Aalto University
School of Science
Degree Programme in Information Networks

Jooel Friman

Data Utilization in the Enterprise
Case Study

Master’s Thesis
Espoo, June 2, 2019

Supervisor:  Professor Eila Järvenpää
Advisor:     Klaus Lindberg M.Sc. (Tech.)
Aalto University
School of Science
Degree Programme in Information Networks

ABSTRACT OF
MASTER’S THESIS

Author: Jooel Friman
Title: Data Utilization in the Enterprise
Date: June 2, 2019
Pages: vii + 94
Major: Information Networks
Code: SCI3047
Supervisor: Professor Eila Järvenpää
Advisor: Klaus Lindberg M.Sc. (Tech.)

Enterprises collect more data than ever. Still capturing value from these existing data masses can be difficult. Companies most valuable data, master data, can easily go underutilized. Development in computer science has enabled technologies for utilization data such as data mining, artificial intelligence and analysis tools. Speed of technological change gives possibilities for enterprises to utilize their data resources. Nevertheless, organizational and contextual barriers are the ones that hinder the use of master data and data in general.

The motivation for this thesis is to find a way to utilize enterprise master data sets for sustained competitive advantage. The literature review focuses on contextual data quality, data-driven organizational practices and structures and strategic opportunities that data provides for sustained competitive advantage. This review is a base from which the research framework is formulated. This research studies on how data can be used to build capabilities, to develop automated value increasing processes and to develop organizational learning and innovation.

The results suggest that to utilize data more effectively, enterprises should build boundary resources internally between business units to build a shared understanding of how data can be utilized. Enterprises have to dissolve contextual data quality barriers and build technological and learning capabilities to make data usable for knowledge workers. This thesis provides a framework for assessing contextual data quality, building data to knowledge resources and understanding differences between managers and knowledge workers. Even though research is done in one Finnish knowledge-intensive enterprise, many issues are shared between different organizational contexts. Using the model provided researcher, management and organizations can study utilization capabilities in contextual settings to gain sustained competitive advantage.

Keywords: data quality, master data, knowledge based view, boundary objects, data utilization, contextual barriers, strategy, enterprise architecture, enterprise

Language: English
**Tekijä:** Jooel Friman  
**Työn nimi:** Datamääritymisen yrityksessä  
**Päiväys:** 2. kesäkuuta 2019  
**Sivumäärä:** vii + 94  
**Päähäne:** Informaatioverkostot  
**Koodi:** SCI3047  
**Valvoja:** Professori Eila Järvenpää  
**Ohjaaja:** Diplomi-insinööri Klaus Lindberg


Tämän työn tarkoituksena on löytää tapoja, jolla yritykset voivat löytää tapoja hyödyntää niiden olemassa olevaa ja ydintietoa. Kirjallisuuskatsaus keskittyy kontekstualaiseen datan laatuun, dataohjautumaa, organisaation toimintoihin ja rakenteisiin, sekä strategisiin mahdollisuuksiin, jotka tarjoavat pysyvää liiketoiminta. Tämän katsauksen pohjalta luotiz viiteeksi, joka ohjasi tutkimusta. Tutkimus selvitti miten dataa voi käyttää kyvykkyyksien rakentamiseen, arvoa lisäävän automatisoitun prossesseihin ja organisaation oppimiseen sekä innovointiin.


**Asiasanat:** datan laatu, ydintieto, tietoperustustuinen näkökulma, rajaobjektit, datan hyödyntäminen, kontekstualiset esteet, strategia, yritysarkkitehtuuri

**Kieli:** Englanti
Acknowledgements

To Mom

Espoo, June 2, 2019

Jooel Friman
# Abbreviations and Acronyms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>Artificial Intelligence</td>
</tr>
<tr>
<td>API</td>
<td>Application Programming Interface</td>
</tr>
<tr>
<td>BU</td>
<td>Business Unit</td>
</tr>
<tr>
<td>CRM</td>
<td>Customer Relationship Management</td>
</tr>
<tr>
<td>DIKW</td>
<td>Data-Information-Knowledge-Wisdom hierarchy</td>
</tr>
<tr>
<td>EA</td>
<td>Enterprise Architecture</td>
</tr>
<tr>
<td>EBMgt</td>
<td>Evidence-based Management</td>
</tr>
<tr>
<td>ETL</td>
<td>Extract Transform Load</td>
</tr>
<tr>
<td>IS</td>
<td>Information System</td>
</tr>
<tr>
<td>IT</td>
<td>Information Technology</td>
</tr>
<tr>
<td>KBV</td>
<td>Knowledge Based View</td>
</tr>
<tr>
<td>MD</td>
<td>Master Data</td>
</tr>
<tr>
<td>RBT</td>
<td>Resource Based Theory</td>
</tr>
<tr>
<td>RQ</td>
<td>Research Question</td>
</tr>
<tr>
<td>RW</td>
<td>Real World</td>
</tr>
<tr>
<td>VRIN</td>
<td>Valuable, Rare, Inimitable and Non-substitutable</td>
</tr>
</tbody>
</table>
Contents

Abbreviations and Acronyms v

1 Introduction 1
1.1 Background and Motivation 1
1.2 Problem Statement 2
1.3 Structure of the thesis 2

2 Theoretical Framework 4
2.1 Knowledge Based View of the Firm 4
2.2 Data - Information - Knowledge - Wisdom 6
2.3 Data Quality in the Enterprise Context 7
2.3.1 What Is Data in Organization? 7
2.3.2 Dimensions of Data Quality 10
2.4 Getting Value from Data 11
2.4.1 Data Is the New Oil 11
2.4.2 Cost of Poor Data 14
2.4.3 Data as Material for Value Creation 15
2.5 Data Driven Management 16
2.5.1 Call for Data Governance 16
2.5.2 Creating Evidence-based Culture 19
2.5.3 Competing on Utilization 21
2.6 Research Framework 23

3 Methods 25
3.1 Research Approach 25
3.2 Case Organization 27
3.3 Informants 28
3.4 Data Collection 30
3.5 Data Analysis 35
4 Results
4.1 Master Data from Enterprise Standpoint ................. 41
4.2 Master Data from Individuals Standpoint ............... 43
4.3 Data Quality Barriers Depend on the Function .......... 45
4.4 Culture for Sharing Knowledge is Needed ............... 48
4.5 Information Asymmetries Leads to Different Ways of Understanding ........................................ 50
4.6 Lack of Capabilities Hinders the Use of the Data .... 52
4.7 Summary of the Results .................................. 55

5 Discussion
5.1 Answers to the Research Questions ...................... 59
  5.1.1 Complexity of the Context Predicts Utilization Difficulties ........................................ 59
  5.1.2 To Understand Master Data, Understand the Organization Management Model .................. 63
  5.1.3 Create Boundary Resources to Facilitate Data Utilization ........................................ 69
5.2 Theoretical Implications and Suggestion for Further Research ............................................ 73
5.3 Practical Implications ..................................... 74
5.4 Limitations .................................................. 77

6 Conclusions .................................................. 80

A Interview questions ......................................... 91
Chapter 1

Introduction

1.1 Background and Motivation

The world is getting more fast phased due the accelerating change [Kotter, 2014]. One key driver in the accelerating world is the vast and increasing amount of data available. Data has become the most valuable resource for the organizations, but there are great utilization difficulties in firms [Erevelles et al., 2016; Mithas et al., 2013]. At the same time, as organizations gather more granulate data than before, capabilities to analyze the data to information and further to knowledge are becoming competitive advantages. In this new business landscape that enterprises focus on it is crucial to understand what kind of competitive advantages and challenges availability of valuable data brings.

Enterprises had used data typically to improve internal processes, products and services. Data is becoming more and more glue in between services and products [Porter and Heppelmann, 2014] as well as internal processes that enterprise needs to adapt to [Porter and Heppelmann, 2015]. This creates pressure to learn and utilize learnings effectively across the organization to shorten the development cycles [Porter and Heppelmann, 2015; Ries, 2017]. To have shorter development cycles, enterprises need also change its organizational structure and have a shared understanding of what different data means. If the enterprise does not have this data-driven culture, it can create confusions inside the organization. Advances in computer science and information systems are also enabling organizations to automate low variability knowledge work, which changes the enterprise’s ways of working.

Understanding how to use explicit knowledge resources that the organization poses are not just IT related issues. Using data and information to build organizational capabilities and to make decision strategically is a source
of competitive advantage [Choo, 1996]. There is a lack of research on what contextual things exist on developing understanding, processes and decisions from data. In this thesis, we assume that data can be used in three strategic ways: to build capabilities, to develop automated value increasing processes and to learn to make better decisions. Having this frame, we studied the most valuable data sets of case organization to determine what contextual barriers exist and how management can overcome these.

1.2 Problem Statement

This thesis aims to enhance data utilization in a large company context. In this thesis data utilization is defined as organizational use of data to create useful and effective practises that captures value from data. As the company being knowledge intensive, the role of data management, information handling and management of knowledge plays an integral role in business processes. The goal of this thesis is to find the boundaries of data utilization and ways to increase the value of existing data in the organization context. Therefore the problem statement consists of three research problems as follows:

- **RQ1** What contextual barriers exist in the utilization of data in the enterprise?
- **RQ2** How the value of master data differs among management and experts?
- **RQ3** How management can leverage data for competitive advantage?

This thesis focuses on the management of the company most valuable data sources that are metrics used to steer company such as financial data, customer data and asset data. Therefore the only the management side of enterprise is under review.

1.3 Structure of the thesis

This thesis is organized into seven different parts: introduction, theoretical framework, background, methods, results, discussions and conclusion. The first part highlights the background and motivation for the research and introduces the research questions.
CHAPTER 1. INTRODUCTION

Next part discusses on theoretical framework on which the research is based on. It introduces the knowledge-based view of the firm and how the theory describes sources of competitive advantage. Second sub-chapter introduces the concept of adding value and understanding into data in data-information-knowledge-wisdom hierarchy (DIKW). This sub-chapter also defines data and create a structure of why data should be developed. This part also covers the literature background, that firstly explores what type of data organization has and how it is used and stored. It also introduces features that are common with enterprise data and the concept of master data. Also, the different dimensions of data quality are reviewed and argued why contextual quality is essential in enterprises. Secondly, value capturing and processes that enterprises use to utilize data are explored. It shows that in the enterprise context, data processes are complex and technical. In this section also, the cost of data and business cases for data-driven developments are a review shortly. Lastly, the literature review focuses on organizational management in the data-driven context in forms of governance practices, evidence-based management and decision making. The background is used in the creation of the research framework that is introduced in the last chapter.

The third part of the thesis describes the methods used in the research. In this part, the research approach and research processes are discussed. After the research approach, description of the case organization, informants and contextual factors that were present in the organization. To have clear understanding on how the research was conducted, the data collection methods are described in the detail as well as the way of doing the data analysis.

Fourth part introduces the results and fifth part synthesis the findings in accordance with the research questions. Discussion chapter also gives implications for further research, suggestions for practical implications and shows the limitations of the research. Lastly, the conclusion concludes the thesis with closing remarks.
Chapter 2

Theoretical Framework

2.1 Knowledge Based View of the Firm

Views, where resources are seen as the source of the firm's success are dominated management literature in past decades. In these views, resources are seen as assets or things, that can be used in the production of services or products. A key aspect of resource-based theories (RBT), is the assumption that firms resources are the source of sustained competitive advantage. Competitive advantage creates value in the market that competitors could not emulate, so creating a barrier based on the firm’s positioning [Wernerfelt, 1984]. To gain sustained competitive advantage, the firm needs to have resources such as knowledge and capabilities that are not readily available in the market [Barney, 1991]. According to this assumption, firms competitive advantage does not change if there are no significant market changes. This creates significant entry barriers for stable markets, as the resources are already acquired by existing firms in the market.

A resource is commonly seen creating sustainable competitive advantage if it fulfils VRIN-attributes (Valuable, Rare, Inimitable, Non-substitutable). [Wernerfelt, 1984; Barney, 1991; Hoskisson et al., 1999] Resources are valuable if a firm can leverage them to be more effective or efficient in implementing the chosen strategy [Barney, 1991]. Barney [1991] even argues that resources have to be valuable as even to be considered as a resource. Value, therefore, comes from the environment, where the firm operates and from the opportunities that it can seize [Hoskisson et al., 1999]. Rarity is an embedded feature in all resource-based theories. It emphasis the notion of imperfect markets, where the resource will not and cannot be homogeneously distributed [Wernerfelt, 1984]. Inimitable means that there are costs asso-

\footnote{In management literature firm, has been used for companies and enterprises.}
CHAPTER 2. THEORETICAL FRAMEWORK

Associated with acquiring resources. Typically it is linked to capability building as processes, ways of working and information flows are hard to imitate. Finally, the framework notices the role of substitutes. Resources should not have substitutes, or it should have network effects with compliments.

This means that a resource can create service or product with a unique value proposition.

Figure 2.1: Resource Based Model of Competitive Advantage [Mata et al., 1995, p. 494]

Mata et al. [1995] summaries this resource based model according to three criteria (see Figure 2.1). The first criteria are that the resource has to be valuable to gain an advantage. If the resource is not valuable, it can be a liability that will consume other resources, creating a disadvantage to the firm. The second criteria argue that only some firms can have the resources in the market. If resources are homogeneously distributed, it will result in competitive parity. Last criteria evaluates both inimitableness and substitutability. This means that if enterprises without the resource can acquire the resource and capability to use the resource without great disadvantages, the resource is therefore mobile. Even thought many IS system can be acquired from markets the capability can differ significantly across companies.
For example, a business process can leverage capabilities that only one firm poses and it is highly embedded into the organizational culture. In general, resources can be categorized into two groups, tangible and intangible, based on the nature of the resource. Tangible resources are typically physical assets such as factories, machinery and inventory. Firms brand, used technologies, as well as human capital, are excellent illustrations of intangible resources. [Russo and Fouts, 1997] There is also a growing body of literature that recognizes the importance of intangible resources as the source of the firm’s competitive advantage. For example, Grant [1996] emphasizes the role of knowledge in organizational learning, usage of technology and managerial initiatives. In the knowledge-based view (KVB), knowledge is not seen as an easily deployable static resource. This view expands RBT and argues that knowledge creates firms core capabilities. Knowledge is contextual, dynamic and related to human beings [Nonaka et al., 2000]. KVB argues that firms are knowledge creating entities that configure, organize and utilize knowledge. Therefore, capabilities and knowledge are the main sources of sustainable competitive advantage [Nahapiet and Ghoshal, 1998; Nonaka et al., 2000; Hoskisson et al., 1999].

2.2 Data - Information - Knowledge - Wisdom

Data-information-knowledge-wisdom hierarchy is a widely used representation of knowledge stages of understanding. This model works as a guiding principle in information management, information system design and knowledge management [Rowley, 2007]. Hierarchy implies that every time a border is completed to a new state, higher understanding and meaning is achieved.

Data is a representation of the real world (RW) objects and events in the sensed environment. Information is data in context 2. Ackoff [1989] points out that difference in data and information is the function. When data can be described it turns to information as it adds value to understanding and answers questions [Rowley, 2007]. Information can also be transferred without loss of purity or meaning [Kogut and Zander, 1992, p.386]. This requires understanding the description method (e.g. technology) used. Knowledge is knowing how to something is done (know-how) and the ability to transfer information to instructions. According to Rowley [2007] knowledge can be acquired from instructions, experience or learning from others. On the top of

2In this thesis, information and data are used interchangeably as the focus of the inquiry is the enterprise context
the hierarchy is wisdom, that is the ability to increase effectiveness (see Figure 2.2). Wisdom requires judgment as it evaluates why something is valuable. Zeleny [1987] argues that wisdom is the skill to know why something happens or why something should happen in future. On the other hand, effectiveness means the ability to achieve the desired objective. In the enterprise setting, this might be the strategy. According to Rowley [2007], intelligence is the ability to increase efficiency and therefore ability how things should be done. Intelligence does not require learning per se, it requires measuring and analyzing how status quo can be preferred. In that dimension, intelligence can be transmitted from knowledge easier than wisdom. Intelligence is inferior to wisdom in its value.[Rowley, 2007]

![Diagram showing the DIKW hierarchy]

**Figure 2.2: The DIKW hierarchy [Rowley, 2007, p.164]**

In general, the value of understanding increases from bottom to top, from data to wisdom. At the same time, the contextual aspect deepens and individualizes. This affects the representation of knowledge. Data can be represented as it is. It can be structured and formatted. Data can be mined for insights with algorithms and computer programs.

### 2.3 Data Quality in the Enterprise Context

#### 2.3.1 What Is Data in Organization?

Defining what data is not a trivial task. It is meaning can differ in by structure, organization and usage. Understanding ontological and structural views of data is necessary to understand the underlying boundaries of data-driven
value creation. In the applications of computer science, the term ontology has the meaning of a standardized technological framework in which information is stored and organized [Panov et al., 2009]. From this standpoint, data can be seen as attributes modelling Real World (RW) system or an object that is stored in some information system (IS) [Wand and Wang, 1996]. Designing the IS depends on the user’s perception of the RW system. Therefore deficiencies in IS design will lead to skewed data from the RW object.

