CHALMERS
UNIVERSITY OF TECHNOLOGY

Data as an Intellectual Asset

A study of how data assets drive value for digital-born companies and industrial companies undergoing digital transformation

Master's thesis in Entrepreneurship and Business Design

MICHELLE FRANSSON
MARTA SADRIU

DEPARTMENT OF TECHNOLOGY MANAGEMENT AND ECONOMICS
DIVISION OF ENTREPRENEURSHIP AND STRATEGY

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2021
www.chalmers.se
Report no. E2021:089

REPORT NO. E2021:089

Data as an Intellectual Asset

A comparative multiple case study of how data assets drive value for digital-born companies and industrial companies undergoing digital transformation

MICHELLE FRANSSON
MARTA SADRIU

Department of Technology Management and Economics
DIVISION OF ENTREPRENEURSHIP AND STRATEGY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2021

Data as an Intellectual Asset
A comparative multiple case study of how data assets drive value for digital-born companies and industrial companies undergoing digital transformation
MICHELLE FRANSSON
MARTA SADRIU

Supervisor: Bowman Heiden, Chalmers
Examiner: Ulf Petrusson, Chalmers

Report no. E2021:089
Department of Technology Management and Economics
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone: +46 31 772 1000

Gothenburg, Sweden 2021

Data as an Intellectual Asset
A comparative multiple case study of how data assets drive value for digital-born companies and industrial companies undergoing digital transformation
Michelle Fransson, Marta Sadriu
Department of Technology Management and Economics
Chalmers University of Technology

Abstract

Describing data as "the new oil" has become a commonly used statement to describe the rising societal and business importance of data. The analogy pinpoints the need to access the embedded value potential of the exponentially growing data produced by, for instance, connected devices. Thus, data has become a new type of property - an invaluable asset which is produced, processed, transferred and stolen. Yet, there is a clear need for better understanding of how to utilize, protect and take control of data assets in order to leverage it. A major problem is the lack of a mutually agreed language or vocabulary in relation to data. This inevitably leads to communicative difficulties, which in turn results in obstacles regarding management of data.

Through a comparative multiple case study, this paper investigates how data can be utilized as an intellectual asset. This includes an investigation of four companies to determine which data assets exist, what mechanisms are being used to control and protect data, and how data creates value. In addition, the four companies are divided to represent digital-born companies and industrial companies undergoing digital transformation. Thus, the study also includes a comparison of the two types of businesses in relation to the investigated topics.

The findings show that there is a lack of structured ways to manage data within the companies, and that there is a need to implement a data vocabulary in order to facilitate data utilization. By identifying which data assets are utilized in the investigated companies, and thereafter assessing them from the control and value perspectives, a Data Asset Framework was created. This indicates that the same type of data assets exist within both digital-born companies and industrial companies undergoing digital transformation, which implies that it is possible to create a mutual vocabulary for data across industries. The Data Asset Framework enables companies to increase their understanding of how data can be collected, transformed and utilized within the company, as well as how control of data needs to be taken into consideration. The major differences between the two types of investigated companies lie in the extent to which data is being utilized and incorporated into the business model, and how well-developed the control strategies around data are. An important takeaway from this study is the correlation between value creation and control, and the significance of ensuring both aspects in order to leverage data as an intellectual asset. By developing a common data vocabulary and establishing a mutual understanding of data assets in the business setting, it is possible for companies to access new sustainable competitive advantages, strengthen their market positions and unlock new value creating opportunities.

Keywords: Data, Data Assets, Control of Data, Sustainable Competitive Advantage, Data Value Creation, Intellectual Asset Management, Data Asset Framework, Intellectual Property, Data Value Chain

Acknowledgements

This master thesis project is the result of a joint internship collaboration at two global companies, conducted during the Spring of 2021. It has been an invaluable experience, and we would like to express our gratitude for this extraordinary opportunity. Our hope is that this thesis will help shed some light on the mystery of data.

We would like to express a thank you to our supervisors at the two companies, who have helped, guided and supported us throughout the project. We are very grateful for the time invested in us and this project, and we consider ourselves fortunate to have had the opportunity to learn from such knowledgeable people in the field. Their expertise and insights have had a tremendous impact on us and will, without a doubt, affect us in all our future adventures to come.

Furthermore, this research would not have been possible to accomplish without the valuable input from all the participating interviewees. We would therefore like to thank all of the interview candidates for their participation and contribution to this research.

We would also like to direct a special thank you to our faculty supervisor Bowman Heiden, from Chalmers University of Technology, for the invaluable guidance, encouragement and patience. This thesis would have never come together like it did without your support.

Lastly, we would like to thank our families and friends for all the love, support and motivation that has helped us carry through and complete this thesis.

Michelle Fransson, Marta Sadriu, Gothenburg, June 2021

Table of Contents

List of Abbreviations

AI Artificial Intelligence

DBC Digital-born Company

DRM Digital Rights Management

GDPR General Data Protection Regulation

ICDT Industrial Company Undergoing Digital Transformation

IA Intellectual Assets

IP Intellectual Property

IPR Intellectual Property Rights

R&D Research & Development

List of Key Definitions

Control Mechanisms - A mechanism that controls an asset, either from a legal, technical or business perspective. Typical control mechanisms can be technical control (e.g. cybersecurity), market power, secrecy, right based property, contract-based property.

Data Assets - Assets in the form of information from hardware and/or software that enables value creation in a firm, either as it is or through processing and/or analytical activities.

Data Resources - A core resource in a digital-born company or a company that is undergoing digital transformation, which enables potential in driving value from data in the business. The resource is in a digital form, commonly retrieved and stored as various data points.

Digital-born company - A company which was born digital, and utilizes or leverages data as a key business resource.

Industrial Company undergoing Digital Transformation - A company which was not born digital nor has fully undergone a digital transformation, but is in the process of digitalization.

Intangible Assets - Assets that are not physical goods or services which can be assets such as knowledge, goodwill, brand recognition, licensing agreements or intellectual property rights.

Intellectual Asset - An intangible asset that can be captured into an intellectual asset, that enables the ability to create business value from the given asset.

Intellectual Capital Management - Management of intangible assets in a business setting.

Intellectual Property - Results of inventions or creative work where these can be captured and utilized as intellectual property rights in form of a patent, copyright, trademark, design right and trade secrets.

List of Figures

List of Tables

1. Introduction

This chapter will present an introductory background in the research field and the research problem including research questions. Furthermore, the introduction will include the purpose of the study, the project scope and delimitations, as well as an overall outline of the thesis disposition.

1.1 Background

The rapid increase of data-driven business models and innovations is generating a global realization of the economic value and functional utility of digital information (Mayer & Ritter, 2018). This confirms the fact that data has become a new type of property - an invaluable asset which is produced, processed, transferred and stolen. This creates a need for better data management within businesses, which is fundamentally based on the following three key background issues:

- *The industrial paradigm shift* - A change in focus from tangible to intangible assets, dependent on the shift from an industrial to a post industrialized economy.

- *Data and its strong correlation to business value* - Organizations' abilities to properly collect, manage, analyze and utilize data is becoming an important source of competitive advantage and business value.

- *The importance of taxonomy for business value* - A shared vocabulary and categorization for use across organizational functions facilitates communication and leads to better analysis as well as value creation.

