introduces himself and company F3:

- F3 is an insurance broker. They are the administrative middleman of large insurance companies, and they can sometimes act as an insurer with a mandate. F3 gets commission while the risk remains with the actual insurer. Furthermore the company creates portfolios to spread risk among different insurers.
- Interviewee's function is manager data management within F3.

2. Reasons for involving KPMG

In this section, the reasons for involving KPMG were discussed.

What were the main data-related challenges M3 faced leading to the involvement of KPMG?
F3 was about to acquire another company. This was a large acquisition; both companies were roughly the same size. This forced the company to rethink their architecture on different levels: the company itself (departments et cetera) and then processes. Processes are supported by systems, which in turn run on data (customer data in this case). To be efficient and to operate well, this data needs to be in order. Before the acquisition, data was given no real thought. KPMG came to perform a maturity assessment: where do both companies stand, who have best practices and how can those be used to design a new company? This scan resulted in a new team, the data management team.

Maturity levels were rather low, can you give some examples of deeper effects this had on the company?
Levels were indeed low, most actions were ad hoc. Some examples:

- Customer service did not always function well because of missing data.
- When data is not in order you cannot link turnover to products, so you don't know which products work well and which need improvement.
- Coupling working companies to their mother organisation can be problematic, which can lead to financial reports that are not properly executed.

3. Organisational changes

This section elaborates on the concrete changes implemented to realised M3's ambitions.

What concrete steps have been taken to accomplish your ambitions?

- Ad hoc activities have been transformed into processes; they are now executed in the same way with the same tooling. They have been standardised and are thus more efficient.
- Ownership was established. Now, if there's a problem you know who to go to solve it. It forces people to take responsibility and in that way enforces better execution of processes. Interviewee notes that data management is always guiding, not executing. If you remove the burden of cleaning the incorrect data from the operational layer, they will be even less motivated to do it right.

    - Interviewee notes that they work closely together with managing directors when it comes to ownership.

      MDs have quite a senior function, is collaborating with them easy?
      No. It is important to realise that F3's primary process (i.e. the lucrative one) is not data management. Data management is secondary and supports the primary one. When data is in order, things will go more smoothly. But this is not necessarily priority, so it can be difficult to be "heard". Something that helps is helping people realise the importance of data. Concrete examples help a lot in this.

Have you reached your goals? Do you have any further ambitions on the horizon?

- Extending the team; there are simply not enough FTEs, which causes some departments to still go without proper data management.
- Increasing levels of automation, so moving towards those high levels of data maturity. A large part of the complexity of data management is the staggering amount of data you are dealing with. You simply can't do everything manually. Interviewee notes that it is important to always incorporate some human checks.

How would you classify these efforts on a heatmap of impact vs effort?

- Low hanging fruit: governance, standardised processes, etc.

- High effort, high impact is for example implementing CRM.
    - Interviewee notes that a lot of customer data comes from a third party. This decreases their responsibility concerning data quality (it moves to the third party).

4. Reflection

This section reflects on the changes that have been implemented.

In hindsight, what have been the most important changes?

- Doing the maturity scan. This really showed the low levels of F3, which was an eyeopener for management, which in turn led to their sponsorship for setting up data initiatives. This is something F3 keeps building on.
- Governance is vital, if you don't have this you have nowhere to go with your solutions.
- There is of course more, you need all elements of the circle in the end. But if you have this, you'll get there eventually. These two are really crucial.

Have you ever experienced difficulties in the collaboration between IT and business stakeholders?
Data management is always the responsibility of the business, but due to the fact that you have to incorporate automation, you will also need IT. Data management sort of sits on the border between business and IT. Something that you often encounter is that business does not always have the tech insights that allow them to judge about the developments of tools.

What would you say are the main differences between manufacturing (M) and financial services (FS) when it comes to data management?
FS is more connected to the government, as they are kept standing in times of crises. This comes with a lot of compliance and legislation. From these aspects there data management is already shaped: you need good DQ, lineages, etc. You need to be able to justify your actions to a supervisory body. For M this is less the case, so they can be much more creative with their data. For FS data management is more controlling, reporting focussed. For M you can use it for more interesting purposes beyond this, use it to innovate and reshape your business.
The common ground is of course the same, they have the same problems, lacking data, incorrect data, etc.

Appendix D – Validation of capabilities in interviews

Below it is specified which capabilities were validated in which interviews. For capabilities that were not verified, reasons are given. Note that these are coloured orange. The first table depicts the capabilities identified in the manufacturing industry, whereas the second table contains the ones derived from financial services.

Table D1. Manufacturing capabilities.

Manufacturing capabilities	Validated in
The ability to organise the company in a suitable way	M1
The ability to ensure effective decision-making for change initiatives	**Only present in M4**
The ability to promote the business rationale of high quality data to increase awareness and acquire sponsorship for change initiatives	M1, M3
The ability to align data initiatives with business objectives	M1, M3
The ability to actively involve data in new projects	M1, M3
The ability to create a data-friendly company culture	M1
The ability to facilitate the cross-pollination of good practices by supporting cross-functional collaboration	M1, M3
The ability to develop and maintain a data strategy, to document and communicate it throughout the organisation	M1
The ability to create a shared data language throughout the organisation by developing and communicating definitions, standards, policies, and rules	M1, M3
The ability to monitor and subsequently improve data management activities by measuring internalised pre-determined KPIs	M3
The ability to determine to optimal way in which data should be organised within the business	**Only present in M2 and M4**
The ability to ensure good communication and collaboration between data, business, and IT stakeholders	M1, M3
The ability to appoint the right central data governance bodies	M1, M3
The ability to communicate governance bodies' roles throughout the organisation	M3
The ability to recruit the right talent	M1, M3
The ability to monitor data quality based on pre-defined metrics and targets, to report on these findings in a standardised manner, and to continuously improve data	M1, M3
The ability to centralise and standardise data processes	M1, M3
The ability to embed controls in data processes	M3

Table D2. Financial services capabilities.

Financial services capabilities	Validated in
The ability to promote the business rationale of high quality data to increase awareness and acquire sponsorship for change initiatives	F3
The ability to align data initiatives with business objectives	**Only present in F1, F2, and F4**
The ability to create a data-friendly company culture	F3
The ability to facilitate the cross-pollination of good practices by supporting cross-functional collaboration	F3
The ability to organise the company in a suitable way	**Only present in F4 and F5**
The ability to define a roadmap with prioritised short- and long-term projects, depending on the organisation's ambitions	F3
The ability to develop and maintain a data strategy, to document and communicate it throughout the organisation	F3
The ability to create a shared data language throughout the organisation by developing and communicating definitions, standards, policies, and rules	**Only present in F1, F2, F4, and F5**
The ability to determine to optimal way in which data should be organised within the business	F3
The ability to centralise data management activities	F3
The ability to monitor and subsequently improve data management activities by measuring internalised pre-determined KPIs	F3
The ability to monitor data quality based on pre-defined metrics and targets, to report on these findings in a standardised manner, and to continuously improve data	F3
The ability to appoint the right central data governance bodies	F3
The ability to communicate governance bodies' roles throughout the organisation	F3

The ability to recruit the right talent	F3
The ability to centralise and standardise data processes	F3
The ability to develop and communicate frameworks for guiding data processes	**Only present in F1, F2, and F3**
The ability to stablish a central security policy aligned with regulatory requirements	**Only present in F1, F2, and F3**
The ability to establish a central data life cycle management process	F3
The ability to establish organisation-wide document management standards, regularly review these	F3