In this view data is a set of variables that have coded qualitative or quantitative values. Wand and Wang [1996, p. 88] makes a clear distinction between direct observation made by the user and pre-designed representation that is added to the information system. Data added to the information system is typically quantitative as it needs to be machine readable. As user processes their direct responses the data is usually qualitative by nature if it does not have designed codification to form quantitative structure. As the values are coded they form a structure. Level of data structure is typically grouped in three class of growing structure: Unstructured, Semi-Structured and Structured as shown in Table 2.1.

Unstructured data does not have a clear structure or model. It includes images, video and text files that could not be stored in relational database systems [Blumberg and Atre, 2003]. According Blumberg and Atre [2003] most enterprise data is unstructured. In the future, nearly 80 per cent of enterprise data will be unstructured and the velocity of data doubling in every 18 months [Sap, 2012] For example, it can be stored in emails, photos and slide shows that do not easily searchable and contains little meta-data. This makes unstructured data unsearchable and hard to link into the enterprise context. Semi-structured data usually leverages unstructured data. It has the same form, but it grouped by some similarity. For example, customer emails that are collected to the enterprise’s customer relationship management (CRM) system are in a semi-structured format. Semi-structure data is the end product of many data mining techniques. Even though it could not be stored in a relational database, it can be mapped with similar data to form data sets, such as document libraries and management systems. Semi-structured data binds together qualitative and quantitative aspects of data and is, therefore, a right place for semantic searches.

Structured data is stored in the IS as an object having attributes. It is in explicit form and it has a schema. Structured data can also be used in automated business processes because it is machine readable and can be stored to relational databases. It is typically created by sensors, computers, logs, monitoring and automated processes. As structured data measures changes in the RW system, it can be used in forecasting and in time series analysis.
CHAPTER 2. THEORETICAL FRAMEWORK

<table>
<thead>
<tr>
<th></th>
<th>Unstructured data</th>
<th>Semi-structured data</th>
<th>Structured data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage of the Data</td>
<td>Documents, presentations, photos and emails</td>
<td>Document collections, Wiki sites and</td>
<td>Information Systems and Databases</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Repositories</td>
<td></td>
</tr>
<tr>
<td>Nature of the Data</td>
<td>Qualitative</td>
<td>Mixed</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Use of the Data</td>
<td>Learning and Communication</td>
<td>Information Search</td>
<td>Measuring, Automation and Forecasting</td>
</tr>
</tbody>
</table>

Table 2.1: Data Structures and Usage in Enterprise Context

Enterprise data should be connected to business processes and goals. Goals and processes define how data should be managed to align enterprise with its strategic goals. In an investigation into information transform, Vayghan et al. [2007] found that most enterprise data can be categorized into three main groups: master data, transaction data and historical data.

Otto and Schmidt [2010] defines master data as a set of business object characteristics, that are common throughout an enterprise. Objects can be the things, people and the places that are involved in the enterprise’s business processes. This view is supported by Vayghan et al. [2007] who writes that master data includes a collection of core enterprise’s data entities such as customers, offering and employment data. This type of data does have low change frequency and it is nonvolatile (see Table 2.2). Master data objects are typically created only once but used multiple times [Haug et al., 2011; Haug and Arlbjørn, 2011]. These objects are independent because they represent main business objects, but are linked by all business transactions. Transactions data represents events in the company such as invoices, orders, payments and records [Haug and Arlbjørn, 2011]. Transaction data describes enterprise operational events and is linked to master data and historical data. Historical data is a collection of logs and past events that are linked to transactions and master data. This includes object states between the time series. It records changes in the enterprise and is the source for advanced analytics and decision making.

What data really is, a representation of RW in a codified form. It can have different structures and its usage differs based on the data type. In the enterprise context, data can also be categorized by meaning for the business, state and independence.
### Table 2.2: Enterprise Data Categorization (based on Otto and Schmidt [2010])

<table>
<thead>
<tr>
<th>Data Category</th>
<th>Velocity</th>
<th>Variety</th>
<th>Existential independence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Master Data</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Transaction Data</td>
<td>High</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Historical Data</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
</tr>
</tbody>
</table>

#### 2.3.2 Dimensions of Data Quality

Good quality of data is necessary for enterprises to operate. Still, there are different needs from the data between different enterprise functions. Wand and Wang [1996, p.86] proposes, that as the quality of the product is dependent on the design and production process, so the quality of data depends on design and production processes of data generation. Therefore problems in data quality come from flaws in design or business operations. However, this view focuses only on internal issues in data. It does not address problems regarding data usage. Problems in data quality can be categorized by design problems and operation issues. Wand and Wang [1996, p.91] identifies incomplete representation, ambiguous representation, and meaningless states as the major causes of design deficiencies.

Incomplete representation means that one real-world entity is not adequately represented in the information system. For example, the name field of an entity does not accept characters that are not used in English. In other words, this means that all lawful states are not mapped to the information system. According to Wand and Wang [1996] ambiguous representation is where two or more RW states maps to same IS state. Having attributes in the same form, but being different states such as home and work phone numbers illustrates this point clearly. In an ambiguous situation, many RW state maps to one IS state. Meaningless states is a state in an information system that does not have lawful RW state. This is common in enterprises where information and data are widely dispersed [Vayghan et al., 2007].

Wand and Wang [1996] argues that operational deficiencies happen when RW state is mapped to the wrong IS state. RW state can either be mapped to meaningless IS state or mapped to a wrong state. In the first case mapping from IS back to RW is not possible. In the second case, user can map from
IS back to RW, but it leads to a wrong RW state. According to Wand and Wang [1996] data has four intrinsic properties that good quality data has to suffice. Data has to be complete, unambiguous, meaningful and correct. DeLone and McLean [1992] argues that good quality data has also to be free from bias as innate property. Freedom of bias is different from previous properties in respects that it emphasizes the independence of the contextual aspect.

In enterprise context problems arise when business processes are not aligned with the design of IS. Even half of the knowledge workers time is used searching, identifying, correcting and validating data [Redman, 2013]. Strong et al. [1997] argues that to solve complex organizational data quality problems more holistic approach for data quality is needed. They argue that data quality depends on the context in which the data is produced, rather than only problems in design and operations. Watts et al. [2009] suggests that data quality dimensions should be studied in a context where data is used.

Not using data for analysis, automation and decision support are a big problem for organizations. Strong et al. [1997] argues that even data that has intrinsic properties right, might not be used as data consumer eventually decides whether or not to use data. As organizational structure can affect the completeness of data. An organization can have multiple sources of the same data. Customer does not know what the master, the most valuable source, of the data is. Also in a situation where data production method is subjective consumer can question the objectivity of the data. These two factors affect the believability of the data [Strong et al., 1997]. If believability causes problems, trust in organization’s applications diminishes. When an organization’s data customers stop using the application the network effect decrease the added value of the application. When the customer knows that others are using different application she switches to the one she is most comfortable with if possible. In the worst case situation, good IS design is not used, because the organization could not standardize data production or centralize data sources.

2.4 Getting Value from Data

2.4.1 Data Is the New Oil

As digital and information products are coming more commonplace in the organization context, the amount of data available also increases. Many tasks are handled in the digital environment by a laptop, smartphone or a tablet.
Employees need to be connected to systems to do their work and update IS. This flow of information needs to be managed to fuel organization growth and value creation. Data has become the fuel of decision support systems, software robots and marketing systems to name a few. This chapter focuses on how existing data should be processed and what kind of business and technical implications it has.

Like oil data also needs to be processed to fit the selected purpose. Enterprises face issues in data capturing and data utilization. Data capturing is a crucial part of the data pipeline as capturing method and technologies affects the business model. Data is usually an outcome of a business process, for example, customer acquisition will be stored in customer registry for later use and used in future when needed. When an organization has different processes and systems the data has to be consolidated to be utilized. Typically enterprises have their own operational databases, but the role of external data sources has grown as enterprise-level cloud infrastructure develops. To use this data it is commonly put into the ETL (Extraction, Transformation and Load) system (see Figure 2.3).

ETL consist of three phases: extraction, transformation and load, where the name ETL comes from. In extraction phase data is consolidated from different sources to one location. Data should be cleaned and tested at the end of the extraction process to ensure data quality. After extraction, data has to be transformed and calculated for new tables and database structures. Transformation can be costly to compute if data is bad quality, it is in huge quantity or transformation is computing heavy. In load process transformation data is stored to be used in data marts and application programming interfaces (API). End-user can query data programmatically or via user interface. According to Kimball and Caserta [2011], data marts should be based on a single business process such as customer analytics or sales modelling. If multiple sources of data are needed, end-user queries them from multiple data marts. The core idea of ETL and ETL like systems is to make data highly available.

What kind of process do enterprises use for data flows out of the scope of this thesis. Still, it is essential to realize the logical implications what end-user will face. The selected process defines the meta-data of data marts data. It also affects reliability, timeliness and correctness of the data. Data warehouses also work as a source of shared "truth" as the data is validated as a company asset. Therefore enterprises need to standardize, document and enforce data flows with company practices. Data models should be aligned with the enterprise business model. Traditionally this has lead from IT alignment, but as the company grows and matures in size, making IT system decision in business units can be harmful. According to Ross et al. [2006] coordinating
applications, systems and infrastructure in enterprise level will cut IT costs, but also make data more accessible. They found that companies that had digitized their main processes have 25 per cent lower IT costs and higher management satisfaction.

A primary concern of IT alignment is how to shorten the business planning and implementation cap. This alignment addresses both effectiveness and efficiency [Luftman, 2000]. In other words, is the enterprise doing the right things and doing things the right way. Typically companies try to solve this problem via Enterprise Architecture (EA) process. Ross et al. [2006, p. 47] defines EA as "the organizing logic for business processes and IT infrastructure reflecting the integration and standardization requirements of the company’s operating model". EA should create shared meaning of long term operations practices that enterprise uses. EA combines business, processes, integrations, software and technology architectures as an aggregation of hierarchies [Winter and Fischer, 2006]. Many of these fields have traditionally been business or IT tasks and not overseen by the whole organization. This might be the reason why, most companies have limited understanding between IT and business. Even though IT is emerging as process enabler, it is seen as part of cost efficiency rather than part of effectiveness. [Luftman, 2000, ]

The EA process is hard to maintain because many changes in enterprise happen in business units. If changes are governed by the enterprise business units will have less flexibility even though enterprise-wide flexibility increases [Ross et al., 2006]. Enterprises have to make strategic choices about how they define the operational model and how new processes and systems are implemented.
2.4.2 Cost of Poor Data

Cost of poor data is hard to measure and identify, but it can have significant negative impacts on the organization. Data management is not just system integrations and automated tests. More commonly it is humans updating existing records that have the most value and need contextual understanding. If the data is not volatile (see Table 2.2) as enterprise master data, changes will have big impacts in process or even business quality. For example, customer invoicing information is crucial data. If information changes and it is not updated to the system, there will be a delay in invoicing process or worst case invoice will go unpaid or automated system works imperfectly. It has direct operational cost to the enterprise. This kind of errors in master data will lead to multiple errors in transactional data as transactions are based on master data. Errors in historical data, on the other hand, effect on strategic decisions, that are based on past experience.

Haug et al. [2011] highlight the need for labeling operational and strategic as different categories of data related costs. They also suggest that cost should also be labeled if they are direct or hidden. As the same data is used in various tasks, the quality for secondary uses might fall when the quality is improved from the primary viewpoint. This would be hidden cost as the cost is not foreseen in the planning phase and realizes only as second order effect. For example, enterprise financing unit might use financial data to plan future investments. If they change the way database attributes are calculated, it might not change the way how management makes decisions based on new reports as they lack a deeper understanding of changes in the context. This type of changes can even compromise decision making in managerial level [Redman, 1998, p.81]. Direct cost is mistakes in primary usage. From the historical standpoint, some functions have valuable and business critical data that they own [Vayghan et al., 2007]. In many cases, they are not the ones that use and creates the most value for the data. The concentration of data to one unit can have hazardous consequences if the cost of increasing data quality to secondary solutions is high.

In the enterprise context, cleaning poorly managed data can be a tremendous cost. An enterprise might not have capabilities for re-engineer existing data, data production and data flows. Typically poor data is also challenging to implement into new systems and warehouses [Redman, 1998]. If benefits for cleaning the data are hidden, such as second-order effects, getting involvement and funding from other stakeholders can be hard. Data cleaning should be done from the business value point of view. Lavalle et al. [2011] found that lack of understanding of how to use analytics and data to improve business was the leading obstacle for data-driven adoption in their study. Data clean-
ing can be seen as an initiative where quality is increased to the desired level. Companies should start small and narrow scope to get value in comparison to big multi-year projects.

In contrast, as data quality increases enterprise faces maintenance costs on top of data quality improvement costs. Costs occur from preventing, detecting and repairing flawed data, but also assuring data quality in the entry, processing and in use phases [Eppler and Helfert, 2004]. As the enterprise has matured its data quality it should focus on optimization of the data quality costs and quality trade-off. Data quality initiatives lower the process cost, but also gives enterprise ability to pursue multiple opportunities as good quality data can be utilized in multiple flows. Still the cost of poor data quality is highly context-dependent compared to the expenses of data quality initiatives [Batini et al., 2009]. Therefore enterprises should adopt risk tolerance and portfolio management as a part of their data management process.

The most significant costs that the enterprise will have for poor data is the opportunity costs. In many digital markets, the winner will take most of the market share [McKinsey, 2016]. In networked markets, few platform-based businesses emerge as the winner, taking practically all of the market [Eisenmann et al., 2006]. If data and utilization capabilities are not acquired as platform-based business models take markets, the enterprise can lose the market for good. Context-dependent and good quality data is the material of value creation in digital products and services. To compete for the long run, enterprises have to build data-driven resources.

### 2.4.3 Data as Material for Value Creation

Data can create value in an enterprise by cutting costs, creating new business opportunities, monetization and enhancing decision making and learning. Porter and Heppelmann [2014] argues that data-driven technologies will enable new smarter products that automate processes. In the enterprise, this could mean, for example, software robots that schedules meetings, automate billing to right cost object and so forth.

Automation brings enormous cost savings but takes time. According to Porter and Heppelmann [2014] companies should start gathering data from their products via sensor and external sources. After data gathering, software and connection could be added to first control and afterwards to optimize the performance and output. In their model, the product autonomy will lead that operators watch over the system rather than individual products. To achieve automation, data needs to be gathered, analyzed carefully and utilized with precision. The higher the level of automation, the level of data
quality also needs to be higher and more context dependent. Probably not all data quality dimensions are known in advance, so redesign and quality control needs to be continuous.

Data has tremendous value as a source of knowledge creation [MIT Technology Review Custom, 2016]. Enterprises can use data and information to get insights with mining the data, validating business opportunities with analytics or learning new skills. If a firm can map what brings the most value to their customers, they can focus on the core capabilities needed to delivering value. Data should be used to analyzing what the organization is doing at the moment, how effective it is producing value.

In their seminal article, Prahalad and Hamel [1990] argues that corporations should focus on building their core competencies. They define core competencies as "the collective learning in the organization, especially how to coordinate diverse production skills and integrate multiple streams of technologies". In growingly knowledge-based work doing the right things rather than doing things right is a source of competitive advantage. Gathering information and data from RW typically is not the problem. Competencies for communicating the findings and realizing the importance of what data provides to the whole organization can be difficult. For example, Dominic and David [2012] points out that companies typically have the data needed for learning, but managers do not know how to information could be used. Also, it is necessary to acknowledge that, core capabilities can fade if not used [Prahalad and Hamel, 1990, p.81]. On the other hand, as they are used, they are getting better as the organization learns and do not tear in use. This is the same for every knowledge related activity is that then data, information, knowledge or wisdom.

For most companies, data is the single biggest asset, and it tends to create even more data when used [MIT Technology Review Custom, 2016]. This emphasizes the need to develop competencies for analyzing and using data. According to Brynjolfsson et al. [2011], companies that use and relies on data in their decision making performs best in terms of performance and productivity.

### 2.5 Data Driven Management

#### 2.5.1 Call for Data Governance

Data governance can be simplified as a set of processes that ensures the enterprise data quality [Sarsfield, 2009]. Data governance is also the enterprise-wide ways of working that implies who makes the decision and how. Vayghan
et al. [2007] points out that data should be governed as an enterprise asset. This includes implementation of policies, standards and guidelines.

First steps in effective enterprise data governance are to have understood what data quality means in the enterprise context and what kind of elements the most valuable data has such as volume, variety, veracity and velocity. Enterprise also has to categorize what data is the most valuable to its current business (Master Data) and what are the secondary use cases that will create new value in the future. This kind of operational planning needs competencies for developing utilization of data sources, but the in-depth perception of business trends, opportunities and threats related to data are more critical. Therefore data governance should be mainly business-driven as the role of quality is enforced by business processes. In essence, in knowledge-based companies, data can work as an operational asset but mostly as a strategic asset for future operations and products.

Data governance has similarities with IT governance as data is typically stored in IS. As many enterprise functions might use the same system for accessing the data but need different things from the data quality. They have varying qualifications for data governance and have different needs in tools and access. This creates pressure both strategic and system level planning and design. Wende [2007] suggest that also organizational aspects such roles, policies and issue resolving plays a crucial part in ineffective data governance. Enterprises should build business process infrastructure on top of the IT infrastructure. This means that as big organizations have legacy systems that cannot be transformed into new functionalities easily, strategic planning of IS is needed for developing functionalities for new use cases.