"Data as the new oil" is a statement that has become commonly used to describe the rising societal and business importance of data (Hartmann et al., 2016). The analogy aims at pinpointing the need to extract, utilize and commercialize the embedded value potential of the exponentially growing data produced by for example the internet, social media, cloud computing and mobile devices. Indeed, with the rapid approach of industry 4.0, more organizations focus on managing and leveraging their intangible assets in addition to utilizing their tangible assets. This way of thinking could be directly connected to the paradigm shift from an industrial economy, where production of goods was the key component in transactions, to a post industrialized economy where development of knowledge and services is becoming increasingly significant (Bell, 1976). To achieve this, organizations must undergo a digital transformation where intangible assets are captured and utilized. Thus, a firm's capability to adapt to digitalization is dependent on its access to data, as well as related permission and ability to process and analyze it (Ritter & Lund Pedersen, 2020). Furthermore, a firm's ability to manage information and extract knowledge is viewed as an essential competitive advantage, and many organizations build their core business on their abilities to collect, analyze and draw insights from data (Cavanillas et al., 2015). Thus, there is a need to extract more value from data and to incorporate this in a business setting, (Xie et al., 2016). There is, however, a considerable difference in the various types of industry settings. In a digital-born company, the ability to extract value from data is the main core of the business since the data is the main product of the company. In contrast, industrial companies have had production of physical products as the main focus for a long period of time. However, digitalization has created a shift in focus to incorporate data in the industrial way of thinking. Thus, being able to manage and utilize data is becoming an imperative need for companies in order to secure their existence and gain competitive advantage. An organization's ability to strategically communicate and share data is impacted by the existence of a common vocabulary of the key business resources (Earley Information Science, 2019). A taxonomy can help companies to structure and organize knowledge, grouping assets based on mutual characteristics and explaining the relationships between these characteristics (Nickerson et al., 2012; Cook et al., 1999). This is especially true for intangible assets, such as data, since these are inherently difficult to define and manage which leads to difficulties in the value extracting processes.

1.2 Prior Research

A literature review was conducted in order to assess and investigate the literature in the chosen research field. The review showed that a large number of academic papers examine different ways of categorizing data resources within businesses. The literature also investigated the business value connected to categorization and utilization of data. Furthermore, the literature review showed that control mechanisms related to data assets were explored to a minor extent. Many of the reviewed articles emphasized creation of a taxonomy for data from many various perspectives. In a business setting, a taxonomy can help organize the knowledge in a field, to facilitate the understanding of relationships between the information and related concepts (Nickerson et al., 2012). There are many articles discussing the relevance of creating taxonomies to better understand the information gathered in a business. However, few of the investigated articles solely focus on how data creates value in their proposed taxonomies. Hartmann et al. (2016), Hannila et al. (2019) and Rizk et al. (2018) are researchers who have focused on developing a taxonomy for data driven businesses to help understand the value of data. However, it was noted that even these researchers focus on different aspects, where Hartmann et al. (2016)and Hannila et al. (2019) focus on the structure and source of the data and implies that this is the value creating aspects. In comparison, Rizk et al. (2018) emphasizes that it is important to capture the activities and utilization capabilities in relation to data in the given taxonomy. Cavanillas et al. (2015) discusses how data can be used to create value for a business and how these activities should be integrated throughout the big data value chain. Petrusson (2004) discusses various perspective in how intellectual property can be used to create wealth and value from an intellectual value chain perspective. Furthermore, several structural control mechanisms are proposed that can be utilized in relation to various intangible assets such as data. To conclude, the reviewed literature discussed the three concepts data as a resource, value of data and control of data, but only to a minor extent. This indicates that there is a clear lack of investigation in this specific field.

The majority of the reviewed literature focuses on fully or near-fully digital companies or technology start-ups. Few investigate the nature of data assets and related value creation within other industries, such as industrial companies. Thus, research related to the differences between digital-born companies (DBCs) and industrial companies undergoing digital transformation (ICDT) is lacking when it comes to value creation and control of data assets. The specific dimensions that will be investigated in this research study include what types of data resources exist, how they create value and how they are controlled within the two previously mentioned types of businesses.

1.3 Problem Statement

Big data is growing exponentially and, simultaneously, data is being recognized as an invaluable asset. However, there is a clear need for better and mutual understanding in how to utilize, protect and take control of these assets in order to gain and maintain the sustainable competitive advantage they provide.

In society today, the resource based perspective is changing to a more digitalized setting. With society becoming more digitalized, data will become increasingly more important for firms to manage. In turn, as more firms begin to utilize and leverage upon data, it will become important to create value from data from all perspectives within a firm (Xie et al., 2016). Many industrial companies in society today have difficulties in strategically managing data, as well as leveraging and utilizing data to a large extent. This is due to the fact that they are still industrially oriented and not focused enough on becoming digitalized.

As data will become more utilized, it will also become more important to find a common vocabulary to enable communication about data in various industries and industry applications. Managing data and speaking the same "data language" could benefit companies by enabling them to leverage and utilize data in the best way possible (Rizk et al., 2018). This in turn can enable more value creating opportunities for companies, where using the same language can help companies to understand the value that data can bring. There is a need to create a common vocabulary, in order for data to be the new oil and to enable greater value creation in firms (Rizk et al., 2018).

1.4 Research Purpose

This thesis will analyze what types of data assets exist, how they create value and how they can be controlled in a business setting, in order to facilitate communication about data. The thesis project will be conducted at four companies; two digital-born companies, referred to as DBCs, and two industrial companies undergoing digital transformation, referred to as ICDTs. The research will therefore provide a comparative analysis between the two types of businesses regarding how data can be leveraged as an intellectual asset. The project will result in the creation of a framework which describes one way of categorizing data assets, as well as the related value driving aspects and control mechanisms.

1.5 Research Questions

The research questions will guide the research and are structured with the aim of fulfilling the research purpose and study, (Bryman & Bell, 2011). The main research question was formed in order to set the foundation for the research purpose. Four sub-questions were developed as a basis to answer the main research question.

Main Research Question

How do data assets drive value for digital-born companies versus industrial companies undergoing digital transformation?

Research Question 1

What data resources do companies have that are important for value creation?

Research question one aims to understand the role of data as a resource in a business setting and how data resources can be described and utilized. The findings of this research question will set the foundation for understanding how data can be utilized in a business setting.

Research Question 2

What control mechanisms can be relevant to use in relation to data assets and why?

Research question two aims to investigate which control mechanisms can be used to control data assets and what value the control mechanisms can bring in relation to data. The control mechanisms used for this question will be structural control mechanisms such as technical control, market power, secrecy, right based property and contracts based property.

Research Question 3

How can data assets create value in a business setting?

Research question three aims to investigate how data assets can create value and which utilization areas that exist for value creation from from data.

Research Question 4

How does the value creation and control mechanisms related to data assets differ between digital-born companies and industrial companies undergoing digital transformation?

Research question four aims to create understanding of how data assets are managed differently in digital-born companies or industrial companies, specifically in relation to how data assets can create value and control in the various companies. This research questions will be a comparative investigation between the two types of companies, i.e. DBCs and ICDTs.