One of the main tasks of governance is to ensure that operations have the needed resources. Endpoints and access to data and information should be selected carefully to align the organization towards its goals. As data is abundant resource, but hard to manage, the governance should define which data is gathered to enable ongoing and future processes. For example, management can decide that customer satisfaction data needs to measure in all services. In ideal case customer data owner would create boundary objects, such as APIs and forms, to make it easy to integrate this data flow to the CRM system.

Boundary objects create shared syntax or language how data, information and knowledge is represented [Carlile, 2002]. Boundary objects should be embedded in the organization culture. They work as a common starting ground for all employees in the enterprise context [Fischer and Ostwald, 2001]. Useful boundary objects define the way of work on how individuals in the organization can cooperate and share knowledge. This enables the internal innovation and management can create policies on how open the enterprise
ecosystem is by creating sufficient boundary objects. Boundary objects in governance space are policies, role definitions that direct responsibility and rights in the enterprise. According Carlile [2002] most effective boundary objects are "objects, models and maps". This means that knowledge hardly transfers from IS and other repositories such as wikis and shared network drives to the end-user. In these cases, syntactic knowledge is transferred, which is not easily implemented if the user is not familiar with the representation [Carlile, 2002]. Data-driven boundary objects should communicate how business needs match with data governance possibilities.

In essence data governance is matching enterprise business model with the enterprise data model. Data governance guides how data-driven processes are implemented into the business and how it should be designed. In Figure 2.4 is presented on how enterprise systems design effects on business model design from a data governance standpoint. Metadata and source standards are embedded in the IS.

Enterprise architecture combines and guides this to match the data needs and system level requirements of the organization. Data acquisition happens as a part of business processes. Finally, data utilizes other business processes where the quality and security needs are matched. As discussed earlier data

![Diagram](image)

Figure 2.4: Matching data governance in the continuum of Enterprise data model and business model

is highly context specific and depends on the use cases, as well the security needs from the data comes from the use cases. For example, highly identifiable and structured data has higher security standards than an unstructured non-critical document for example. The growing use of data also creates legislative pressure and in future companies have to look at data governance from the compliance standpoint.

In essence data governance is mitigating data related business risks, but it also has enormous upside when coupled with the business model redesign. Generating good quality data for new and valuable use cases transforms
operations and business processes. As enterprise faces need for changes in the business processes also the capabilities and needs for data management changes. These changes are becoming in in growing speed and enterprises need to continuously access how much internal change is required [Kotter, 2014]. Balancing with capabilities might focus overly on IS. Sikdar and Payyazhi [2014] argues that the overemphasis of information technology and underemphasis of business processes, communication and organizational dependency is one reason why business process redesign and changes fail. They also point out that most redesign programmes have not been integrated into corporate change strategies. Enterprises have not been able to define how data is to be utilized in business processes. This may be a consequence of data governance is not incorporated into enterprise strategy.

2.5.2 Creating Evidence-based Culture

Using research evidence in management has long roots [Briner et al., 2009]. The underlying assumption has been that management and practitioners can make a better decision by combining scientific research with a contextual understanding of the firm. Still there is evidence that managers typically do not use evidence while making their decisions and rather focus on their past experiences [Pfeffer and Sutton, 2006; Rousseau, 2006]. Briner et al. [2009] points out that evidence-based management (EBMgt) is not just implying academic research to the enterprise setting. It is about using different types of information in making decisions and plans.

EBMgt in the enterprise is combining management expertise and judgment with evidence such as data from local context and research evidence. This is important notation as competencies that enterprise hold as in managers expertise and judgment skills might be biased or not suitable in changing the environment. Managers should do continuous and systematic knowledge search outside their organization as there might have new research suggesting more suitable solutions. In simpler terms companies should seek knowledge outside their boundaries. These boundaries might even exist inside the firm. For example some employee might have needed skills and knowledge to make the decision, but there is no know-who information available. In the overall question of data utilization in decision making how much the decision is based on the expertise and how much analysis made from the corporate data. [Hansen et al., 1999] argues that organizations should either focus on codification or personalization in their knowledge management. Personalization means that knowledge is mainly stored in employees and into their skills. This focus emphasizes tacit knowledge and know-how. Codification can be defined as documenting learning to the system that stores,
spreads and allow reuse of the document.

In enterprise which focuses on personalization, the knowledge in-need is gathered into employees. Consequently, the decision should be made by persons that have the most knowledge, contextual understanding and implementation skills. Enterprises that focus on personalization needs to facilitate conversation and the exchange of tacit knowledge [Hansen et al., 1999]. From the data utilization perspective personalization highlights analytical skill at least to be effective in the domain perspective. Personalization is suited to situations when knowledge is not easily reusable or when the environment is highly dynamic and needs innovation [Greiner et al., 2007]. Codification spotlights the reuse of knowledge and relies more on information technology [Hansen et al., 1999]. This means that employees documents heavily and create boundary objects that are reusable. For example, sales teams could use other team’s sales bids for to save time and resources [Haas and Hansen, 2007]. Greiner et al. [2007] found that in situations where efficiency plays a critical role codification is the preferred knowledge management strategy.

To effectively use data as evidence, an enterprise should define a system of how decisions are made. W. Edwards Deming introduced the PDCA (Plan, Do, Check, Act) cycle in the 1950s that have become the cornerstone for continuous processes improvement [Moen and Norman, 2009]. The model emphasizes observation and adjustment of the process when new evidence emerges. It integrates the scientific method and incremental improvements. Ries [2011] modified the PDCA cycle to his Build-Measure-Learn feedback loop intending to build into concrete measurements that validate learning with data (see Figure 2.5. The loop starts with an idea or hypothesis that are tested with building a minimum viable product (MVP). The idea of containing business risks into small tests and prototypes has been existing long time.

Traditionally decisions are made by executive teams. Eisenhardt [1999] argues that executive teams should focus on multiple possibilities at the same time while planning solutions and systematically consider multiple future states. Combining the traditional approach with a hypothesis-driven approach, executives can work with experts and validate learnings. For example, if an enterprise is considering multiple strategic choices it could map visioned outcomes with different possible futures. The most valuable option is probably not the one with the highest Return on Investment (ROI), but the one that can be validated to bring the most value in the shortest possible time. Strategic planning takes typically two to four months of [Eisenhardt, 1999]. In that time frame organization should gather data, expertise and make the judgment what hypothesis are tested and are most likely bring most value as validated learning. Planning should be an iterative process
rather than a one-time project.

Top executives commitment to using data is crucial if making the change for the evidence-based organization. Change effort starts from management, and culture change is made by changing the way organization works [Kotter, 1995; Thomas H. Davenport, 2006]. Being able to change the organization to evidence-based management should start to implement and use data-driven practices. Eisenhardt [1999] gives a great example where top executives were responsible for gathering information in specific domains and reporting it. This highly personalized way enforces top executives to use existing data sources and to point out which are used as the source of "truth" in decision making. Using the most accurate data and analyses start from the management and should be an evolving process to match the changing environment.

2.5.3 Competing on Utilization

Enterprises work as a nexus of data. They collect data from their customer, their operations and their financial. As in the previous chapter outlined enterprises are using data to improve decision making. On the other hand data can be used for rapid implementation of new ideas, services and products to increase. The primary competitive advantage that effective data implementation brings is the reduced time to value [Lavalle et al., 2011]. The more valuable the data is, the more integrated it should be to the value chains. In short, data can be utilized in three categories: Capabilities, Automation and Decision support (see Table 2.3). This system of classification helps distinguish different aspects of categories.
Table 2.3: Sources of competitive advantage of data driven enterprise

Data can be a source of insights that should be turned into knowledge and further enterprise-wide capabilities. Insights should answer what has happened or what is happening, so the organization can adapt or learn from the change. Delen and Demirkan [2013] argues that descriptive business analysis helps define business opportunities and problems. This works as a stepping stone for building organizational capabilities and learning. Capabilities analysis focus on organization and looks at what procedures do not generate value or learning. Effective capabilities development facilitates knowledge sharing with techniques such as data visualization (boundary objects) and warehousing data to IS (codification).

Automation can be seen as a clear value generator for as cut transaction costs and standardize the service value. Automation-related data is prescriptive by nature but also has predictive and descriptive roles. The focus in data-driven automation is on integrating information systems and machines into a system of systems where integrative parts are done by machine rather than human [Porter and Heppelmann, 2014]. Integrating IS as data sources together with machines implementing the process can be optimized by learning algorithms. The ongoing boom of Artificial Intelligence (AI) and new uses of data mining techniques had enabled organizations to analyze big sets of data and to make automated decisions. Optimized machines can make algorithmic decisions based on previous data and spot outliers. This would reduce variance in the process and therefore enhance the quality. Enterprises
should therefore seek processes and functions where the reduction of variance has the most impact to the competitive advantage.

Decision support relies on the existing capabilities of the organization and the understanding of the operating context. Effective decisions help organizations to allocate resources effectively. This is the main source of sustained competitive advantage if the enterprise is able to allocate resources into more valuable structures. If a company can gather data that holds true in VRIN-attributes they can leverage the position. A good example is CRM systems, that utilize customer data. Enterprises can gather customer feedback from their product to enhance capabilities, but more importantly target existing customers for reselling.

2.6 Research Framework

In his chapter we m synthesis the literature findings to coherent research framework. We found the research framework to be The framework consists of three assumptions that guide the whole research. Using the framework was seen to be necessary as the research studies intersection of three fields: information technology, strategic management and organization studies. The first assumption focuses on technology and argues that it is possible to get RW objects and events modelled to the IS and other repositories (see Figure 2.6).

First assumption: Enterprise can gather data to repositories and information systems

The second assumption refers to learning and developing the organization from the use of acquired data. This assumption is based on the data utilization must be possible in the organizational context. In this thesis, data utilization is defined as the organizational use of data to create useful and effective practices that captures value from data. The second assumption also states that data can be a source of learning to utilize company data better.

Second assumption: Data can be source of building capabilities, automating process and decision support

The third assumption relies on the second. To have utilized data, and to be able to develop capabilities can be a source of competitive advantage in organizations where the value is created mainly by the knowledge work of the employees. Also organizations can build technical systems and knowledge
resources that have the VRIN-attributes and therefore can be sources of sustained competitive advantage.

*Third assumption:* In knowledge intensive organization capabilities, automation and decisive decision making can be sources of sustained competitive advantage.

These three assumptions form the research framework that is detailed in Figure 2.6. The explicit statement of assumptions allows the reader to evaluate the research framework critically and evaluate the credibility of the findings. The formulation of the research framework focuses the study to knowledge-intensive organizations and sets boundaries for in what context the research is discussed. This helps to describe the studied organization and applicability of the findings.

![Figure 2.6: The research framework](image-url)
Chapter 3

Methods

3.1 Research Approach

In the case study where research happens in only in one context, full objectivity cannot be achieved. Still, research should inspect phenomena as objective even if phenomena is open for misinterpretation. Dobson [2002] argues for using critical realism as the philosophical standpoint in information system research as it provides a consistent and useful basis for connecting subjective and objective views of reality. According to this viewpoint, the reality is a result of social structures and cannot be interpreted independently of the actors affected in the process [Dobson, 2002]. This thesis uses critical realism as the guiding research philosophy. Also, it gives room for the researcher own interpretations and pre-assumptions. This gives the researcher views more weight, which is preferred in this study, as the researcher had worked for the company two years before starting the research.

This thesis approaches research problems in a way to find out relationships in the enterprise context. The inductive approach is the preferred choice in case studies. A significant advantage of the inductive approach is that it

<table>
<thead>
<tr>
<th>Research philosophy</th>
<th>Critical realism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research approach</td>
<td>Induction</td>
</tr>
<tr>
<td>Research strategy</td>
<td>Case Study</td>
</tr>
<tr>
<td>Time horizon</td>
<td>Cross-sectional</td>
</tr>
<tr>
<td>Techniques</td>
<td>Qualitative</td>
</tr>
</tbody>
</table>

Table 3.1: Summary of research philosophy (based on Saunders et al. [2009, p. 138])
emphasizes theory building and the context dependency. Its flexible structure permits changes in the research process if needed [Saunders et al., 2009]. A cross-sectional case study approach was used to allow an in-depth outlook to a knowledge-based service company. Case study strategy focuses on understanding the contextual dynamics in the research setting. [Eisenhardt, 1989]. In the research, we used multiple methods for acquiring qualitative data. Qualitative research is common in the case studies and this study we gathered data from multiples sources to triangulate findings from interviews, but also to give members a to possibility comment findings. Still, the primary source of data is semi-structured interviews that are detailed more in the upcoming chapter. This thesis follows a research approach that we assumed would best to answer the research questions in the contextual study. Summary of the approach can be seen in Table 3.1.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{research_process.png}
\caption{Fig.3.1: Phases of the research process}
\end{figure}

In research process defining research questions and research approach was the first phase (see Figure 3.1). Having the inductive approach in the second phase of research focused on the literature review. The review focused on data in the enterprise context. Main areas were on how data can create value when utilized and how competitive advantage can be created by creating knowledge resources from data. The review also focused on internal processes
and practices needed for utilization. The aim of the literature review was to create the research framework suitable for the empirical organization context. We found that based on literature, data can be used in automation, decision-making and building capabilities as presented in the previous chapter. As the case focus was in company master data, we also identified theories on how data governance should be entailed. In literature, the roles of enterprise master data management were divided into many different sets. Still, three typical roles of owner, administrator and user were presented in most of the theories (see Vosburg and Kumar [2001]; Wende [2007]; Sarsfield [2009]; Botello [2013]). This three roles detailed the focus of inquiry that worked as theoretical sampling used in the data collection. Data collection and data analysis were the next phases of the research process and will be in gone through in more detail in the following chapters.

3.2 Case Organization

The case organization was a knowledge-driven growth organization. In the last five years, the organization has grown from 250 employees to over 350. Enterprise has two locations in Finland and headquartered in Helsinki metropolitan area. As the case organization works on information technology and high technology focus industry, most tasks are knowledge-intensive and need high expertise. Enterprise has a long operating history, and therefore, it has multiple systems. Many information systems built in-house for specific problems. In recent years enterprise has acquired commercial IS and would like to utilize them more effectively. These larger IS projects made it possible for the whole organization to utilize data sets that have only been in use in some part of the organization.

The organizational structure of the case organization was functional. Teams were structured based on the function such as marketing, IT management or finance, also business units had their function based teams such as service, consultation and product. Typically, teams consist of between 8 to 15 persons and formed so that one manager managed the whole team of experts. On top of managers were directors that formed a board of directors. Directors might also participate in multiple steering groups that combined managers and experts by some more important theme across organizational functions. The case organization was also structured into different business units and to one support unit, that had all support function teams. The di-

\footnote{In literature, this role is typically detailed as steward to highlight data quality and understanding. In this study we combine the roles as case organization does not have data governance model.}
rector was responsible for her business unit and was member or chair of some steering group. Sometimes directors even managed experts directly, this was especially common in support functions such as sales and marketing. In the research, many informants were in support functions as the master data supports the whole organization and works as a data source for management.

3.3 Informants

Patton [2015, p. 259] claims that defining boundaries for the case study is necessary as it determines the focus of inquiry. In this study, the focus is on organizational behaviour of knowledge workers in different roles based on the three roles identified in the literature. We selected informants in a way they either own or administrate any defined case organization master data set. As the informants are as well as master data users, all organizations data management roles (owner, admin, user) are entirely covered in the sample. In the study, we coded these roles into two organizational groups: managers and experts (see Table 3.2). Both managers and experts are typical users of the master data sets, but there is a distinction of administration as experts administer the data set and managers or directors are owners in the case organization. This distinction between the two groups was also necessary that individuals could not be recognized from the results.

Directors have clear management power as they lead the business unit and are on the board of directors. They are therefore able to major resourcing decision on how to govern the data in organizational wide or make decisions on acquiring IS. Their day to day tasks are not so driven by the IS; they more use reports generated from the systems. In the case organization, almost all managers had a master’s degree or equivalent and more than twenty years of work experience.

Managers have resources based on budget and objectives that they want the team to achieve. Typically they use information systems daily and use the data to develop their function, service or product. There is huge variability on how much data set is used by managers depending on the level of automation and technical expertise needed. All managers had a master’s degree or higher education level and at least ten years of work experience.

Experts were also highly educated, only the junior specialist did not have a bachelor’s degree or equivalent level of education. Most experts had master’s level education. Experts with senior role had more than ten years of work experience suitable for the role. Juniors typically had less than five years of expertise suitable for the role. Expert’s tasks were heavily based on the IS. They updated, analyzed and modified data and data models. Some
<table>
<thead>
<tr>
<th>Role</th>
<th>Coding</th>
<th>Master Data sets</th>
<th>Years in the case organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Director</td>
<td>Manager</td>
<td>Vendors and Cost objects (Owner)</td>
<td>15</td>
</tr>
<tr>
<td>Senior Specialist</td>
<td>Expert</td>
<td>Vendors and Cost objects (Admin)</td>
<td>30</td>
</tr>
<tr>
<td>Director</td>
<td>Manager</td>
<td>Customer (Owner)</td>
<td>3</td>
</tr>
<tr>
<td>Specialist</td>
<td>Expert</td>
<td>Customer (Owner)</td>
<td>1.5</td>
</tr>
<tr>
<td>Director</td>
<td>Manager</td>
<td>Staff (Owner)</td>
<td>20</td>
</tr>
<tr>
<td>Senior Specialist</td>
<td>Expert</td>
<td>Staff (Admin)</td>
<td>8</td>
</tr>
<tr>
<td>Director</td>
<td>Manager</td>
<td>Contracts (Owner)</td>
<td>2</td>
</tr>
<tr>
<td>Specialist</td>
<td>Expert</td>
<td>Contracts (Admin)</td>
<td>16</td>
</tr>
<tr>
<td>Manager</td>
<td>Manager</td>
<td>Assets (Owner)</td>
<td>7</td>
</tr>
<tr>
<td>Junior Specialist</td>
<td>Expert</td>
<td>Assets (Admin)</td>
<td>0.5</td>
</tr>
<tr>
<td>Manager</td>
<td>Manager</td>
<td>User registry (Owner)</td>
<td>7</td>
</tr>
<tr>
<td>Senior specialist</td>
<td>Expert</td>
<td>User registry (Owner)</td>
<td>1</td>
</tr>
<tr>
<td>Manager</td>
<td>Manager</td>
<td>Service catalog (Owner and Admin)</td>
<td>13</td>
</tr>
</tbody>
</table>

Table 3.2: Background of the informants
experts even developed and built architecture for the IS. Most of experts task were tied to master data sets by the function’s processes or by development tasks.

Overall, informants were highly educated and many had worked for the case organization for multiple years (see Table 3.2). While they have worked in the organization they might have had different roles, the role therefore signifies the expertise for specific role. They all worked on the company headquarters in Helsinki metropolitan are, but in different teams to one another. Even though the research focus was on the organizational context, it is common to focus informants as the unit of analysis to integrate and synthesize responses into a coherent case.

3.4 Data Collection

The main data collection method was interviews in the research. Many interviews were group interviews and even decision making workshops or meetings. Therefore multiple research data collection methods were used in the research. This is typical in inductive studies as validation of the findings are necessary for building theory [Saunders et al., 2009]. The multiple methods made the research time consuming, but also made possible of using many types of data in research and extensive triangulation.

The data collection begins with the researcher facilitated half day open organization-wide workshop, where the master data sets of case company were decided and approved. The outcome of the workshop was the list of eight master data sets that were the focus of inquiry in the research as the case organization saw those as most valuable (see Table 4.1). After the selection of master data sets the procedural data collection begun as two-phase research (see Table 3.3). The objective of the first phase was to create a maturity model to understand better which contextual data quality problems arise in the case organization and to find the master data sets owners and admins. Sampling was done purposely in a way that managers who were responsible of the functional excellence of the data sets were invited and asked them to invite all members in their team that used or administered the data or IS. After the sampling six focus groups were formed\(^2\).

The focus groups took part in one-hour workshops. There were four to six participants in a group. Focus group method was chosen as focus groups are an effective way of collecting data and to get diverse perspective [Patton, 2015]. In the focus group it was possible to concentrate on particularly

\(^2\)Master data set and interviewee count differs as one manager and one expert handled two master data sets


<table>
<thead>
<tr>
<th>Research setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Phase</td>
</tr>
<tr>
<td>2nd Phase</td>
</tr>
</tbody>
</table>

Table 3.3: Research data collection

master data set individually. This made observation of interactions between participants viable. The main point was to identify participants in the second phase of the study and the contextual maturity of the master data set in focus. The sampling and focus group were selected purposefully to build the inductive research process. The focus groups and section purposes are detailed in Table 3.4.

First in the focus groups, the research was introduced and the way how the organization had detailed master data. After the introduction processes for the identification of roles begun to make theoretical sampling possible in the second phase. This was done to find theoretically relevant cases as suggested by Eisenhardt [1989]. This was also an organizational process as those roles were meant to be used in future in governance tasks. This section was fast in some clear cases, but the debate for the roles in some cases took the majority of focus group interview time. The ones identified either to owner or admin took part in the second round of interviews. After the identification, focus group participants were probed with open-ended questions on the roles and usability of the master data set and data flow diagrams, if those were available for research. Even though questions were open, the data flow diagrams guided the conversations to data utilization barriers, processes and automation. For these main themes, the researcher took notes and for the highlighted differences between managers and experts.

After the focus group had described the master data, participants were presented the maturity model. Maturity model was a tool to find issues and barriers relating to master data sets. It had four dimensions *fit for use, clear roles, accessibility* and *governance* (see Table 4.1). First the meaning of dimensions was detailed to the participants and then they had a chance to give comments and examples on how the dimension worked in the case organization context. Comments were documented based on the informant and the master data set into spreadsheet software. As an outcome of the
<table>
<thead>
<tr>
<th>Method</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purposive sampling</td>
<td>All identified stakeholders were invited to attend</td>
</tr>
<tr>
<td>Research focus group interview</td>
<td>Effectively highlight diverse perspectives on which structured interviews can base upon. Having a high focus on the particular master data set and role identification.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction to research and master data work</td>
<td>Introduce the research and define the organizational terminology for the master data</td>
</tr>
<tr>
<td>Identification of roles</td>
<td>Identify who is the owner and administrator of the master data set</td>
</tr>
<tr>
<td>1st theme: Barriers in data usage</td>
<td>Take notes of examples of barriers in the formulation of semi-structured interviews</td>
</tr>
<tr>
<td>2nd theme: Data flows and processes</td>
<td>Find the automated use cases and probe for possibilities that data can be used upon. Map for the maturity</td>
</tr>
<tr>
<td>3rd theme: Automation</td>
<td>Map identification to maturity model</td>
</tr>
<tr>
<td>Presentation of maturity model (see Table 5.1)</td>
<td>Identify organizational maturity in master data utilization</td>
</tr>
</tbody>
</table>

Table 3.4: Research data sampling and collection for research focus groups

First phase, the case company master data owners, administrators and users were identified (see Table 3.2) and invited to the semi-structured one to one interviews.

**Semi-structured interviews** were selected to the data collection method for the second phase because the researcher can ask a series of predetermined questions. This gives the researcher more control over the theme of the interviews. [Given, 2008, p.810] This method goes more in-depth with the topic and selected themes. The semi-structured approach also gives the participants a more open way to focus their own ideas and opinions as well as the ability to focus on the most important problems. For this reason, Creswell [2013] suggest using a qualitative approach to access tacit and emergent knowledge from the organization. The research was also able to combine concepts and create a semi-structured interview guide based on the focus
group interviews. The semi-structured interviews were structured in a way that three main themes of contextual boundaries, data utilization and capabilities were investigated, because those themes arise from the first phase and had mediated from the literature review. The structure was also formatted in that way that the comments that were most detailed in the focus groups were given higher emphasis. The leading idea was to find clear real world examples that help to answer the research questions, but also to formulate a grounded theory on data utilization. This was done by the critical incident technique where the events and experiences from informants are defined clearly and consequences are distinct [Flanagan, 1954; Saunders et al., 2009].

Flanagan [1954] argues that activity detailed as critical incident must occur in a situation where the intent of the activity is clear as well as the effects. Even though the method was used to observe situations, it can also be used for experiences and self-report [Given, 2008, p.158] as used in this study.

This technique and question structure were tested with one expert that was outside the study to validate the suitability. Minor adjustments were made on how the audio was recorded and time management as all interviews were reserved exactly one hour per participant.

An individual interview started with an introduction to the research and background question and recap of the first phase (see Table 3.5). General interview structure was the same for both managers and experts, but had huge variance between individuals. After the introduction and background questions, informants were asked to reflect on focus groups and point out things that they saw as important for the study. This was done for the reason, that experts could share their understanding without to be exposed to their managers and vice versa. If participants had any critical incidents those were gone through with high detail, if not the conversation followed the structure. In cases where incidents arise those were probed with open-ended questions and at the end with "why"-questions to find attached meanings from participants. All incidents were marked into the research notes with a time stamp.

The whole research data collection was done in six weeks. In the first week focus group interviews were held. In the second week second phase interviews were booked and focus group notes were analyzed. Also, the semi-structured interviews were structured in the second week. In the third week interviews of the second phase started. There was only one interview a day as every interview was detailed on based on the research notes from the first phase. Also notes were analyzed and highlighted based on themes. This was done to ensure that enough data was collected from all the themes from both experts and managers and not asking saturated questions. The second phase of interviews ended at the beginning of the fifth week. In the fifth
<table>
<thead>
<tr>
<th>Method</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theoretical sampling</td>
<td>Identified informants based on literature review and sample themes based on the focus groups</td>
</tr>
<tr>
<td>Semi-structured interviews</td>
<td>Identify critical incidents from the themes.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction to research and data utilization</td>
<td>Introduction to the research and what utilization means as well as to tell that these interviews will be recorded.</td>
</tr>
<tr>
<td>Background questions</td>
<td>Gather relevant background knowledge of the informants</td>
</tr>
<tr>
<td>1st theme: Contextual boundaries</td>
<td>Understand the case organization context, information needs and barrier</td>
</tr>
<tr>
<td>2nd theme: Data utilization</td>
<td>Understand what brings and hinders value in enterprise context</td>
</tr>
<tr>
<td>3rd theme: Capabilities</td>
<td>Understand value driving skills and capabilities for competitive advantage</td>
</tr>
</tbody>
</table>

Table 3.5: Research data sampling and collection for semi-structured interviews
week the maturity was analyzed and the model updated. In the sixth week all focus groups were invited to the workshop where they could comment one last time the enterprise master data maturity. These conflicting comments were documented in the research notes.

3.5 Data Analysis

Understanding the research data is paramount in linking gathered data to the previous findings [Given, 2008]. There are multiple variants of qualitative research associated with multiple ways of data analysis. In this study interview transcripts, research notes, archives, emails, technical documentation and different kind of memos were used in the data analysis (see Table 3.6). Due to the vast amount of data from the case organization, finding the relevant information can be troublesome. Therefore this research had a detailed procedural approach on which the is based on the grounded theory method. According to Corbin and Strauss [1990] systematically and sequentially doing the data analysis and the data collection capture all potentially relevant immediately as they emerge. This is the main reason for selecting the grounded theory method as it emphasizes the role of continuous data analysis. The grounded theory is a flexible set of systematic guidelines that are aimed towards inductive theory construction [Strauss and Corbin, 1967; Given, 2008]. Based on the grounded theory, data collection and data analysis are done simultaneously and are interrelated [Eisenhardt, 1989; Corbin and Strauss, 1990].

According to Corbin and Strauss [1990, p.419] data analysis begins when the first data items are collected. In this study this was the start of the master data steering group and the working group as well as the open organization-wide workshop. From the master data groups "fit for use", "roles", "accessibility" and "governance" were identified as contextual concepts by the case company management (see Table 4.1). Many other technical concepts such as "data warehouse", "integrations" and "information systems" emerged as well as capabilities driven "knowledge", "know-how", "skills" and "seniority". Contextual concepts were used in analysis and were the basis of the creation of the maturity model and combining findings from previous research. Still, in inductive research it is important that the model created best fits for the material rather than previous research. This point comes clearly in studies building formal theories. In grounded theory concepts are main units of analysis, they are labels that are used in comparing incidents that have the same kind of characteristics.

The same kind of labelling was used in the focus groups. Even though
<table>
<thead>
<tr>
<th>Source of data</th>
<th>Nature of data collected</th>
<th>Time of data collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focus groups</td>
<td>research notes on comments and themes mapped to the individuals</td>
<td>After the focus groups</td>
</tr>
<tr>
<td>Semi-structured interviews</td>
<td>Audio, transcripts and research notes</td>
<td>After the second round of interviews</td>
</tr>
<tr>
<td>Master Data steering group</td>
<td>Enterprise level action items and plans regarding data sets, Material presented in the group, Memos</td>
<td>6 events, once a month through the research project, material available after events throughout the research project</td>
</tr>
<tr>
<td>Master Data working group</td>
<td>Enterprise architecture planning and reviews of data practices, Memos</td>
<td>15 events, about once in two weeks, the material available after events throughout the research project</td>
</tr>
<tr>
<td>Information system documentation</td>
<td>Technical documentation of enterprise master data flow</td>
<td>Throughout the whole research project</td>
</tr>
<tr>
<td>Information systems</td>
<td>Samples or subsets of master data sets</td>
<td>Throughout the whole research project</td>
</tr>
<tr>
<td>Member checking event for master data views (see Table 5.1)</td>
<td>Comments, emails and notes</td>
<td>After the round of interviews</td>
</tr>
<tr>
<td>Documentation of management practises</td>
<td>Internal guides, organizational policies and board of directors memos</td>
<td>After the round of interviews</td>
</tr>
</tbody>
</table>

Table 3.6: Research data sampling and collection
the focus groups were also a role identification meetings, many new concepts came up as probing critical incidents, such as: "accountability", "believability\(^3\)", "governance", "skills", "security", "tools", "data overload" and "processes". In focus groups there were also many comments that detailed the quality of the master data set based on concepts stated before. These comments were written down and given ad hoc concepts such as: "poor data", "not used in the organization" and "no roles" An interesting finding was that the same technical concepts that mentioned in the steering group did not arise in the focus group. This steered the analyzation focus even more to the organization context rather than to technical solutions. Technically this was conducted so that comment and decisions on the maturity model were documented to the research notes. Also the few critical incidents were detailed into the notes. The researcher made a brief memo after the focus group on which concepts arise on interview and which concepts should be asked on which informants. This memo was compared to the other memos after the next focus group interview to use the same concepts.

After the data collection of focus groups maturity model was detailed with three level maturity hierarchy. These data quality categories were formed by the concepts and formed in a way that they were appropriate to all four concepts detailed earlier. This was done by finding within-group similarities and intergroup differences as suggested by Eisenhardt [1989, p.540]. The categories were detailed by the research problem and organizational context. There was also little literature on the contextual maturity of data, so it was reasonable to create a new contextual model. The categories selected were **satisfactory, insufficient and attractive**. Categories mean a grouping of concepts which have the same underlying similarities based on comparative analysis [Corbin and Strauss, 1990]. Findings are detailed in the fourth chapter and Table 4.1. Of course the model was updated and re-evaluated after each data collection phase, but only minor fixes were done based on the member checking.

Before the data collection of the second phase could start, the informants’ roles had to be checked based on the framework. We wanted to sample informants on how they handled particular data sets in the enterprise context. This was quite straightforward as the case company had made the decisions based on the same criteria in the focus groups. Only one manager needed to be grouped into either *expert* or *manager* as he was both admin and owner. This was based on the data set usage and the admin load assessed by the technical structure of the IS where data was stored. The manager’s data

\(^3\)The word believability was used rather than trustability to make distinction as there was a belief that data quality was poor, but systems in general were trusted.
coding was matched as a manager (see Table 3.2).

After the role checking, the guide for semi-structured interviews was formatted (see Figure 3.2). It was done grouping concepts together to three categories: data utilization, capabilities and data boundaries. Technically this was done by analyzing focus group memos and maturity model with master data steering group and master data working group memos. In this phase validity of informants comment was not triangulated with other documentation. These three categories were then matched with incidents, concepts and issues that arise from the focus group phase to have a formal structure. Even though the coding for two groups was done on based on the data usage, the semi-structured guide was modified based on issues the informant raised in the focus groups. All interviews were focused on finding critical incidents in detail.

The theoretical sampling was done with questions asked by theme as well as the sampling the informants. Also the underlying interview guide was the same for both managers and experts. The guiding idea was to sample and seek data so, that those three categories would saturate as incidents would repeat. This meant that all themes had incidents from both managers and experts. In the formation of categories, questions were also structured so that for same data set the same dimensions were asked. If in the first interview of the data set an informant had a critical incident regarding a dimension, this dimension was asked in detail in the second interview of the data set. Also the critical incidents were probed cross-dataset. For example, if Customer data set owner had a critical incident in capabilities dimension, this was asked in detail in the interview. After the interview, the interview was analyzed and new questions for probe the situation was created and asked either the data set expert or other set owners. If the incident re-emerged it was marked as saturated and to be validated with triangulation. Technically this was done so research memo document, where the incidents were marked, was quickly analyzed after the interview. All critical incidents were mapped to spreadsheet file with concepts that emerged from the interview. Then questions were formulated if necessary to validate and saturate the findings from the other informants.

When the actual research data collection, the data that was not part of organizational processes, triangulation and analyzation phases begun (see Figure 3.2). In this phase all semi-structured interviews were transcribed into text documents and critical incidents from research memo were identified and new incidents were searched. The new critical incident was labelled with existing concepts and added to the spreadsheet document. After the conceptualization, the concepts in a specific category were mapped to groups. These groups were then analyzed and triangulated with organizational data
CHAPTER 3. METHODS

Figure 3.2: Summary of simultaneous data collection and analysis

- **Master data steering and working group**
  - **Data Collected:** Concepts of data utilization difficulties from management perspective
  - **Analysis outcome:** Creation of initial concepts

- **Open organization-wide workshop**
  - **Data Collected:** Identification of master data set. Definition what master data is in the organizational context.
  - **Analysis outcome:** Master data maturity model initial creation and mapping the stakeholders. Categorisation of concepts

- **Focus groups**
  - **Data Collected:** Roles of stakeholder, maturity model concepts, new concepts and critical incidents
  - **Analysis outcome:** Categorisation of concepts into semi-structured interview themes. Updating maturity model

- **Semi-structured interviews**
  - **Data Collected:** Critical Incidents, new concepts and transcripts
  - **Analysis outcome:** Analyzation of similarities between data set utilization based on the categorization

- **Triangulation and analysis**
  - **Data Collected:** The case company internal documentation and member checking event
  - **Analysis outcome:** Identification of the emerged theory and answers to research questions, validation of findings
CHAPTER 3. METHODS

(see Table 3.3) and cross data sets to validate the findings. As this analysis was done, data set maturities were checked to be uniform across the whole organization. This was important as the informant had no shared understanding of the contextual quality of the data.