1.6 Delimitations and Scope

The study will be limited to the investigation of four companies, two digital-born companies and two industrial companies undergoing digital transformation. Two of the companies will be investigated more in-depth, whilst the other two will be used to provide a broader perspective and minimize biases. All findings will be derived from the comparative multiple case study of companies as well as a literature analysis.

The project will focus on analyzing R&D data assets, i.e. assets related to the development of products or services. The analysis will therefore not include data assets relating to fields such as financial services or marketing data. Due to time constraints and the complexity of the legal field, this research study will not include an extensive focus on the management of personal and sensitive data.

The thesis will be underpinned by an intellectual property management logic, where data assets will be captured and analyzed in the same manner as intellectual assets. Capturing data assets includes identifying, defining, and analyzing strategic control mechanisms and value creating opportunities. Petrusson's (2016) IAM Framework will be used to incorporate this IP management logic, where the scope will be limited to the claim process. The researchers will touch upon the positioning and deciding phase, but this will not be emphasized in this thesis. In regards to this delimitation, the study will focus on an internal investigation and not thoroughly analyze the external field.

The investigated legal framework for the project will be limited to the EU jurisdiction due to the General Data Protection Regulation, GDPR, which provides strict regulations in relation to data privacy and integrity. Thus, companies operating in the EU, such as the participating companies in this study, must adhere to these laws.

1.7 Thesis Outline

The disposition of the research paper is set to include seven main chapters, as well as references and appendices in the following presented order.

The first chapter consists of the *Introduction*, which includes the background and introduction to the research study, prior research, problem statement, research purpose, research questions, and delimitations and scope.

The second chapter describes the *Theoretical Framework*, which includes the four main fields of theory utilized in the research study.

The third chapter describes the *Methodology*, which includes an epistemological & ontological analysis, research strategy, research method and an assessment of the research quality.

The fourth chapter presents the *Empirical Results*, which includes the main empirical findings from the comparative multiple case study, in relation to the four research questions.

The fifth chapter consists of the *Analysis*, which includes an analysis and synthesis of the empirical findings in relation to the theoretical framework.

The sixth chapter consists of the *Conclusion*, which summarizes the research study and main research findings.

The seventh and final chapter presents the *Discussion*, which includes a discussion regarding contribution of research, research limitations as well as suggestions for future research.

2. Theoretical Framework

This chapter will include and describe key theoretical concepts that will be used in the analysis of the research study.

2.1 The Value of Data

This section will include theory in relation to what data is and how it can be used, and how data can drive value for a business. The theory will be used to analyze how R&D data assets are used in businesses and will be applied and utilized in the comparative analysis of how data is used in DBCs and ICDTs. This theory will lay the foundation of data as a concept and how it can be utilized in a business setting as a source to gain competitive advantage.

2.1.1 Data as a Concept

Our societies are moving into a new era where products and services will become more digitized and the volume of data will grow exponentially (Cavanillas et al., 2015). Data is collected and gathered from various companies products and services, such as connected devices, sensors, internet utilization and mobile device integration. As more products and services become digitalized, larger amounts of data is collected which increasingly leads to an exponential growth of the volume of data, (Hartmann et al., 2016; Cavanillas et al., 2015). This prominent growth and volume of data is more commonly referred to the terminology *Big Data* (Liang et al., 2018; Cavanillas et al., 2015; Hartmann et al., 2016; Kaynak & Yin, 2015). Big Data can be described in many ways, where the current most common explanation is through the three V's; *Volume, Variety and Velocity* (Liang et al., 2018). The Volume characteristic relates to the size of the data, which can range in many various sizes such as terabytes, zettabytes or petabytes. Velocity relates to the speed of the data stream and how fast the data is streamed, collected and generated. Variety relates to broadness of the different formats and array that the data can take. An example of variety is the different devices, applications and interfaces that data can be retrieved from. Some researchers define big data as the four or five V's, where *Veracity* and *Value* is added as two additional characteristics, (Hartmann et al., 2016; Kaynak & Yin, 2015; Cavanillas et al., 2015), see Figure 2.1 for full illustration of the 5V model. Veracity refers to quality of the data both in terms of documentation and the origin of the data. Value refers to the data's ability to transform into business value.

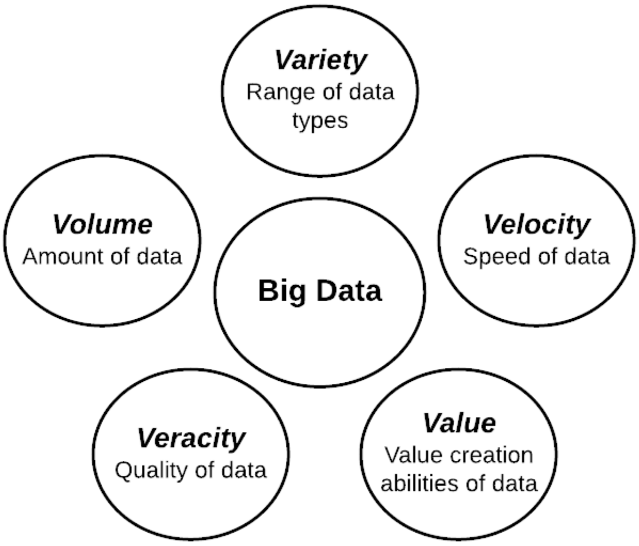


Figure 2.1: Illustration of the definition of Big Data based on Hartmann et al. (2016), Kaynak & Yin (2015) and Cavanillas et al. (2015)

Big Data is estimated to be one of the main drivers in the fourth industrial revolution, which evolves around creating smart cities and factories to produce more intelligent products and services (Kaynak & Yin, 2015). In general, data can be utilized in many various ways and can unlock and create value in many different areas. Examples of such is product and market development, predictions in market demand, efficiency in operations and production of products, more efficient decision making and creation of better customer experience through personalized experiences. Many researchers discuss both the potential of big data as well as the challenges that it implies, where the utilization of data differs from different industry sectors (Kaynak & Yin, 2015). It is predicted that industries within computer, electronic products and information management will gain a substantial value and utilization of big data.

According to a report developed by the Organization for Economic Co-operation and Development (2014), healthcare, public administration and educational industry sectors are some of the industries where implementation of data analytics can have the highest impact. These industries withhold a large portion of occupations that perform manual and not computerized tasks. The industries are collecting a large amount of data and conducting manual analysis and statistical observations, in which data analysis can be helpful to enable higher efficiency. In the manufacturing industries, it has become increasingly more attractive to utilize sensors in the production facilities when developing products to enable Machine-to-Machine communication to optimize the product production. Furthermore, OECD (2014) describes that sensor data can monitor and analyze products efficient and also optimize the operations at a large level, where sensor data can also help to optimize after sales market by enabling predictive maintenance services.

In this research study, the research will be limited to focus on R&D data which will be referred to as all types of data related to the development of a company's products and services. Big data is becoming more incorporated in companies' innovative processes, where companies strive towards utilizing more solutions that are highly dependent on big data, such as including solutions that are either algorithm or software based, connected or IoT (Internet of Things) integrated.

2.1.2 The Definition of Data

Data is a rather vague and difficult term to define and describe (Liew, 2007). It is also correlated and used interchangeably with the term information. Otto & Aier (2013) describe this as a current lack of a clear and accepted understanding of data and information as terminologies. In a knowledge economy that is simultaneously transforming into a digital economy, key fundamental concepts are knowledge, information and data (Liew, 2007). Many researchers have tried to define the relationship between data, information and knowledge and how they interact and can be used in relation to different contexts. Ackoff (1989) is one of the researchers who has defined the relationship between data, information and knowledge as a hierarchy of wisdom, where wisdom is at the top of the pyramid whilst understanding, knowledge, information and data descend in that order, see Figure 2.2.

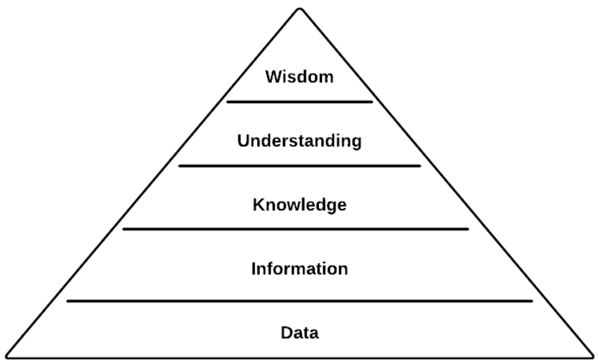


Figure 2.2: Illustration of Ackoff's (1989) hierarchy of wisdom

Ackoff (1989) describes data as *"symbols that represent properties of objects, event and their environment"*. Data is explained as a product of observation, where observing is described as a way to form a sense and draw conclusions from data. Information is described as *"descriptions and answers to questions that begin with words such as who, what, where, when and how many"*, and that is collected and inferred from data. The large difference between data and information is that data does not enable large value if not processed and developed into relevant forms. According to Ackoff (1989), there is a functional difference between data and information, where data is commonly reduced when transformed into relevant information. Furthermore, Ackoff (1989) describes knowledge as know-how that enables information to be transformed into certain instructions. Knowledge is also described as a way to control a system and to make it more efficient. There are two ways to obtain knowledge, either by transferring knowledge as instructions from one person to another or by obtaining the knowledge from experience. Understanding is described as the facilitator and accelerator of learning and adaptation, where knowledge is the base as to why and how a system exists and functions.

Wisdom is described as an evaluation of the understanding and is the top step of the hierarchy. Ackoff (1989) describes the relation between wisdom, understanding and knowledge by explaining that knowledge and understanding can enable intelligence which can enable abilities to increase efficiency, and that wisdom through this can enable abilities to be effective. Efficiency is based on a logical principal which can be applied by being programmed into an automatic system based on principles that are general and not personal, in comparison to judgment. A distinction between effectiveness and efficiency can be identified, where Ackoff (1989) differentiates wisdom from understanding, knowledge and information. Wisdom adds value to the context, which requires human judgment in order to be valuable (Ackoff, 1989).

Liew (2007) describes the relationship between data, information and knowledge as key building blocks in intellectual capital management and in knowledge economies. Data is defined as captured and stored recordings from symbol or signal readings. Information is described as a type of message that contains a meaning or input that can be used to form a basis for decision making processes. Knowledge is described in three terms as the *know-what*, which relates to cognitive or recognitive abilities, *know-how*, relating to the capacity to act, and the *know-why*, which relates to understanding the knowledge that is contained or resides in the mind. The relationship between the three can be described as seen in Figure 2.3, where data relates to information through its ability to become processed, analyzed and converted into information. When data is contextualized, i.e. put in a meaningful context as information, the data is captured and stored. Similarly the relation between information and knowledge relates to information that has been understood and internalized by a human intellect, which enables the information to be turned into knowledge. This knowledge can thereafter be externalized, either through an illustration or verbally, and can also be described as information.

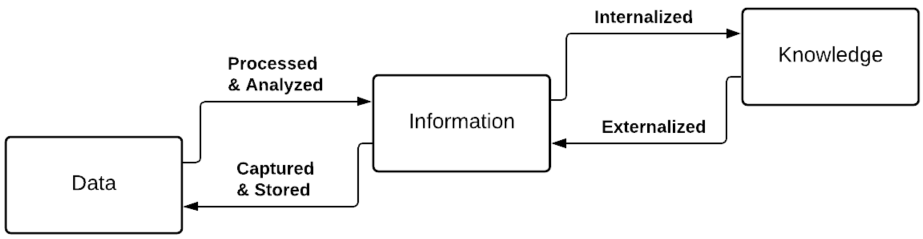


Figure 2.3: Illustration of the definition of data according to Liew (2007)

To summarize, both Liew (2007) and Ackoff (1989) explain the three perspectives in relation to each other, and both agree that data is the first step or building block that must be accessed or retrieved in order to enable value creation from it. Thus, data can be evaluated, processed and used in order to gain more thorough information, understanding and knowledge which facilitates decision-making within a business. This theory will be relevant to use in the research since data is a difficult term to comprehend and there is a need to understand the term in a more contextualized way. It will also be important to understand how data, information and knowledge are integrated when analyzing data assets, since not all types of data can create value from only being data.

2.1.3 Data Monetization

There are many articles that explain how to create value from data in a business setting. This research study will focus on theories and models based on Baecker et al. (2020). The research provided by Baecker et al. (2020) proposes twelve generic strategies for creating value from and monetizing upon data within a firm. The research assumes value to be real and quantifiable, and can be represented as both monetary value and other economic benefits. Baecker et al. (2020) describes this difference by explaining that data can be monetized by either improving or creating processes and products that rely on data more efficiently. This value can thereafter become "real" value through quantifiable, consecutive effects or returns by repeated and increased purchases, reduced costs and/or higher market share. The following section below will describe the twelve strategies for how to create value from data based on Baecker et al. (2020) .

Description of the Data Monetization Strategies

Asset Sales

Asset Sales involves selling data as an asset through three various strategies to third parties. The first one includes direct sales of data to customers to give them complete control of the data. The second is to give paying customers access to data, in order for the company to keep more control of the data. The third is to utilize a DaaS-model (data-as-a-service), sell data to customers based on the customers' preferences. These utilization strategies can create economic value such as additional sales or provisions of data, creating new revenue streams and extending the customer base.

Business Process Improvement

Business Process Improvements involves creating value from data by improving and optimizing the existing business processes. This monetization area has many various strategies that includes how to create value from data. An example of how companies can utilize their data to create improved business processes is to optimize the current processes, to monitor process performances and to create and improve the transparency in the processes. Other mentioned strategies for the data is to use data as a basis for decision making or to improve safety measures in the business processes. This can create economic value such as reduced costs, increased sales, increased productivity, identification of inconsistencies and fraud, and better support for decision making.

Product or Service Innovation

Product or Service Innovation involves creating value from data by developing new products and services. This strategy utilizes data by either retrieving insights from the data and incorporating these into new products or services, or feeding data into products or services to enable extension of the services or new features. This can create economic value such as new revenue streams or new company business segments.

Product or Service Optimization

Product or Service Optimization includes utilizing data as a basis to create value by optimizing or improving the existing products and services in a firm. An example of this is the constant data collection streams from consumers utilizing the company's existing products and services, which enables continuous improvements and optimizations. This can create economic value such as improving the customer experience, improving brand reputation or direct increased sales.