Before the final analysis, the maturity model was shown to all stakeholders for verification and for comments. In this session the categorization of the maturity model data was presented in detail and participants were able to ask questions. Creswell [2013] argues that member checking is a vital part of data validation. We found in especially in this research that studies contextual aspects of data utilization, having shared understanding with all stakeholders is necessary as the results are based on this context. In this session maturity model was approved and there were no changes made to it afterwards. The maturity model worked thereafter as a tool that further analysis relied upon.

Lastly in the analysis phase grouped concepts were created as formal categories. Some categories that have only a few concepts were disqualified as not saturated. The research found four categories, where was multiple critical incidents. From the incidents direct quotations were taken to allow the reader to value the validity. The findings are presented in the next chapter. After the formation of findings, findings were analyzed based on the research questions, research framework and previous literature. In this phase concepts that re-emerged were discussed and theory modelled for enterprise data utilization.
Chapter 4

Results

4.1 Master Data from Enterprise Standpoint

In this chapter, we focus on findings from the round of interviews, especially from the focus groups. The interviews mapped the maturity level of enterprises master data management and usability. The chapter also sheds light on how management choices affect on utilization effectiveness of the master data.

Master data refers to core entities that are consistent and used across the business processes of the enterprise. Typically master data entities are products, customers, staff and organizational information. Master data varies from one industry to another and even between the enterprises as enterprises in the same industry can have substantially different business models. The case organization defined master data as following:

"Master data is the consistent and uniform set of identifiers and extended attributes that describe the core entities of the enterprise — and are used across multiple business processes. There is a single agreed source for each master data attribute."

As the definition points out, the core entities and business processes should be aligned. This data was seen as an asset that could be utilized from existing information systems.

At the time of research, management of case organization had long debates on what entities should be selected to the master data list. If the entity is listed, it should have a single source of access and also hold organizational "truth". This would mean that business units do not have their own repositories, for example, for vendors, but relies on an enterprise vendor list. In this example, only vendors that are found in enterprise master data source are
### CHAPTER 4. RESULTS

<table>
<thead>
<tr>
<th>Masted data set</th>
<th>Fit for use</th>
<th>Clear roles</th>
<th>Accessibility</th>
<th>Governance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Staff</td>
<td>Insufficient</td>
<td>Satisfactory</td>
<td>Satisfactory</td>
<td>Satisfactory</td>
</tr>
<tr>
<td>Cost objects</td>
<td>Satisfactory</td>
<td>Attractive</td>
<td>Satisfactory</td>
<td>Satisfactory</td>
</tr>
<tr>
<td>Contracts</td>
<td>Insufficient</td>
<td>Insufficient</td>
<td>Satisfactory</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Assets</td>
<td>Satisfactory</td>
<td>Insufficient</td>
<td>Satisfactory</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Customers</td>
<td>Insufficient</td>
<td>Insufficient</td>
<td>Satisfactory</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Vendors</td>
<td>Insufficient</td>
<td>Satisfactory</td>
<td>Insufficient</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Service catalog</td>
<td>Insufficient</td>
<td>Satisfactory</td>
<td>Satisfactory</td>
<td>Satisfactory</td>
</tr>
<tr>
<td>User registry</td>
<td>Satisfactory</td>
<td>Insufficient</td>
<td>Satisfactory</td>
<td>Satisfactory</td>
</tr>
</tbody>
</table>

Table 4.1: Master data teams views on how functional data sets are in the enterprise

This calls for resources and capabilities for implementing master data to a single source.

In Table 4.1 is collected results of the interviews with master data teams combined with memeber checking. Findings were collected using three criteria: insufficient, satisfactory and attractive. In insufficient situation, the primary organizational function could not be with operated with ease, roles where not defined and had no apparent owner, information was not accessible for experts and policies and decision-making structure on data set was not clear. The satisfactory term is used in situations where there are clear and detailed use cases for the data. There also exist integrations with other data sources, and at least some task is done automatically. Satisfactory roles mean that there is clear ownership of who is responsible for managing the data, and understanding on how the decisions are made based on the roles. For accessibility to be satisfactory, data set users should be able to access data in some form. From governance perspective, satisfactory can be defined with clear roles, policies and there is a person or a function that is responsible for making decisions.

The attractive label highlights the ease of use and capability to build on top of the data set. For example, in fit for the use case, that would mean that new services and products could use and trust the data set. To have an attractive fit, the other dimensions need to most likely to be attractive also. This would mean transparent role management, accessibility by the
user interface and API. It would mean that as many services use the same source, changes into the metadata, system or risk management are done by the governing party in the organizational level to align data utilization with corporate strategy. To sum up, insufficient data set hinders operational efficiency. In a satisfactory level, existing processes can be optimized and even automated, but only on attractive level innovation and new capabilities can be built.

Using the categorization we can see that biggest problems are in Fit for use, Clear roles and Governance. Many reasoning for fit for use was given, most often insufficient labelling was due to the fact, that multiple data sets of the source existed or the master data source could not be trusted. Only in staff, data set insufficient labelling was selected as not all existing processes could be done with using the master data source without changing either underlying process or data set schema substantially. Compared to fit for use dimension, clear roles was seen as quite positive. Even though it was not clear who was responsible for a specific task, teams assured that task typically gets done. This might lead to some friction and highlights internal selection bias. Comparing this to the insufficient roles tend to lead higher lead times and unplanned work according to the first round of interviews.

Problems in governance are driven by how much business power is embedded in the master data set. Due to the legal compliance, staff, user registry and cost objects are at a satisfactory level. In the preliminary round, the governance was much about compliance. Only the service catalogue was out of compliance viewpoint. This was due to the fact, that even though the data was available with governance function, it was not fit for internal usage. In essence, the services seemed to be least valuable master data as it was not seen as having value as business units typically sold their own services. This underlies the power or value that needs to be embedded to master data to be useful for the organization.

### 4.2 Master Data from Individuals Standpoint

In the previous chapter we focused on how the case company sees master data on team and master data set level. This chapter focuses on how experts and managers see the master data what topics surfaced from a structured interview.

After the first round, it was imminent, that governance was a great issue in the case organization. In semi-structured interviews many experts said that they did not know what they were allowed to do with data, also the role of data management and IS access blurred. One expert detailed the
CHAPTER 4. RESULTS

confusion in this way, when asked how one can get access to data:

"You can ask the data as an Excel file. I do not know what our license policy is. And how easily you can be given access rights?"

According to the expert, data is exported from many systems by different employees and stored to their personal computers. Typically the analysis is mixed with some other data from other systems. There is a lack of governance on how data can or should be used and what processes will use them enterprise-wide is not clear. Another expert highlights the problem as follows:

"In some cases, you have to store the data, but what data? Here in [case organization], we have non-existing data insertion and deletion process. I have pointed this out [six months ago].”

The expert noted, that these kinds of procedures and practices should be adopted into the whole organization as the data set is used by the whole company. The expert expressed concerns to the point that the management does not see that as an important fact and they do not drive the adoption of data management practices. To compare experts views with managers, there was a clear difference. Experts do not know how to cope with demands that are put onto them. Problems can arise from multiple sources, but many experts emphasized friction to their working with data. One manager even said a bold claim:

"In our group everything works. [...] Dedicated information systems functions really well.”

The claim is understandable from the managers perspective. The IS are built and selected the team task in focus, but adding new or enterprise-wide task is troublesome. Still, both managers and experts see that governance of the processes should be done from enterprise management. This might be due to the line organizational structure and highly task-oriented way of working. Some experts expressed frustration that internal customers were not able to communicate their requirements in technological terms or they did not understand the consequences of processes variation. One expert said that his task related to one master data set are mainly due to people not knowing or following the process:

"Should not add thing afterwards by hand. Then the number would identify [data entity] well and it would follow [the whole business process]. „
CHAPTER 4. RESULTS

Education for processes or system were only little and most of the learning should happen on the job. In the interviews, many managers said that their recruitment has focused on task-specific skills and information specific skills was not taken into account. According to managers, finding experts that have both experiences to a particular system and task is hard. At the same time experts see, that coworkers are fully resourced and finding help in systems usage is troublesome. This might be a part of the cultural issue as one manager puts it:

"Here come and goes 50-80 employees in a year. We have not been able to change the culture as we wanted. We must start to simplify and streamline our basic processes."

The case company has been growing in past years, that had put some pressure on support processes. While the company has been scaling the processes has not kept on the phase and there are a lot of possibilities for making processes better. That was common theme managers, that most managers shared in the interviews. Still, there was a clear distinction on that most managers did not see the need for development as it would cut resources from existing operations. Development efforts were seen to be an organizational effort.

These results highlight the effect of governance in the case company. Management seems to think development as buying better IS and thinking experts will learn by doing. Experts acknowledge that the case company has acquired good tools, but that is not enough as data generation processes do not work. The clear result compared to the company-wide view is the need for data generation process and simplification.

4.3 Data Quality Barriers Depend on the Function

In the previous chapter, we focused on comparing how managers and experts see master data management in their company. In this chapter, we focus on comparing contextual data quality issues that were presented in the interviews. We also study the views on how what barriers are present while data is utilized.

Data to be utilized needs to be accessed. In many cases having access rights is not enough as it takes technical skills and systems detailed to process. When asked what is the biggest barrier of information utilization, one manager pointed out that getting access to the best tools is troublesome:
"Combination of wiki and [ticketing system]. It is not efficient. Drives me crazy when they do it. Access denied for any other place. I would like to collect data in a more central place. [Management] should jump into roles of that people."

This manager argued, that the necessary data have to be mined from multiple locations and if management needed some information it was not valuable enough compared to the workload needed. She also noted that, the reasoning for getting some information was not detailed most of the time or was due to some reporting promise to customers. These promises could be different from a contract that made this even more troublesome. In some cases knowing internal business processes is hard, when the customer information is handled by business units. Information moves too slow to be processed. One manager points out that, the big barrier of knowing is not knowing where all data is or how big the transaction costs are:

"... customer data is in multiple locations, user data is in multiple systems. Data does not transfer in electronic form even though it soon 2020. It is transferred in a paper if it is even transferred"

Data not being in a central location and not being trustworthy is causing barriers. Many functions circumvent the barriers by building their own lists and registries. IS for handling these registries are selected based on function needs. According to one manager:

"There will be practical issues on how to get for example marketing, to use one source that they do not have their own listings. How to get sales to use one source. [...] There have been many competing sources and all the time we are getting more"

Even though this highlights the need for governance and standardization, one underlying barrier is not addressed directly. As an expert in a function typically has no previous experience IS management and that is one assumption why the search for the new tool is probably done by their context. This has resulted in multiple systems for multiple tasks which makes integration harder. This should be an organizational choice and not driven by some function. One manager underlines the problem regarding on what basis organization structures IS:

"It would be great that the information could be stored to one location and everything could be integrated, usable offline with easy to use the tool. But then there are these security things that we need to consider. That has been the biggest problem here. There is every kind of passions."
Security is seen as very important in the case organization, but it seems to be put first over the business value. A good example of this playing out is work environment change project which could not have used data from the meeting room as it was not recorded due to security reasons.

![Barriers to data utilization in functions](image)

Figure 4.1: Data barriers found in function context

Access rights related problems were one of the most noted reasons for both managers and experts. This typically emerged as long wait times to getting access to the system and on the other hand as raw data exports. The documents were shared via email which created new problems in understanding data. In many cases, many experts were needed to combine multiple documents to one working document. Also in data-heavy documents relied a lot on skills of asker to analyze the material. In high-velocity data sets using external data or combining sources out of IS were barriers why functions could not utilize data to second and third order use cases. These are a collection of problems that functions face as can see summarized in Figure 4.1. The common nominator is the poor accessibility meaning, that is unclear how to use, get or handle data in cross functions or even the same function. In many cases, functions would be more effective utilizing existing document editing skill to compared to learning IS, argued one manager even. Creating procedures on how to create simple transformations were seen to support function role, that did not exist in the case organization. In general, getting access and do transformations was seen usually too troublesome and workarounds were therefore commonly used.
CHAPTER 4. RESULTS

4.4 Culture for Sharing Knowledge is Needed

Knowledge-based work emphasizes the need for recreation and sharing of knowledge. Sustaining competitive advantage comes from the dynamic capabilities, that case organization poses as it is transitioning more and more knowledge and service intensive direction. Creation of new knowledge is, therefore, the utmost important, but the value only materializes when something is delivered. Case organization has had some problems in how the value is delivered one expert says:

"Getting things moving is hindered by the organization’s way of working. We wonder and wonder. When we work with a commercial party they are wondering how can a procurement process take over a year. On the other hand, people have quite big loads of work."

When asking managers and experts to define the culture words such as "open", "undefined" and "not shared" stand out. Only one expert mentioned that, sharing of knowledge in the culture section. Comparing this to his manager’s response had huge asymmetry, as the manager said that BU lacked skills and ambition for sharing knowledge. There are designs in the organization that clearly help and hinders knowledge sharing as pointed out from one expert:

"Sometimes I hope encouragement for sharing knowledge. When you know you have done something really well, you could book time and show to the rest of the team. But... The good and bad thing in open floor plant is that you can go next to someone and ask why this is not working.

Sharing knowledge was not rewarded in the organization. In some cases, holding the data is even seen as beneficial as not giving access holds resources and gives individuals more power in budget negotiations. Data is not used for cost-benefit analysis and maturity of data analysis is seen not to be a priority. Therefore, knowledge search happens mainly via interpersonal relationships. One expert highlights the problem when asked how to seek information inside the organization:

"When you have a small hunch that I would need to find this kind of information from wiki. I would have no idea. Then you go to the next door neighbour to ask if she knows who is responsible for these kind things. I have no clue about which system I would go
to look for roles and responsibilities. Could be that some would say that it has always been the one system. [...] there has been too much information.”

As the unstructured data would have been available is not findable. One main reason being that the organization has many overlapping IS for the same kind of information. For example, they have at least intranet, internal wiki, internal hosted semi-open workspace and two internal hosted open workspaces and to combine that many employees use different IS to circumvent the security and use "easier" solutions such as Google Apps. This had lead to a situation where information has been added in many buckets, and according to one manager this is due to the ambitions set by top management:

"The drive is missing altogether. [Case organization] is completely scattered. The information scattering between systems is getting worse. On employee level, there is a huge amount of people that have been doing the same thing very long time."

The second point emphasizes the need for process improvement. In the interview, the manager focused on the point, that processes in place do not support data utilization. Employees in BU are not able to gain new skills and re-allocating resources is hard as allocation happens in organization-wide and is not agile. Another manager argues, that changing employees team or BU is troublesome as practices can vary a lot:

"How you deal with work varies a lot. There is no unified way of doing things"

Lack of unified ways of working makes it harder for the expert to develop skills even inside the team. When making organizational changes, older structures in a team can last for years. This kind of situation played out in one BU:

"Other team members have substitutes. I could not be their substitute even if needed. If both of them would be a way it would be risky."

This expert has been in the team over a year and really comfortable in his role, but sharing knowledge of ones work in team level has no process. Only the ones that are each other’s substitutes shares the overlap. Few interviewees saw this as a huge problem why things are not developed as small coalitions that have been each other’s substitutes do not want to change the process as it would need resources for learning new processes.

Role of organizational culture in what is encouraged, what kind of actions are preferred and how much learning is emphasized seemed to have a huge
part in data utilization. If sharing of data is not encouraged and power is structured on IS in budgeting and resource allocation, the company cannot utilize data for learning and analysis. The case organization has not to build a process for facilitating information and learning. The culture emphasizes quick wins over long term development, this might be due to the lack of measuring and analyzing on how data should be used and what data the company has. Also, managers and experts have different views on how systems and processes work and should work.

4.5 Information Asymmetries Leads to Different Ways of Understanding

Bigger companies have clearly more complex organizational structures, which affects the understanding between BUs. This same effect can be even found between teams when talking about the enterprise’s information flows. Lack of understanding people needs and other teams business processes leads to information asymmetries. Information asymmetry means that one party has better information or have capabilities to gather information more effectively. High information asymmetries can create knowledge silos to the organization due high dependencies of knowledge. One expert answered to what are barriers of information flow:

"I do not know. For me, it seems that everyone is in a hurry with their stuff. And then there is the human factor. You just do not think, that in somewhere people would need this information, that there are dependencies."

Dependencies seem to be the underlying topic in many interviews. One problem was that it was hard to know who worked with the same kind of tools or structures as most IS were handled by one team or BU. Sharing information was not measured in the case organization and ways of measuring work in data and system related task can be troublesome. When tasking expert into working with a task, there should be key performance indicators (KPI) to hold the task into business value as one expert puts it:

"You can be really slow making only one task in three days. Then someone changes the design and [IS] makes the task on every piece of things you do. That is an example of KPI that lies. This is really hard."

The problem what the expert is detailing is common to many knowledge workers in the organization. Many tasks are driven by some measurement
given by the manager and based in hourly billing cycle or internal accounting, but the underlying value is not taken into account. One manager stated that reason for unified practices between and therefore information asymmetries are the versatile reporting practices in the case company:

"We had made agreements with the customer how we report things. In many contracts, there are old agreements that are renewed in the same way. We cannot automate most of those individual reports."

This is one reason why master data is not used effectively as it does not support custom reporting. The company had many policies that were documented in the intranet, according to one expert. Still, many things are done by heuristics and individual practices. Changes in policies and guides are not enforced and communicated from upper management, For example, one expert said:

"It is a real problem that you can not find anything from the intranet. I do not know do people even bother to look. It is that way that people come to ask because they have not found the guide. Then I send a link to them via email."

Concerns regarding working practices were more widespread in the case organization. As a lack of process individuals used emails as a source of guides and working practices. In master data related task the uncertainty what employees are doing is not clear as managers cannot control as many have nontechnical backgrounds. Talking about this issue a manager said:

"... doing the tasks are so much of this, not process-like and without any IS. Is email a system or if it is counted as a system. Hah."

Information asymmetries of working context between managers and experts varied a lot. In cases where there had been the best understanding of work context, the information was less used in the company. As the data was less used, managers tend to have a better view of task related to experts work. Expert tend also to have been working in the same role in several years and in many cases even with the same manager. On the other hand, when data was used in many company business processes, the accessibility and usability of data was the paramount feature and delivered clear business value.

As we the interviews were done from the data sets that the company has classified most valuable, is it reasonable to assume, that company could see enormous business value in the data utilization. The manager’s comment below illustrates frustration that substantial minority hold:
"Management also understands that something needs to be done, but they lack commitment. There might be some cynicism because they have tried to fix these things so many times before. Always worrying about the cost and everything additional is out of the picture. Arguing takes a lot of time. We lack straightforwardness."

This was a clear finding as it was view hold most firmly by data-driven business units. Some felt that case company focused too heavily on costs and many valuable things were left undone as it was not the priority of upper management. A small number of participants demonstrated use cases how they could cut costs, but that would take extra effort. When asked why they would not implement the propositions, the participants were unanimous in the view that the resources that they would free would not be allocated back to the team. They argued that is due to the problems in understanding the fragmented service portfolio of the case company and working practices. Commenting on different services and how a case company is managed, one interviewed manager said:

"It is it, that not everyone understands what we are doing and why we have so many different services. We just talk that we need to productize these services."

Interestingly, the understanding differences were observed in almost every interview. If not between expert and her manager, between teams and business units. Common themes with expert and manager were technological understanding, different prioritization of tasks and lack of process. Between the teams and BUs main theme were unified working practices and collaboration barriers.

4.6 Lack of Capabilities Hinders the Use of the Data

In this chapter, we focus on the capabilities that the case organization has and would like to acquire. The underlying assumption of the research has been that capabilities in the knowledge-intensive organization are the main source of sustained competitive advantage. Building data related competencies and dynamic capabilities should also be in focus of companies as the possibilities for data utilization wideness.

Capabilities seem to be distributed widely in the case company. In many teams, there was clear distinction what is "technical" and therefore out of
the scope of data owning the team. Many teams do not have the capabilities to manage company-wide IS. It was surprising to result to understand that, using advanced features that relied on data processing, data transformation and data loading were seen as IT processes rather than business processes. In one case, the expert thought that:

"There should be internal support for IS. We should have the support nearby in the house. For example, in a situation where we are moving documents or data around."

Data related business processes were seen to be IT processes in teams where employees had been working multiple years in the case company. On the other hand, employees that had previous experience from different industries or were new to the company saw IS and data more related to the underlying business process. This might be due to the resourcing and practices used in the case company. The case company has been acquiring capabilities as it has been growing in rapidly in past few years. Existing business units have not been able to build capabilities while the company has been growing argues one manager:

"Yes, while [Case company] has grown [business unit] tasks and also the amount of information and reporting should be growing, but we have not been able to do that. We have not been able to address our [business unit] needs let alone the needs of the whole organization."

As the manager’s team is not able to provide benefits for the data, the data has been unused and features of the IS was not used almost at all. According to both the manager and expert, the main data was handled and processed in spreadsheets. This processed data was not shared, documented or validated throughout the organization. The way the data was presented to decision support was not clear to management as the processing method was not clearly defined.

Surprisingly, only a minority of interviewees used multiple data sets and were interested in using them. Combining data cross-organization could give added benefits in decision making and open new possibilities. One manager summaries the point as follows:

"Tasks are typically done the same way always. Having new angles on how to do things is seen as an extra push. We have a lot to do in process thinking. We are doing good work, but not everything we do is very structured."
Managers and experts have asymmetries on how data could be utilized. Experts think work more related in the ease of doing tasks. One interviewee even pointed out that automating task would lead to more unpleasant task such as monitoring, using unfamiliar IS and reporting. These new automation related capabilities are the cornerstone in utilization as found in the previous chapter. When experts do not have sufficient learning and management is not able to see the skill and utilization gap, the automation of task might lead to unanticipated situations. One situation in the case organization was that new IS was bought to automate some tasks and replace the old one. The end result is that the organization has overlapping systems. As one manager said:

"It is very common that data is in two separate systems. They can be in the same also, but the systems as they are bought do not support reporting or it is very clumsy. That is actually the main reason To be really frank, we also do not have skills in those areas."

Not having capabilities to manage reporting, analysis and organizations specific tasks in the new system is troublesome from a capabilities standpoint. If capabilities are not present when new IS is fully employed to the organization, specific BUs and teams will use existing capabilities for circumventing the new IS. In case organization were multiple cases where IS workflow or analysis function was not used or configured. The typical workaround was to use a spreadsheet in some shared location in cloud or internal network drive. Talking about this theme an expert said:

"We should have a lot more data analysis. We have the self-service business intelligence tool - kind of. It is not enough, that you just add values to a spreadsheet. It does not tell anything. Everyone collects the data they want, but we never integrate or mix those in any way”

Asymmetries are big when the context is not presented to the data receiving party. As making analysis, implementations or tasks in general from the data, an expert needs a high-level understanding of the context or data needs to be highly structured. If analysis lacks structure or the structure is not familiar to the receiver, information is not passed to management. Asymmetries in understanding structures were troublesome finding, some managers could not understand data models even though they have acquired IS for specifically solve problems that were in organizational enterprise architecture or in the company data model. On the other hand, delivering
business outcomes via IS can be as troubling if the business process cannot be effectively communicated to experts. This asymmetry deepens problems related to resourcing, enterprise strategy and data utilization. The expert’s comment below illustrates the asymmetries in understanding:

"From the IS point of view we have no problem. I would even say that in that category of IS, it is the Rolls Royce, It is not cheap, but really good and powerful. Not a one time have I come up with the situation that you could not do the job with it. But, the biggest problem is the resources. Or how the project is managed. Bit of this kind, that we do not know what to do and when. It all stems from resources. We had made a lot of good things. Great stuff, but the due the resourcing they are not finished in many parts”

As the expert suggest more resources, but the impact might be even greater if the efforts would be focused on the high-value task. This expert’s manager argued that big problems in his work were due to the really diverse tasks that the team were doing. Even having resources is not enough if the resources do not have specific capabilities. The clear finding was that when organizations are using data-heavy processes capabilities on handling information asymmetries is highly valued. In knowledge work detailing outcomes can troublesome tasks for managers in data related terms. On the other hand, expert’s might not know how the new configuration to IS changes the complex business process.

In many cases, the needed technology was capable of the use. If the main reason for not using data was lack of skill, training would in many cases solve the issues. Taking into account the business environment, the landscape becomes more complex. Management needs to balance with exploiting existing capabilities and building new ones. In case company many managers saw that building new skills in teams was not a plausible option. They rather hired new employees to bridge the skill cap. The main founding was that, interviewees that had an understanding of how their actions enabled others work saw data as a valuable resource.

4.7 Summary of the Results

In knowledge-intensive organizations data is both the input and output of the most processes. In between the processes data gains value as it is transformed, processed or sorted. When most processes are driven by data, or knowledge in general, data quality links greatly in the process quality. The
more organization automates, estimates or forecasts, the better the contextual data quality needs to be. In this study, we found that data quality in the enterprise is highly contextual and what data utilization barriers companies do have.

Initial findings on master data were that it is very easily distributed in unmanaged fashion. Master data is used to manage the organization and therefore functions store that data in some form to conduct operations. We found that functions produced or transformed data in many forms, that was unmanaged. A typical example being when an expert or manager asks for data, they are not guided to the IS or application layer, they are rather given data export or custom report in spreadsheet form. Lack of detailed process and documentation has resulted in the data to be untrustworthy. The organizational barrier to using data is to get employees to trust the data sources. This will force the organization to have some governance practices to ensure data policies. This has been illustrated in Figure 4.2 as the basement layer that leads to data quality gap.

Data quality gap from master data standpoint means the problems that data uncertainty and low-quality cause for organization management. In many interviews data in IS was taken as a fact and as a ”source of truth”. The analysis made from those sources needs to take into account how the data is produced, how valid it is to make a decision in the underlying context and what are the applications for the data. Designing organizational priorities so that data governance can achieve detailed quality and maturity level was seen to be problematic at least in regarding master data. Making data higher quality need resources and detailing what the dimensions of how data is evaluated in the organization are. In this research, we made preliminary findings on how contextual dimensions were seen in functions. From that, an interesting finding was how enormous meaning roles has in the enterprise context. Roles based on handling the data and the management structure has a big effect on utilization.

Second organization data quality barrier was the management ability to select value-adding applications and manage them. Many experts pointed out, that the application or IS was really good, but there was a lack of resources or understanding how to develop it further. For example, in finance function the development was mostly done by external consultants and organization lacked capabilities for identifying and implementing value adding features. This emphasizes the need for managers to understand the application in detail and how it integrates data flow to the business process, according to the interviews. Experts noted that many internal systems lacked roadmap on how things will progress in the future and what are the features that will be implemented. On the other hand, IS that were highly integrated
and had little customization worked most effectively. Those functions had the most detailed processes and most organization-wide use cases. These results imply that IS and application selection plays a crucial part in success, but also the capabilities needed for IS or application management are vital.

Figure 4.2: Summary of the findings: why data was not utilized in the case company

Third organization barrier that was encountered was "poor applications" as they were called by experts. In cases where the value of the application or IS was not seen by the experts, workarounds were invented. This strongly indicated information asymmetries as managers tend to get reports from IS were things were different compared to working documents such as spreadsheets. This causes several misalignment and burden to management. In one example manager even disputed data from a master data system to be false that he was responsible. In further investigation, it was discovered that the underlying business process did not use IS any other way than as the data generation system and all processes worked in spreadsheet documents. This barrier combined problems in functions, in BUs and in the whole organization. Problems ranged from the way of working and culture to IS documentation and business process descriptions. Interviewees noted these problems to be barriers to data utilization. On the other hand, many under-
lying reasons why documentation or process was not detailed can be traced to a lack of skills and capabilities on implementing the change.

Fourth organization barrier can be detailed as the capabilities gap. Expert clearly lacked skills for data utilization. An interesting finding was that the missing skills were not directly linked to the IS. Needed skills were more in relations how to document work, detail out procedures, find information and learn effectively. The expert did not see that there was enough time to learn new things or in many cases they saw it as an impossible task. In many cases, there was not a starting point on how to begin learning. Some experts that were responsible for the master data sets did not have any formal or informal education for the system. This was noted to be a huge gap as no learning resources were available. Learnings were also not documented and shared in any functions, which underlines the importance of culture in a data-driven organization.
Chapter 5

Discussion

5.1 Answers to the Research Questions

The primary goal of the research was to find data utilization barriers in the enterprise context and increase the value for data through utilization. As the research focused only on data used for organization management, also the strategic implications were researched.

5.1.1 Complexity of the Context Predicts Utilization Difficulties

RQ1: What contextual barriers exist in the utilization of data in the enterprise?

Data utilization in the enterprise context means that there should be value added to the process for data to be utilized. In that context intrinsic data quality is not the biggest problem, instead is the uncertainty of getting value from data. In most cases, the accuracy was found to be relatively good. On the other hand, interviewees saw that data was not "fit for use" to value-adding applications. This supports Wang and Strong [1996] argument that data quality is more than accuracy and should be studied in contextual form. These kinds of utilization barriers were typically due to the contextual data quality or business processes design.

Utilization barriers due the data quality were driven by barriers in data accessibility. If data is in multiple locations and data, not one location holds complete data, and data is not used according to our research. This is was found to be one of the main barriers in starting data utilization for individuals or functions and should be taken into account designing organization data
architecture. These findings suggest that enterprise master data should be unified and adequately defined, documented and measured. Interviewees as data consumers evaluate the data quality concerning the job to be done, not in the relation of accuracy or completeness of the data. This finding was also reported by Strong et al. [1997]. Even though accessibility was a most noted barrier in the interviews, other notable barriers where **Usability**, **Timeliness** and **Understandability** that are related to the data operations design.

Research suggests, that unmanaged data production is clear barrier for data quality and believability (see Figure 4.2) as it makes data to be untrustworthy. One unanticipated finding was that much value could be extracted from sparse quality data if data is in proper form. In situations where data could easily be transformed, updated and loaded by data consumers in one IS, the data quality improved substantially. One IS where data was aggregated properly still was the requirement for successful data utilization, but not guaranteeing it. Successful data utilization needed value-adding business processes.

![Figure 5.1: Competitive advantage of the data utilization process [Vayghan et al., 2007]](image)

Gathering data starts typically from measuring business process and its operational efficiency (see Figure 5.1). Measurements are used focused on reporting and understanding business process more in detail for development purposes. In this phase, the focus is heavily on intrinsic data quality and the
accuracy and completeness of the data model to match the business process. Master data differs in that form, that it is used for managing the whole organization. In this maturity phase, the focus should be on reporting and intrinsic data quality according to Vayghan et al. [2007]. We found that master data sets having the lowest maturity had understandability barriers. If data sets and attributes were not understood in the same way across the whole organization, utilization was hard even when existing IS functioned, and data was in a usable form. Understandability of data is necessary as the context is broad.

Understandability means that every data consumer understands attributes same way and have an understanding of the underlying assumptions made in data production and models. The current study found that when the understanding of data was not clear or if different BU thought data model differently integrating reporting and detailed process was not possible. For example, in case organization defining what customer means and translating that to the organization-wide business process was troublesome. Therefore customer object was not translated into understandable form to the CRM system even though enterprise had made sizable infestations for acquiring the system. Understandability is crucial for master data as it is used for company-wide management. Otto and Schmidt [2010] even used organization-wide usage in their definition of master data. Master data is used for evaluating performance and forecasting future from the management perspective. If most valuable data is not understandable, the company cannot adjust its processes as management would like due to the information asymmetry.

When the enterprise has data that is "fit for use" and usable for reporting and running organization-wide management processes, data can enable process development. Using the research framework (see Figure 2.6), data can be used for building capabilities, automating processes to make better decisions as suggested in the literature review. In this maturity phase, data can be used to manage the business effectively compared to just having necessary data. Company employees have to interact with multiple data sets. For example, in this phase, customer information can be updated and integrated into transactional data sources. Creating this kind of information interaction asks new skills from the employees. This creates enormous pressure for IS to be easy to learn, documented and in short to be usable.

Usability was one of the main barriers in the utilization of data in the enterprise context. As complexity and amount of dependencies increase the skills needed for interacting with new information are needed. In this study, the lack of IS related know-how was found to cause barriers in utilization. Bad usability was due informed to be from "poor applications", but it was
more driven by the gap between existing capabilities and a new way of performing functions tasks (see Figure 4.2) Usability can be seen more as how easy it is to learn, document, use and develop data-driven process and IS. In other words, usability in this research, is the ability to interact with the data set to perform a function. Highly usable systems have been used for years, and dedicated personnel had a deep understanding of what kind of data these systems entail. Understanding data usage and implication coupled with on-demand availability of the data can result in new information products. For example, case organization made sizable savings when utilizing staff data set with recurring billing data and cutting unnecessary contracts.

Still, it is clear that business process development underlies how data is developed to be a strategic asset. Understanding how data flows are changing and what is the velocity of new data, even master data, is necessary for building processes. Velocity is linked to the timeliness aspect. As experts do not have the right data at the right time, the process becomes fragile. This can be noted in situations where resources are cut absurdly, for example, when one expert leaves the company or get long sick leave. The velocity is too high for the rest of the function to handle, and inputs are not recorded into IS. Matching business process maturity, resource allocation and timeliness of the data is new competency that organizations have to develop. The finance function has a good example of high maturity function that has clear governance as reporting for management and stakeholders has a long history and legal obligations. Still, enterprises can have more valuable business processes depending on their context, which should be modelled and highly managed. In data-driven applications optimization can bring huge competitive advantage due to the operational lever \(^1\).