Sales of Data Insights

Sales of Data Insights involves utilizing data as a basis to create value by sales of information or knowledge that is derived from data. This can include both analytical observations, graphical representations or visualizations of the data, which enable better understanding of the insights. This can create economic value such as new revenue streams and new business divisions.

Contextualization

Contextualization includes to create value from data by creating additional value to the customer or to the business itself. Examples of context-related data can be social media data, locations domains and IP-addresses, which can all be used to create value in various ways. An example of contextualization is how companies can recommend certain products and services that are frequently bought together, in order for them to optimize and increase sales. Contextualization does not necessarily utilize data from individuals, but can be applied in a larger context such as specific locations, regions or purchase groups. This can create economic value such as increased sales and price optimization.

Individualization

Individualization involves utilizing data as a basis to generate additional value based on the company value proposition. The strategy creates value for specific consumers or companies on an individual level, by using data to link offers to customer profiles. A company can then, for example, develop customized or personalized recommendations for a specific individual or business, in order to create a better experience of the products or services. This can create economic value such as improved customer experience or increased sales, due to the personalized or customized recommendations or features.

Strengthening and Building Customer Relationships

This strategy includes utilizing data to strengthen and build sustainable customer relationships. This is based on utilization of data regarding customer needs and behaviors. An example is if companies can predict when their products or services will need maintenance, and thus can predict a customer need that makes the customer willing to pay for additional services. This can create economic value such as retaining customers, enhanced customer loyalty and customer satisfaction, increased recurring revenue and increased trust from the customers.

Strategically Opening Data

Strategically Opening Data involves to utilize and open data to other businesses or third parties, to enable more value co-creation or more company visibility. This strategy can be used to leverage data internally across the company to make it more accessible in various parts of the business, as well as in the various parts of the company value chain. However, the strategy can also be used externally where third parties can access internal company data by creating data ecosystems and networks. This can create economic value such as creation of new partnerships, increased company visibility, partnership value co-creation, leverage of business partner capabilities and abilities to share costs between partners.

Data Enrichment

Data Enrichment refers to the companies' abilities to aggregate internally or externally generated data, by transforming or cleaning the data and thereby enabling economic value. An example of this is to consolidate the data landscape at a company in order to make internal data more available within all company divisions. This is particularly valuable for larger companies that might struggle with data-sharing across the company. This can create economic value such as improved internal company value creation, larger availability of data across the company and increased data quality.

Data Bartering

Data Bartering includes to utilize data by exchanging company data in return for other valuable assets such as services, data or other tools. This can create economic value such as new data insights, tools or services that is exchanged in the process.

Data Privacy and Control Guarantee

This strategy includes utilizing data that is gathered from customer interactions, that can be monetized upon, but simultaneously guaranteeing that the customer is in full control of their own data. By guaranteeing this type of control and privacy to the customer, this data can provide increased market shares, increased customer loyalty, and improved brand reputation.

Adapted Version of the Data Monetization Strategies

For the purpose of this study, an adapted version of the above mentioned utilization strategies has been developed, where the twelve categories were converted into five main categories, see Figure 2.4.

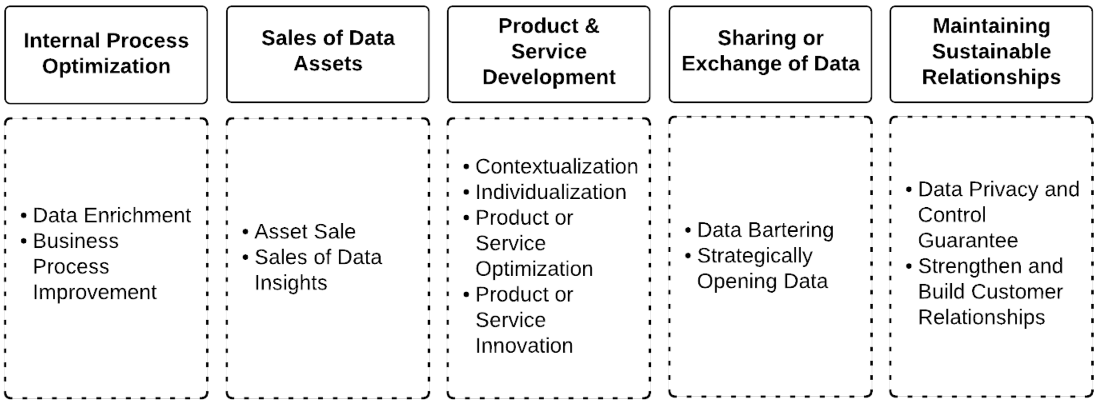


Figure 2.4: Adapted version of the data monetization model provided by Baecker et. al (2016)

2.2 Sustainable Competitive Advantage

This section will describe the concept of sustainable competitive advantage which will be used as an essential piece of theory in this research. Sustainable competitive advantage as a business concept can be defined in various ways and is a research field that has become increasingly explored within strategic management (Barney, 1991; Huang et al., 2015). A sustained competitive advantage can be addressed through two major perspectives; the internally analyzed perspective of resources and capabilities, referred to as Resource Based View (RBV), and the externally analyzed perspective of market positions, called Industrial Organizational theory (IO) (Huang et al., 2015; Powell, 2001).

2.2.1 The Internal and External Perspective of Sustainable Competitive Advantage

There are many different ways to describe what competitive advantage is and how organizations can ensure that the competitive advantage is *sustained* in relation to competitors. The two main perspectives, RBV and IO theory, reconciles that competitive advantage is a firm's ability to establish a superior profit in comparison to its competitors, where the main difference is how the profit is created and achieved (Huang et al., 2015). The IO theory describes that superior profit is achieved by the firm having a stronger market position in comparison to the industry competitors, and that this could be created through, for example, economies of scale. Recent researchers however, tend to focus on explaining the external analysis of a firm's performance with Michael Porter's five force model, which is a theory that has originated from the IO theories (Barney, 1991). The RBV theories propose that sustainable competitive advantage can be achieved through a firm's abilities to preserve resources and capabilities with distinctive characteristics (Barney, 1991; Grant, 1996).

Barney (1991) explains that there are two fundamental assumptions that are used to differentiate these two perspectives, where the theories related to IO perspective assumes that the firms' resources are homogeneous. This indicates that firms utilize and control the same resources in the same strategic way in order to pursue the same business strategy. These theories also assume that the resources and business strategies in a firm are mobile across the industry. In comparison to RBV, the resources are assumed to be heterogeneous and immobile across industries (Barney, 1991; Huang et al., 2015). The research related to which resources that are used to create a sustainable competitive advantage is described to either be focused on a firm's ability to isolate opportunities and threats, describing the strengths and weaknesses in a firm or to analyze how these four perspectives can be used and matched together to choose and develop business strategies (Barney, 1991). With this reasoning in mind, Barney (1991) proposes a framework that describes that sustainable competitive advantage can be achieved by implementing business strategies that exploit the internal strengths by acting and responding to the market trends and opportunities, while simultaneously counteracting external threats and trying to avoid internal weaknesses. This framework, see Figure 2.5, illustrates a firm's ability to utilize both their internal and external resources in order to create a sustainable competitive advantage.

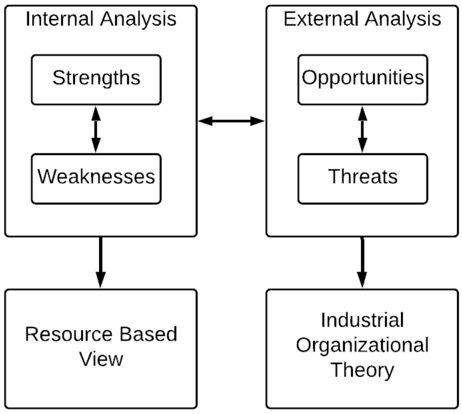


Figure 2.5: Illustration of the framework created by Barney (1991)

This research study will apply theory in relation to RBV as a starting point, since the study will be underpinned by the IAM framework and a comparative, internal study. RBV is assessed as a suitable theory to use since it focuses on analyzing the internal resources within a firm, which also regards and corresponds to the stated research delimitations. Consequently, IO theories will not be included to a large extent, since the research is mostly based on an internal analysis of how data can be used as an asset within a firm.

2.2.2 Resource Based View and Knowledge Based View

This section will in short describe Resource Based View (RBV), but also include the concept of the Knowledge Based View (KBV), which according to Grant (1996) could be perceived as a development and extension of RBV. The KBV is assessed to be an essential theory to utilize in this research, since utilizing data as an intellectual asset could be defined and explained as knowledge and information in the firm.

The RBV can be described as directly related to a firm's abilities to achieve a sustainable competitive advantage by the means and strengths of protecting rare and valuable resources, as well as isolating these mechanisms in order to minimize the risk of imitation by competitors (Wang et al., 2009). There are many ways to define a firm's resources in accordance with RBV. Wang et. al (2009) describe the RBV as a cluster of heterogeneous resources that include operational processes, products, tangible and intangible assets. Barney (1991, p.101) on the other hand, defines the RBV resources as *"the firm's ability to control and enable the resources to develop and implement business strategies that improves the firms efficiency and effectiveness by utilizing resources such as assets, capabilities, organizational processes, firm* *attributes, information, knowledge etc..."*. An essential consideration to have in mind regarding RBV is that not all resources have the capabilities and potential to enable sustainable competitive advantage (Barney, 1991). The RBV perspective describes that sustainable competitive advantage is achieved when superior performance is based on the resources being *Valuable, Rare, Imperfectly imitable and Non-substitutable* (Barney, 1991; Huang et al., 2015). Thus, the resource needs to be 1) *Valuable*, by enabling exploitation and creation of opportunities and/or neutralization of threats from a competitive environment, 2) *Rare*, both in terms of the current resources and the potential competition, 3) *Imperfectly imitable*, i.e. no competitor should be able to imitate the resource and 4) *Non-substitutable*, meaning there cannot be any equivalent substitute that is valuable nor rare nor imperfectly imitable, (Barney, 1991, p.105-106). However, if the resource is "only" valuable or rare, the resources cannot be a source of *sustainable* competitive advantage, only competitive advantage. According to Barney (1991), sustainable competitive advantage can be obtained by achieving all four key attributes of a resource, where the competitors need to find it difficult to create substitutes and/or imitate the resources. If the firm is not able to fulfill all four key attributes, this is defined as temporary competitive advantage (Huang et al., 2015; Barney, 1991).

The Knowledge Based View (KBV) can be described as a paradigm shift from an industrial economy to a post industrialized economy, where there has been a change in focus from production of goods into a larger focus in the development of knowledge and services (Bell, 1976). Industries thereby change the focus to not only being resource oriented, but becoming more knowledge oriented. Spender (1996) describes the theory behind KBV as a firm's ability to think beyond resources and production-functions, to enable more insights and experiences from different perspectives. Grant (1996) explains that knowledge is the core of a firm's resources which essentially leads to the KBV reasoning, where knowledge is viewed as the most strategic resource in the firm. Furthermore, Grant (1996) explains that since knowledge has a specific set of characteristics, it can only be fully owned by the person who possesses the knowledge and not by the organization, and that knowledge can primarily be utilized by the person possessing the knowledge. By utilizing the KBV, this amplifies the importance of utilizing the human resources in the firm in order to gain sustainable competitive advantage, and also highlights the role of utilizing data as knowledge and as the core of the firm to integrate and coordinate around.

2.2.3 Utilizing Data to Gain Competitive Advantage

In this study, data is considered an intangible resource utilized in the knowledge economy. As society is becoming more digitalized, more organizations are focusing on resources that are intangible or intellectual by themselves. With this change in focus, it is assumed that utilizing knowledge and information in the form of data will become more important for a firm to gain competitive advantage on the market. This will also create a need in understanding how to manage and utilize data as an intellectual asset within an organization.

There are many researchers who state the importance of utilizing data as an asset and who predict that firms will increasingly become more dependent on the collection of data in order to develop new products and services, as well as to gain insights about the customers (Hannila et al., 2019; Liang et al., 2018; Cavanillas et al., 2015). Liang et al. (2018) states that utilizing data as an asset within a firm is considered to be one of the main keys to unlock the ongoing digitalization, and is the new strategy to become more productive. Cavanillas et al. (2015) states that big data will be one of the main assets in the future, where a main driver is to utilize data in a firm and to collect and generate data from customers. This data can thereafter be turned into insights through analysis and application of human knowledge and intellect, and be incorporated into the development of the firms' products and services. This type of data gathering can in turn lead to new insights and data collection that other competitors cannot access. This confirms that large volumes of data collection can be used as a new driver to gain competitive advantage. By utilizing data in this way, the volumes of data will increase which, in turn, will lead to data being used as a valuable resource. This enables magnetization of data in digital businesses, and in turn creates a source of competitive advantage, (Baecker et al., 2020). In relation to the above mentioned concepts, i.e. sustainable competitive advantage as described by Barney (1991), one could describe this utilization and collection of data as both rare, valuable, imperfectly imitable and non-substitutable if the data is managed, protected and controlled correctly.

2.3 Intellectual Asset Management

This section will describe the concept of intellectual asset management (IAM) which will be used as an important piece of theory in this research. In order to understand and identify how to protect, manage and create value from data as an intellectual asset, it will be important to understand and identify the core resources that derive and create value in a firm. The theory behind IAM is based on utilizing knowledge as the basis of the organization or a project, as well as the need to capture this knowledge in order to utilize it as various intellectual assets (Petrusson, 2016).

2.3.1 Intellectual Capital and Intellectual Property in relation to Intellectual Assets

The theory behind intellectual assets is linked to theory regarding management of intellectual capital and intellectual property (Sullivan, 1999). Intellectual capital is defined as intangible assets being the value creating drivers and the core resources in a firm. An intangible asset can be defined as knowledge, information or resources that are abstract and difficult to map (Petrusson, 2016). Sullivan (1999) describes intellectual capital as human capital and intellectual assets as the two main elements in a firm, see Figure 2.6. Furthermore, intellectual property is defined as a part of IAM and is utilized to legally protect intellectual assets. The intellectual property rights include legal protection such as patent right, trademark, trade secret, design right and copyright. However, in relation to this research study, not all intellectual property rights are feasible control mechanisms that can protect data. This will be further discussed in Chapter 2.4 Control of Data. Intellectual property can help create strategies for how identified assets can be controlled. However, by also identifying and evaluating the fundamental assets, IAM enables possibilities to understand intangible resources that are used to form a basis to drive value in a firm. This method is particularly beneficial to use in a firm where knowledge is a core resource which, in relation to this research, is very applicable since the research is focused on understanding how data can create value in DBCs and ICDTs.