The third barrier relates to the capability to create value-adding features to the existing data. In the results this was coded as Capabilities Gap (see Figure 4.2) These can be enriching, such as combining master data with transactional data or even monetizing the existing data. To build strategic assets, data has to be integrated into the business context and have to be timed right based on the volatility of the data (see Table 2.2). The results of this research indicate that managers or data owners have to have an understanding of the business process, but also the ability to roadmap IS and process development in the whole enterprise context. In highly data-driven functions manager needs to have both technical understandability and contextual usability to utilize data for decision making, automation and building capabilities. This needs a clear strategy and tactics from organizations on

---

\(^1\)Operational lever underlies situation where scaling operations has diminishing variable costs. This is common in data-driven business processes.
how to seize emerging opportunities to innovate new services (see Figure 5.1). Data has to have an organization-wide strategic plan for asset utilization on top of functions can build their own roadmaps. At the same time, the IS procurement decision is made outside centralized IT and in functions or BU. Having own master data set inside functions develops a two or multi-sided network which the data source aggregates. To govern this data organization-wide for second order effects needs for new capabilities and new ways to manage the information resources.

Last clear finding from the study was needed for building new capabilities. As noted managers and experts working landscape is shifting due the data-driven applications. Managers need to search for more effective solutions to perform functions continually. At the same time, experts need to learn to utilize these applications and organizations needs to create capabilities to rearrange functions based on the RW data. Following the research framework, it is clear that to utilize data, enterprises need to understand the RW context. The underlying business models and data models have to align to match the data flowing to the company. This means typically that organizations need to understand their customers' behaviour or product usage better. The case organization, for example, had used a CRM system to capture customer behaviour. Case organization still lacked capabilities to capture the data. The organization culture on documenting played a big part, most experts did not share information on their task to colleagues if there was not team policy. For example, data and information sharing was done in situations where backup persons were seen as critical. In most MD situations, functions should have the same kind of shared capabilities. Capabilities were also needed in managing and analyzing the stored data for business purposes. Most striking capabilities barrier being that how organizations can translate business opportunities to information products.

5.1.2 To Understand Master Data, Understand the Organization Management Model

RQ2: How the value of master data differs among management and experts?

How data is meant to be used differs between managers and experts. There are differences in viewpoints how specific tasks should be done among managers and experts. This result may be explained by the fact that managers and experts have different views on the value of the data on hand. Managers are in the case organization accountable on how the function performs with little or no control on ways of working or implementing these
functions. In a case of lacking clear data governance model or enterprise architecture, functions have acquired custom IS and developed processes that are different between BU's. These findings suggest that organization management practices and culture have a massive impact on how value is captured. Another remarkable factor that emerged from findings is the allocation of tasks between organization hierarchy, resulting asymmetries in understanding master data value. We found four types of values in our study: Technical value, Operational value, Management value and Strategic value.

Technical value is created into IS and configurations between the different system as well as the documentation and guides. It can be seen as a pre-requirement for building operational ledger. In this view, technical value is linked to the base of building competitive advantage through data utilization, as seen in Figure 5.1. Technical value can only be utilized with the capability to leverage data to perform operations. IS has typically intrinsic technical value\(^2\), that is developed to match the business operations better. Experts are responsible for creating the technical value. According to the study, experts highlighted that managers do not understand what capabilities are needed for building technical value.

In many BU’s IS acquiring decision was made by the manager, rather than leveraging existing systems, new systems were acquired. Our study found that in the case organization, experts commonly felt that they did not have enough technical understanding of the IS to be comfortable with their most challenging tasks. On the other hand, managers views on value-adding features of new IS are rarely met as configurations, integrations or data transformation to the system takes longer than expected. As the technical level and complexity increases managers hardly know how long performing the task will take. Having big and multifunction teams had most asymmetries on technical value. The same effect was seen in the team which were controlled by upper management. However, with small sample size, caution must be applied, as the findings might not be generalizable.

Managers see the creation of technical value as IS implementation. This view is resource-driven, where implementation per se creates value. According to the resource-based view, this only creates value temporary competitive advantages as configurations and systems designs are easily mobile (see Figure 2.1). Experts, on the other hand, sees the creation of technical value as the capability of creating and maintaining robust systems. If management acquires a system that has not transferable capability, capturing technical value needs investment in training or recruitment. Management should maximize the technical value of data utilization with building on top of existing

\(^2\)The reason why the system is bought or created, is to bring value-adding resource.
CHAPTER 5. DISCUSSION

capabilities and tools, if possible. At the same time battling the question, is the opportunity of operational lever high enough for getting technical value needed. In this question the shared understanding of existing skills and capability and potential benefits of the operational ledger should be evaluated. This is supported by the findings, that manager and experts have different views on what is the source of technical value or how it is captured. Both parties should have a shared understanding of possibilities and threats, and to commit to the decision.

The technical value is focused on effectiveness and therefore how the enterprise can perform operations with least effort. The operational value is focused on efficiency and therefore doing correct actions on the correct time. It is defined by the operations that the organization has detailed. In other words that means what kind of practices are detailed on some level. In function or team level there were not many detailed processes and the operational value was driven by knowledge workers expertize. That makes it hard for managers to understand how resource intensive some tasks are as managers only see the end products of work. Best examples are situations where a manager can clearly see tasks as function. In this kind of situations inputs and end products are detailed in the organizational context. Managers need to have substantial understanding of the task to know how intensive it is. In the case organization, this difference of understanding the complexity of tasks was high in teams where managers had experts working with really different functions.

It is also notable finding that, data-driven task are really typically functional by nature. The operational value should be linked to functions or enterprise level process. Enterprise process should highlight different functions and data streams to the end product. As found in interviews the enterprise context is less dynamic compared to functional units. This leads to situations where processes are not updated as functions are updated. It is typical for experts to hold all information on function changes and on how the process has changed as a result. This leads to a management problem as a lack of documentation and knowledge sharing will increase the speed of making the task, but lacks control of seeing what really happens in the task. If function dependencies are not seen in enterprise process, process variation will increase leading to more inadequate value. From an organizational perspective it is crucial to have managers who understand when prioritize speed over operational value. For example, that might be a great choice when making prototyping or testing hypothesis on master data.

Management value is one of the most complex findings due to the breadth of use cases. Data can be analyzed to make adjustments, not just to function where it is generated but also into the wider organization, as the data
can hold universal truth shared on organization-wide. Still it is imperative that management understand how the analysis is made as they are the basis of the decisions. If the decisions are based on standard reporting that is available throughout the organization, experts can see how the management justifies their reasoning. The research findings indicates that experts often did not understand how the decision was made or how they are reasoned. This lead to dissatisfaction and some experts even quoted that they prefer to choose the way they handle task, even though it might be against company policies or best practices.

Management has the ability to use data to manage employees, optimize structures and gather insight for analysis. Management value is utilized when decisions are implemented with high certainty and they create forecast results. Higher the management value the higher the decision accuracy and ability to change decision-based on better analysis of existing or new data as well as implementing corrections as new data emerges. Management value is created when managers combine their shared competencies with data analysis. This creates a need for decision support function that provides fast-paced analysis on the hypothesis that management wants to test. In case organization management support was not structured. Data that management used were typically asked by experts based on their task with really low formulation what the questions are that data or analysis should be answering on. Using data as knowledge transferring boundary object for decision did not work in the case organization as management does not have a clear understanding of how data is gathered, managed and updated.

There were multiple examples of a situation where managers view on data was different compared to the experts view in terms of quality, usability and trustworthy for decision making. The biggest obstacle was knowledge infusion from data to the expert to manager, when situation change. Experts had provided managers with a dataset that was not correct anymore. Communicating the changes was seen really troublesome and often experts do not even know to communicate the changes as they did not know why the dataset was exported.

Another remarkable finding was that managers typically used experience rather than data as making decisions. This also implies that an organization has to define what is the level of data-drivenness that is needed for making decisions or at least how the decisions are documented. Transferring knowledge and analysis was seen to be the biggest obstacle in the utilization of management value.

The last type of value categorization that was used for grouping differences was **strategic value**. Strategic value comes from the ability to manage uncertainty and adapt to future events for organizational benefit. Using
this narrow definition, it was possible to find clear differences in how data was used for creating strategic value between managers and experts. In the case organization managers saw that strategic decision is made by the top management and the value of data comes from making better organizational decisions. This was also true for managers that were members of the board of executives. On the other hand experts saw the strategic value more driven by learning, understanding and changing behaviour. This view can have intrinsic strategic value for the data, if having some important dataset that can be beneficial in the future. One most common example is the customer relations management systems where accumulating data from customers are seen as a strategic asset. From managers perspective CRM system or another type of IS can be seen as a tool for making decisions and forecasting. For example, what products customers buy by segment or which accounts are most profitable is data that helps managers to formulate the strategy.

Experts on the other hand focus more on capabilities. For example, what kind of problems we have to solve for customers, how to collect information in a systematic way and what new things customers will want. Experts can slowly learn new customer needs as data and information changes. In the case organization, great learning from the experts’ point of view was how to change the underlying process or IS to handle better the customer needs. As the master data sets are the most valuable internal data resources that the organization has by definition, many managers ask experts to provide ad hoc answers. This results in one of the biggest communication issues in the organization. Managers do not provide urgency or value-based arguments in case organization why these answers should be provided. Experts see them as extra work and provide typically only data or indefinite answers. This arises certain understanding difficulties as managers do not understand to ask relevant questions on data quality. Understanding what the velocity of data is, is crucial. Managers should be able to understand how often master data objects are inserted, updated or removed. Experts on the other hand should have transparency in what kind of decisions are made with data. This gives them a possibility to redefine the granularity of the provided dataset or analysis.

Second strategic value view difference that caused problems in case organization is the variety of data. Even though master data has a low variety compared to other categories of organizational data (see Table 2.2), understanding the variety is necessary. For example, in case organization master data object customer had little variety with object user. In many services the word was not used interoperable which caused confusion for creating metrics. In large organizations creating uniform master data set is troublesome if the various components strategic value has not been understood. Experts typ-
## Table 5.1: Different views of data value between managers and knowledge workers

<table>
<thead>
<tr>
<th>Type of value</th>
<th>Managers view</th>
<th>Knowledge workers view</th>
<th>Value driver</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical value</td>
<td>Value is created by the effective systems</td>
<td>Value is created by managing system and processes</td>
<td>Operational lever</td>
</tr>
<tr>
<td>Operational value</td>
<td>Value is created by completion of the tasks</td>
<td>Value is created by doing the task right</td>
<td>Efficiency and quality</td>
</tr>
<tr>
<td>Management value</td>
<td>Data has value for validating assumptions</td>
<td>Data should be the basis of making assumptions</td>
<td>Decision making</td>
</tr>
<tr>
<td>Strategic value</td>
<td>Value comes from creating data resources</td>
<td>Value comes from building data utilization capabilities</td>
<td>Organizational alignment</td>
</tr>
</tbody>
</table>

Critically have no view for different varieties of data and data can be gathered in multiple BU. In case organization managers could not gather data if they had no personal understanding of who controls which variety of data. For example, in a situation where manager would like to know which customers are the less profitable. To find these in formations she would have to look for multiple datasets and experts to make the analysis. At the same time, many information needs have short time windows on which the information is needed. Having no clear enterprise architecture for the data flows in an understandable format for both managers and experts was a big communication barrier. In case of the strategic value of data, organizational alignment and understanding what are the most valuable objectives are key drivers as this value will only realize in future.

In Table 5.1 summarizes the differences that managers and knowledge workers have in the case organization about the value of data. An interesting finding was that managers tend to think more in the short term than experts doing knowledge work. This might be due to the organizational context and culture, but still interesting finding to be explored further. On the other hand, experts tend to focus on task and context at hand and trust more on the data. Managers typically trusted more on their own judgment and saw data more on validating assumptions. It is clear that moving to a more data-driven structure the views must align. This brings strategic value into
focal point building data-driven organization. It seems that data flows and
business models should be hardly linked to each other. For everyone the
organizational strategy should be clear as some knowledge workers might
make big changes to the business model. Even though a master data sets
has independence from the others, but a set typically has high dependence
on organizational processes that are then linked to other master data sets.
Having value differences aligned with the business and operating model of
the organization is the starting point for utilizing data for managerial and
technical value.

5.1.3 Create Boundary Resources to Facilitate Data Utilization

RQ3: How management can leverage data for competitive advantage?

According to the resource-based view, an organization can have sustained
competitive advantage if it has resources that are valuable, rare, inimitable
and nonsubstitutable (see Figure 2.1). Even though data is easy to move
and capture, it typically has intrinsic value and it is unique as it has to
be structured with metadata. Metadata has structure that support orga-
nization’s processes and common objects such as customer have different
metadata structure based on the organizational context which makes it non-
substitutable. Like we had argued in previous chapters also high variety and
velocity of data creates barriers for competitors to imitate processes, but also
utilization barriers for the company having the data resource.

Data utilization barriers in the context of collected data are linked to
data quality. Data quality is based on an understanding of how data is
structured. This creates the barrier for imitating a dataset as the metadata
structure is typically linked into the operating model of the company. In the
case of master data it is also nonsubstitutable, as one cannot change basic
objects without changing underlying business models. Understanding data
architecture in an organization context comes from understanding what good
quality data is. Data quality should be linked in the function quality and
operational quality. We found that managers can leverage data more easily
when all workers have a shared understanding of metadata, IS documentation
as well as shared tools.

These are strategic choices on how to utilize data. Understanding the
scalable solutions on an organization is a new strategic paradigm to con-
sider while building capabilities and organization. In master data use cases
every party has a big responsibility when utilization scales. For example,
CHAPTER 5. DISCUSSION

if the invoicing process has inefficiencies it might not be a problem when
an organization has low velocity or variability of invoice types. As velocity
or variability grows, the cost of utilization problems scales also. The fast-
growing function can create huge cost just as the data management process is
not sufficient for scaling the function. Therefore management should leverage
the data from a strategic viewpoint.

In a data-driven enterprise, ways of working and organizational culture
should emphasize the utilization of data. Data should be the main source of
decision making and capabilities should be based on that. Having a data-
driven organization seems to mean also that the organization sees the mea-
measurement of functions and working as a core habit. This brings focus to the
organizational culture. Management should build on a cultural foundation.
If the culture is not measurement-driven, utilization data can be trouble-
some as the value of measuring things is not part of organizational DNA. As
it is common on all digital transformations biggest transformation is in the
culture.

Still managers can build competitive advantage by creating boundary
resources that hold the VRIN-attributes. Boundary resources, in the organi-
izational context, are resources that management uses to govern and link
business units together. They are shared knowledge resources that build
capabilities while using and aligns the organization. A good example in mas-
ter data case is that all master data should be documented in the same way,
accessible with same and shared tools across the organization. Also the meta-
data models should be documented coherently and in organizational fashion.
Having these ways of working, builds create a shared understanding of the
whole organization which makes them extremely valuable. They are rare and
inimitable as they are embedded in the organizations operational model. By
default boundary resources have to be nonsubstitutable as if their meaning
is not understood or applied it diminishes the value organization can cap-
ture. We conclude that management can use boundary resources to create
competitive advantage.

In Table 5.2 different boundary resources are mapped to the DIKW-
hierarchy. Data model and IS that organization has been linked to boundary
resources that enable the usage in an organization-wide context. Biggest
opportunities for competitive advantage is where the speed to utilize data is
high. Even though master data had low velocity, some attributes changed
while data were transformed between tools. Having consistent tools across
the organization, not just on the functional level, makes analysis faster and
easier to share. Metadata models and shared documentation styles help
understand data variety between information system.

Understanding the speed of utilization in the organizational context is a
<table>
<thead>
<tr>
<th>Knowledge Hierarchy</th>
<th>Organizational Context</th>
<th>Data Utilization Barriers</th>
<th>Boundary Resource</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>Data structure and Databases</td>
<td>Data quality</td>
<td>Metadata, IS documentation and Shared tools</td>
</tr>
<tr>
<td>Information</td>
<td>Data model and Information systems</td>
<td>Analysis, Integrations, Transformation</td>
<td>Enterprise data model, APIs, Reporting practices and Task descriptions</td>
</tr>
<tr>
<td>Knowledge</td>
<td>Processes and Enterprise architecture</td>
<td>Data Governance, Security, Resourcing</td>
<td>Organizational structure, Communication platforms, IS roadmaps</td>
</tr>
<tr>
<td>Wisdom</td>
<td>Utilization capabilities, Strategy and Decision making</td>
<td>Building capabilities, Information Asymmetries</td>
<td>Culture, Ways of working, Strategy and Compensation</td>
</tr>
</tbody>
</table>

Table 5.2: Research theory on how data can be used for competitive advantage in the enterprise context
necessary element on building competitive advantage as more information, knowledge and wisdom is created when data is utilized. Individuals can utilize data to wisdom through learning by using individual skills and abilities. In team level matters how they transfer knowledge to each other, the whole team learns. Management can use boundary resources to guide what is learned and in what context. This was remarkable finding that teams worked best together when they used the same set of data and therefore understood the metadata model and used the same tools. In the case organization many units have their own technology tools and systems, tasks are documented differently. A good example of this is the different data visualizations and analysis software between groups. Other groups visualizations could be unusable as they cannot be modified as the other group do not have the specifics tool.

The organization has tremendous potential to utilize learning when the linkage between data-information-knowledge-wisdom is not broken. Wisdom should stem from enterprise values, culture and ways of working. Knowledge is accumulated to reflect the organizational structure and ways of communication. Knowledge is used to do changes as wisdom is used to judge if the changes are correct for the organizations. Information is gathered to analyze what happens in the enterprise in a given moment. Data is more of the raw material that enables experiments and information creation. If an organization can align the whole workforce to use the same data sets, the learning happens based on shared understanding. Management can use knowledge to lead the workforce and with data-driven goals workforce can track their own progress in real time. Goals can also change rapidly if the organization has shared the understanding of how goals are set and measurements are calculated. Measuring gives more responsibility to the employees as the goals can be easily quantified and performance evaluated across the organization. This also diminishes the information asymmetries in organizational functions and structures.