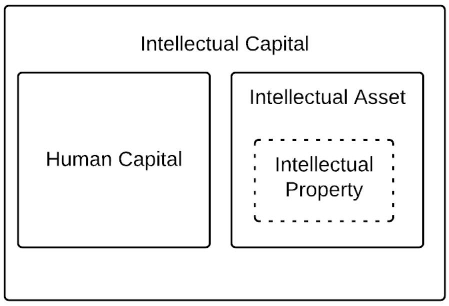


Figure 2.6: Illustration of Sullivan's (1999) description of intellectual capital

2.3.2 The Intellectual Asset Management Framework

Intellectual Asset Management is based on theory where knowledge based firms can be broken down into distinctive intangible or intellectual resources (Petrusson, 2016). By dividing and breaking down the resources, this provides a foundation to enable the firm to find strategies to capture, manage, protect and leverage the core resource base of the organization or project. This method provides means to capture and convert core resources into intellectual assets. Petrusson (2016) developed an IAM Framework that provides supporting methods and tools to evaluate an organizations ability to capture, create value, leverage, utilize knowledge and technology assets, through both internal and external processes. Originally, the model was developed for academic purposes and focused on supporting and creating utilization possibilities from research results. Although the framework can be applied in any industry type, it enables most value for a knowledge based firm where the core resources are intangible resources. In this research study, all firms that are analyzed in the comparative multiple case study are defined as knowledge based firms to a certain degree, where data can be found and utilized as an intellectual asset.

Petrusson's (2016) IAM framework consists of four main building blocks; *claim, position, decide* and *organize*, see Figure 2.7. All areas enable features and supporting tools that can be used throughout the intellectual asset management process. The claiming step, also known as the capturing step, represents the tools to identify, analyze and capture intellectual assets within a firm. The positioning step represents tools to position and analyze the external environment such as analyzing what competitors do in relation to the intellectual asset portfolio. The deciding process, also known as the leveraging step, includes tools that help support and decide various utilization capabilities and opportunities for the identified intellectual assets. The final step, organize, helps organize the firm in a way which enables management of the identified intellectual assets.

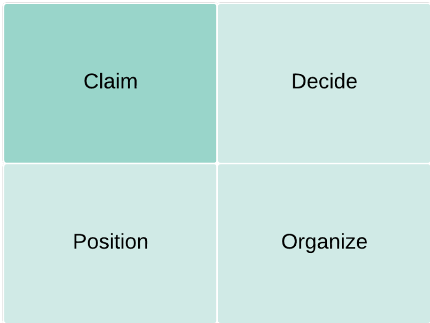


Figure 2.7: Illustration of the building blocks of the Intellectual Asset Management Framework by Petrusson (2015)

In this research study, the claim processes will be used to capture data resources and convert these into data assets through the material gathered in the comparative multiple case study. The study will also incorporate some aspects of the position and decide steps to enable evaluation of the data assets from a business value and utilization perspective. These two steps will, however, be conducted to a minimum extent due to the scope and delimitations of the research study. The decide and position perspectives will be conducted through an evaluation and analysis of the identified data assets to find the best opportunities of leverage, control and creation of business value from the identified data assets. To be able to utilize the IAM Framework theory in this research, the method and process of capture will be described more thoroughly.

2.3.3 Claiming Intellectual Assets

The claiming processes is explained as a method and tool to identity and capture knowledge assets, in order to understand the value of the asset or resource (Petrusson, 2016). The typical and fundamental knowledge assets for research projects can be divided into different categories, see Table 2.1.

Table 2.1: Examples of knowledge asset categorizations, according to Petrusson (2016)

Knowledge Assets	Description
Data	Unstructured information e.g. measurements, raw data, survey data.
Database	Structured and searchable data e.g. spreadsheets, electronic database, tables, matrices etc.
Observation	Conclusion and analysis from empirical data collection and data analysis e.g. correlations, market trends, optimizations, insights.
Theoretical Framework	Models that describe phenomena, relations and causes e.g. theories, frameworks, schemes
Technical Solution	Creative solution to a scientific problem e.g. methods, processes, systems, devices, inventions, technology.
Instruction	Practical guidelines that are provided for certain activities e.g. recipes, algorithms, manuals, scripts
Visualization	A static or dynamic illustration that goes beyond a simple drawing e.g. prototypes, diagrams, graphs, simulations
Software	A computational function or feature with an implemented and organized set of data that can preform certain tasks autonomously e.g. applications, systems, platforms, programs.
Narrative	A structured and coordinated case study with empirically shown account or narrative e.g. literature, interviews, original source studies
Creation	A creative and artistic work e.g. painting, music, drama, artwork.

The claiming process can be described as an internal analysis of the key resources in a company, where it is important to analyze and address relevant stakeholders involved in the company and identify the rightful owner of the intellectual property rights. In this stage, one should ask themselves *"which research results do we have that can be utilized?"*. Petrusson (2016) describes that there are five central steps that should be utilized in order to fulfill the claim process. These are the following: (1) identification of knowledge assets, (2) defining the technical solution, (3) claiming benefits, (4) claiming control position and (5) creation of an asset list, see Figure 2.8 for full illustration.

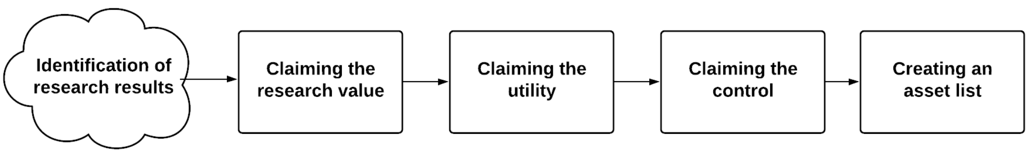


Figure 2.8: Illustration of the claiming process in the IAM Framework, according to Petrusson (2016)

The first process of identifying the knowledge assets, involves analysis and evaluation of what assets can create value and what potential this can create (Petrusson, 2016). In order to do this, it is recommended to use the categorization of knowledge assets stated in Table 2.1. After categorizing all identified knowledge assets, the next step is to define the technical solutions and describe what technical value the assets can contribute and result in. It is important that the description of the technical solution is clear and comprehensible, in order for the knowledge asset to managed and enable value creation for the firm. It is also beneficial to have a clear description of the asset for future collaborations, since this enables easier communication regarding the assets. The next step is to claim the benefits of the assets which includes describing the knowledge assets and research results from a utilization perspective. This step is done in order to understand what benefits and opportunities the asset can offer the firm. In addition, this step also includes analysis of which stakeholders are needed to utilize the asset. Stakeholders in this sense can be, for instance, customers or suppliers. The next step in the claiming process is to claim the control position. This includes analysis and evaluation of the control position of the claimed asset through different perspectives such as intellectual property rights, confidentiality, technical, relational and individual dependency. The final step is to create an asset list that summarizes and gathers all information from the steps in the process and enables the knowledge assets to be disclosed in an intellectual asset list format. The intellectual asset list is a useful tool to enable tagging of the different knowledge assets and linking them to IPR's, background knowledge and utilization aspects. The tagging can also include linking the assets in relation to certain people, stakeholders or organizations. The identified intellectual assets can thereafter also be tagged into different IPR's in order to ensure legal protection and property claims.