In the case organization it was clear that data has not used organization-wide to lead employees or it was not integrated into the leadership system. This is assumable common situations in most enterprises. Management should analyze the competitive landscape of what kind of boundary resources competition is building and acquiring. Most importantly management has to make strategic decisions on how to change the enterprise to data-driven. Making the change has costs as systems need to be integrated, data cleaned and processes mapped to the new data-driven operating model. On the other hand clear value adding structures that create enterprise values also needs to be confirmed as a business case. Management can use the formulated research theory for building and scaling the data utilization (see Table 5.2).
It is clear that highest return comes from transforming the way of working and culture to support capabilities building. To lower the risk of failure in utilizing the data at least the found barriers should be addressed. This can be hard as lack of research on implementation practises and understanding of opportunities can make the focusing of the investments on data quality unclear. Implications of the lack of research and how enterprises should tackle challenges in practice in the following chapters.

5.2 Theoretical Implications and Suggestion for Further Research

In this thesis the main objective was to find ways to build competitive advantage with data resources and data utilization practices. The literature research focused on understanding contextual data quality, data creation and utilization processes and resource-based view theory of competitive advantage. The literature focuses on idea how to capture maximum value from enterprise data sets. Understanding contextual data quality is necessary as the value of data is really hard to evaluate without context. Contextual data quality is the data’s fitness of use in the organizational context. Combining contextual data quality with the theoretical framework of building capabilities, decision support and automation creates a framework for building competitive advantage (see Figure 2.6).

One clear suggestion for further research is how to build capabilities from data. In the enterprise context most valuable data sets do not run in automated systems and are therefore, strongly curated by experts. A natural progression of this work is to analyze ways on how experts interact, learn and use data in the enterprise context. This study has identified that contextual data quality barriers are dependent on how data is managed, how the data value is embedded in the organization culture and how using data-driven processes are encouraged in the enterprise. Considerably more work will need to be done to determine what are the core resources of the data-driven enterprise and how to implement them. In this thesis we used VRIN-model to identify valuable data sets, but as data is an intangible resource and the implicit parts are easily migrated to new systems and even organizations, there is a clear need for a more knowledge-intensive model for knowledge and data resources.

One interesting model for further research can be boundary resources that are contextual representations of knowledge resource being that data, information, knowledge or even wisdom. Boundary resources are typical
in platform based enterprises where they provide access points for external parties [Karhu et al., 2018; Ghazawneh and Henfridsson, 2013]. A typical example is API that external developers can use only using documentation on the enterprise website. This thesis underlies that enterprises have many boundary resources that by leveraging them an individual can learn, create value and optimize systems. Building boundary resources could build lasting competitive advantage for companies if they enable faster learning cycles, higher value capture and accelerated resource creation.

If management is able to guide organizational learning and building of highly valuable knowledge resources the knowledge creation capability can be a source of competitive advantage. In the study we found multiple different kinds of data-driven boundary object such our ways of working, IS, technical tools and documentation that guided learning. This creates new dimensions to study in organizational learning and development. Organizations can create an intellectual property by creating learning tools and practises that enables the data accumulation from customers, products and business processes. Combining this with fast moving, complex and uncertain environments, enterprises have to have capabilities to leverage the boundary resources for sense-making. This can only be achieved with learning and building new knowledge. A further study with more focus on data-driven learning is therefore suggested.

The findings of this thesis also suggest that context-aware data is key for insights but also for accumulating tacit knowledge. As data is structure is explicit, but data in context has high ambiguity on knowledge creation processes as knowledge workers try to understand the underlying meaning. There is a clear need in knowledge management to create models on how tacit knowledge, know-how and experiences are used to lower the ambiguity levels. Also these findings argue against Hansen et al. [1999] point of distributed knowledge management, at least in data-driven organizations needs to concentrate their knowledge into central repositories. A clear finding of the study suggests that building organizational context supports the organizational from the data. Research does not conclude what the beneficial ways to contextualize enterprise architecture to support building data resources are. Future studies on the current topic are therefore recommended.

5.3 Practical Implications

This study underlays the importance of data in the enterprise context as a tool for competitive advantage. There are two main findings that have clear practical implications for enterprises. The first finding suggests that
organizational complexity and lacking clear management model on how the data is used for enterprise management hinders data utilization potential. The second finding is that if the value that can be captured from the data is mainly dependent on contextual quality rather than intrinsic data quality. In enterprise this means for example that having access to company master data sets is more valuable than data being correct.

As enterprises scales its operations, the number of incidents also increases. To balance the surge of incidents, enterprises typically create processes to ensure quality levels. This creates a rigid structure for the organization but makes organization slower for change and makes innovation harder. Creation of processes and other organizational structures increases complexity. Management has to make a deliberate decision on what level of shared understanding all stakeholders should have and what are the shared ways of working in the organization. Because the scaling creates complexity, an enterprise should focus on building context and management model to be more flexible. One core idea of context building is to focus on outcomes. Knowledge workers should know why some processes are made as well as managers. If processes are understood and the outputs measured, knowledge workers can innovate and see how changes in function creates the outcomes more effectively. Therefore the enterprise should learn to define and measure outcomes. Outcomes should be clear objectives that are aligned to enterprise strategy. As the outcomes are defined knowledge workers and managers should have a shared understanding of what the outcome is, not necessarily how to get to the outcome. The manager should be able to measure progress. One practical way to do this is to create metrics that measure progress before monetary metrics such as return on investment.

Ries [2011] argues that an organization should take into account the learnings while accounting or comparing projects. In essence this means that in uncertain business cases it is important to test assumptions of the business cases, before fully committing to those. To utilize data, organization should reserve resources for small projects that can be utilized quickly. Enterprise could create multifunctional taskforces that tackle emergent data utilization possibilities. If the focus is on learning and developing new capabilities, the organization can also use the project to create or enhance boundary resources. When contributing and leveraging new and existing boundary resources organization creates new knowledge and capabilities while gaining more certainty to the business case. In summary practical implications that enterprises can do to remove utilization barriers and enhance capabilities:

- Build organizational context in which experimenting is encouraged to build data-driven capabilities
• Make measuring and contribution a core habit of the organization

• Start with small projects and create innovation metrics to track progress

• Create multifunctional taskforces that understand different contextual aspects of business

• Create boundary objects such as common documentation structures and shared tools to communicate the organizational context

Extracting value from data seems to be driven by process quality rather than intrinsic data quality in the enterprise context. Having clear value drivers and processes can be mapped both in "bad and good" systems and they are usually subjective. Bad intrinsic quality data can be really valuable if it’s from a trusted source and the limitations are understood in the organization. After the value capture begins and the processes related variation and issues emerge due to the intrinsic quality, there also emerges the need to clean the data and data generation process. Therefore organizations should start with processes that are already detailed or easy to automate (see Figure 5.2). To understand possibilities, managers should educate themselves and seeks ideas from inside and outside of the company. Typically easy to automate and low productivity task that has low variations can be found from integration providers. These solutions call for the use of the data governance, but yield quick wins if EA is in place. Hendricks et al. [2007] found that enterprises that invested on IS got a high return on assets and the return increased over time. On the other hand it was evident that sound systems do not fix inadequate processes. Building capabilities to develop these processes should be prioritized before the acquiring of the new IS. This was a clear in the studied enterprise, that many systems were acquired before it was clear what processes they were to handle. Outlining of the processes might have been in place, but the know-how was not present.

Experimenting should be done with shared tools before adopting an IS. The most task can for example be modeled in spreadsheet software first and when a definite value is captured only start to automate. This should be a tactic for the utilization of existing data in low variation use cases with high productivity as presented in Figure 5.2 right bottom corner. When the process is detailed it is possible to automate and integrate to the master data sets to scale it to the whole organization. A common example is the digital marketing revolution where detailed customer acquisition processes are combined with CRM systems. Hard to automate the task with low productivity are clearly not whole organization task and are typically sourced to BU functions.
CHAPTER 5. DISCUSSION

<table>
<thead>
<tr>
<th>Hard to automate</th>
<th>Detailed processes with high variation (e.g. customer service)</th>
<th>Fuzzy process with high variation (e.g. innovation and strategy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy to automate</td>
<td>Using existing data again, no variation (e.g. integrations)</td>
<td>Detailed processes with low variation (e.g. lead generation)</td>
</tr>
</tbody>
</table>

Low productivity                                      High productivity

Figure 5.2: Automation difficulty mapping for internal processes

Lastly organizations high productivity, fuzzy processes with high variation as strategy and innovation processes can only use data to make better decisions and faster process cycles. In that category the shared values and ways of working come to an essential part. Here the understanding of opportunities comes crucial and fast validation of assumptions. Here are practical suggestions on how to utilize existing data based on the research:

- Seek opportunities actively both inside and outside of the company to actively leverage existing capabilities and learnings
- Do a data inventory and start with existing data
- Map business processes on the level of automation or based on the value of data accumulation

5.4 Limitations

When designing a case study of a complex and multifaced organization, there are many emergent and hiding aspects that cannot be seen before the inquiry is done. Understanding also adaptive and changing dynamics can be challenging. To take this notation into account the research focused highly on
the predefined research framework (see Figure 3.1, that focused on looking data utilization as an organizational system for competitive advantage. According to Agostinho [2005] defining framework gives the researcher more focused viewpoint and more validity for the case study. To further evaluate the validity of the study four criteria of qualitative research validity can be used: credibility, transferability, dependability, and confirmability [Lincoln and Guba, 1985].

Credibility refers to confidence in the truthfulness of the findings and result [Lincoln and Guba, 1985]. To ensure the credibility of the research, it had multiple phases and the engagement period with research participants was prolonged therefore. Due to the sampling and multiphased research approach, we were able to include the all stakeholders in the case organization to minimize credibility issues resulting from subjective viewpoints. Participants had individual interviews after the group interview phase, so they could elaborate important items that they were unable to say in the group phase. Some participants had their supervisor also in group interviews so that might have lower the credibility. That was also the reason why participants in the second phase had a semi-structured interview structure. These interviews were recorded and transcribed and triangulated to the organizational documentation of the data set and other interviewees. The master data model was also checked with participants after initial results to ensure credibility. One liability that effects were different management models in functions and views on how data utilization was understood and lack of shared vocabulary in organization-wide.

High contextuality makes transferability complicated in qualitative research. Transferability describes the applicability of the findings into different contexts [Lincoln and Guba, 1985]. In the case studies transferability of results is weak due to the individual data point and contextuality Patton [2015]. To help transferability this research provides a description of the case organization and the research design. Results are also direct quotations from the interviews to give higher transferability. Even though results are not easily generalized, the method of structuring the issue is. Other researchers can use and modify the research model to different organizations to find new data utilization barriers and boundary objects as data utilization challenges are common for every large organization.

Dependability refers to the consistency and coherence of the research to ensure repeatably [Lincoln and Guba, 1985]. The research project was highly systematic and well documented. All phases were documented and triangulated as presented in the methods chapter. Still, only one organization was studied in the context of data utilization, that had a highly educated workforce of knowledge workers. Findings were coherent between mapped roles
CHAPTER 5. DISCUSSION

of participants. To increase dependability and validity systematic paradigm suggested by Creswell [2013] was used. This paradigm emphasizes the use of triangulation, member checking and the audit trail. Auditing can also ensure the confirmability of the research process.

Confirmability refers to the neutrality of the research data rather than the researcher’s personal construction. This means that research should free of biases, motivations and prejudice of the researcher. [Lincoln and Guba, 1985] In the time of research the researcher was employed by the case organization and had worked there a few years before the research project. This gives the ability to understand enterprise context more thoroughly, but also gives bias for the researcher. That is one core reason why the research was highly systematic and emphasized an audit trail. Most interviews were conducted in Finnish so there might be some hidden biases in translating the text to English. Still all interviewees were able to do interviews in their first language (English or Finnish). Direct quotations were also used to give more voice to the points that participants wanted to emphasize and give readers the ability to validate the confirmability.

This research was really context-dependent and focused on the emergent field of data studies, management and strategy. Many terms and concepts have not yet used in the field of management and strategy and are still being formulated and data studies have focused mainly on technical aspects of data utilization techniques. The case method limits the generalizability, but has been an effective method on doing research in cross-field domains. On the other literature search has been wide and many are from different domains therefore.
Chapter 6

Conclusions

The research focused on how the company can utilize data that it collects from the surrounding environment for sustained competitive advantage. Typically data is used internally to make processes more efficient and to leverage economies of reuse with little marginal costs. This is the core idea in master data utilization, where enterprise creates shared objects and data sets to be used organization-wide. On the other hand, the enterprise should create new knowledge and capabilities to sustain gathered competitive advantage. This study set out to find how enterprises can gain a competitive advantage while automating existing processes, making more detailed decisions and building capabilities.

This study has identified data utilization barriers, boundary resources that helps enterprises to utilize data and learn from it and strategic tools for enterprise management. The findings reported here shed new light on contextual data quality research and confirmed many Strong et al. [1997] earlier finding (see Figure 4.1). A striking finding in the study was that understanding what master data object such as customer means for the organization and how this metadata information was meant to be used. When the usage is not explicit for knowledge worker typically the data production is also not understood and therefore not trusted (see Figure 4.2). This highlights the importance of managed organization-wide data collection and data governance. This data contextual quality cap can be seen as a significant factor why data is not utilized. Second, in prior research overlooked theme were the capabilities needed for utilizing data in the enterprise context. Knowledge worker needs individual skills to use information systems, data-driven tools and analyzing skills. Still, that is not enough as the organization needs to convert data into information and further into the knowledge. This needs capabilities to structurize data to a usable form. In the simplest case, this means documenting knowledge work to utilize data in the organization en-
CHAPTER 6. CONCLUSIONS

tirely. Resources should follow the same structure, available with selected tools and easily understandable to the whole organization.

The research suggests that enterprises should build and acquire boundary resources that enable the use of valuable core resources. In literature, boundary resources are linked to platforms and in the use of shared resources such as APIs, licenses and software tools (for example, Karhu et al. [2018]; Ghazawneh and Henfridsson [2013]). In large organizations such as enterprises, master data works as a platform fuelled by information systems. Using the knowledge-based view of the firm organizations can build valuable, rare, immutable and non-substitutable boundary resources that build competitive advantage. In the research, we found multiple boundary resources that help to utilize enterprise master data, many of which are non-technical by nature (see Table 5.2). Concept of boundary resources gives management new strategic tools to enable innovation and learning while at the same time enhancing the utilization of existing data and enterprise processes.

This thesis has been one of the first attempts to examine contextual boundaries that enterprises have in data utilization thoroughly. The new understanding of what contextual barriers and possibilities enterprises have on data utilization should help to in transforming organizations more data-driven. In all transformation, people are the key to success, it was clear finding that changes cannot be fast, but they might need to decisive. Starting small and building the shared context first and processes second might be the most valuable finding of the research. This thesis also had provided a highly transferable framework that can be used to utilize data in a large organization and enterprise context.


Appendix A

Interview questions
Background

Name:
Job title:
Education:
How long in case company:

Context:

What kind of organization case company is?
How you would describe culture?
If CASE ORGANIZATION would become data driven company, what would it look like / what had changed?
How would you describe the decision making ability of CASE ORGANIZATION?

Boundaries:

What information you need to do your work?
Do you need to confirmation that data is correct?
What are the biggest problems in your team?
What limits the use of X data?

- What tools and services uses your data?
  - Is that important?
Do you run reports on top of the master data?

- Why/ Why not?

What are the most common task that involves using your master data set?

- Could they be automated/ why not?

Who uses your data set?

Is there a clear roles?
How is data quality measured?
Lack of rewards?
Is there inefficient organizational procedures?
Is there a lack of management understanding and active involvement?
Is there a top management support?
How organization structure effects on data flow?
Is there good tools?
Is there a routines?

Utilization:
What this data set tells you?
In what way CASE ORGANIZATION has to change to get more value out of data?

What are the most valuable processes in your function?
In what way you have optimized the process?
Where does the biggest costs come from?
What are the biggest opportunities for this data?
What takes most of your time?
Do you find answers to solutions you are looking for?
In what this data is used?
  • What benefits the better data has?
What kind of processes there are?
Is it easy to test things in CASE ORGANIZATION?
  • Do you get enough resources?
How are the cost of the data management defined?
Is the data gathering hard?
  • If this data sets disappeared how long it would take to remake it?
• Is it even possible?

**If you got extra resources to your team what would you do with it?**

How hard it was to implement the system that contains the data?

• Could it be changed?

• How long does it take?

**Capabilities:**

**What are the most valuable resources?**

**If you get more resources how would you use it?**

**What kind of skills are needed in your function?**

• Where you search for these skills?

• Has these kind of skills inside CASE ORGANIZATION?

**Can you recruit this kind of capabilities?**

**What kind of new technologies there are emerging in your area?**

• How does it change your role?

• What kind of skills your team needs in coming years?

In your opinion is the knowledge in your team available in some codified form e.g.

How organization change has affected?