2.3.4 Claiming Data as an Intellectual Asset

As mentioned above, the IAM framework can be applied and adapted to different industry segments depending on the context and the value logic. Since the research topic will be focused on a commercialization setting, it will be beneficial to analyze the framework from this perspective. Hermansson (2020) presents how the tool can be applied and adapted into a commercial business setting. This adaptation includes four new main processes as follows: (1) resources, (2) positioning on the market, (3) leverage and (4) organizing the venture. The different steps are similar to the academic setting, but are applied and constructed to better form and capture assets and resources in a venture and/or commercialization setting. Through this kind of thinking, the framework can result in an overall set of technology assets related to data in commercial IT projects, see Figure 2.9. It is, however, important to note that this type of categorization is not exhaustive, but rather enables a reference of how the framework and categories can change in relation to different organizational contexts.

Data Technology Assets
Raw Data
Databases of Data
Algorithm Data Analysis
Software Invention
Technical Solution
Software System
Algorithm User Interface
Graphical User Interface

Figure 2.9: General data technology asset categorization, according to Hermansson (2020)

2.4 Control of Data

Intellectual assets are, due to their intangible form, inherently difficult to control. Nonetheless, as discussed above, it is necessary to claim control of the firm's value-driving assets in order to secure the competitive advantage that they provide. In order to do so, it is important to understand the construction of innovations. According to Petrusson (2004), an innovation consists of a number of different intellectual building blocks which interact with each other. The first step in evaluating the portfolio of building blocks for a certain innovation is to analyze the explicit property claims. Most innovations are based on various property claims which, as mentioned, can be viewed as structural building blocks that enable structural control, i.e. various control mechanisms. As seen in Figure 2.10, right based property claims, such as IPRs, often interact with other means of control to create a strong structural control position. Furthermore, there is another aspect which needs to be taken into account. With commodities rapidly changing from physical to virtual items, as discussed previously, essential assets become easier to distribute, manipulate and copy (Liang et al., 2018). Hence, data protection emerges as a critical way of securing data ownership and value of data. It is therefore even more important to examine all possible property claims which can enable control. The presented control mechanisms in Figure 2.10 will be further discussed and applied to data specifically below.

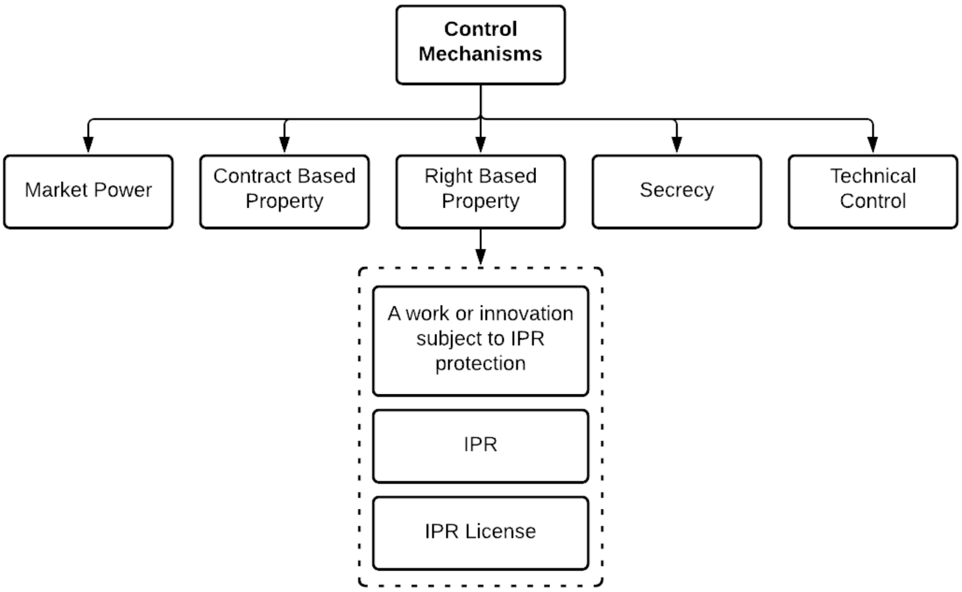


Figure 2.10: Adapted version of structural control mechanisms from Petrusson (2004)

2.4.1 Right Based Property

Petrusson and Heiden (2009) define *assets* as valuable objects and *property* as objects for commercial transactions. Thus, for an asset to be considered as property, it must be trusted as an object of commercial transaction. This is only possible if there is a system of well-established ownership rights supported by a judicial system which can validate the rights. In addition, these systems must be used and accepted by market actors as well as the general society. Commercial transactions can occur even in the absence of well-established ownership rights, however, they will lack the trust required by the financial markets to consider the property as capital. Similarly, De Soto (2003) states that if assets are not recognized as property by the political and financial institutions, they cannot effectively be used as capital. This shows the importance of right based property claims for capturing and utilizing assets such as innovations, but it also shows the dependency on well-defined and accepted societal systems. Nonetheless, a property right grants the holder exclusive authority to decide how the resource and its services should be used (Alchian, 2021). In addition, the holder is granted the right to delegate, rent or sell any portion of the property right. Thus, it is easy to recognize this as an important mechanism of control. However, the IPR in itself, the patent for instance, is only one part of right based property (Petrusson, 2004). This building block also includes the subject matter in itself, e.g. the innovation or accomplished work, and the IPR license, as illustrated in Figure 2.10. For the purpose of this paper, and due to its scope and delimitations, the selected right based property that will be investigated in relation to data are presented in Table 2.2 below.

Table 2.2: Overview of relevant IPRs, adapted from Petrusson (2004) and Petrusson (2015)

Subject Matter	IPR
Novel Technical Invention	Patent
Artistic, Literary or Scientific Work	Copyright
Database Content	Sui Generis Rights
Documented or Undocumented Information	Trade Secret Management

Patent Right

The requirements set out by the European Commission (European Patent Office, 2021a) for receiving patent protection are as follows:

- the invention must have *technical character*;
- the invention must be *novel*, i.e. must not have be known before the application has been filed;
- the invention must involve an *inventive step*, i.e. it must differ significantly from previously known inventions and must not be obvious to a person skilled in the relevant technical field;
- the invention must have *industrial applicability*, i.e. it must be possible to produce or utilize the invention in any kind of industry.

A patent gives the holder the right to prevent others from producing, using or selling the invention without the patent holder's permission (European Union, 2021c). However, patents also imply disclosure of the invention subject to the application. Thus, it is important to consider and weigh the benefits of protection against the risks of disclosure. Patent protection can also be granted for processes, even in cases where the sequences of steps in the process are performed by a computer using software (European Patent Office, 2021b). This is referred to as a computer-implemented invention and the patent may comprise claims to the following, as stated by the European Patent Office (2021b), *"...computers, computer networks or other programmable apparatus, whereby at least one feature is realized by means of a computer program."* However, although the use, storage or application of data might be eligible for patent protection, the underlying data is not (Determann, 2018). Therefore, in relation to protecting the rights to data explicitly, patent law is not considered an effective legal framework.

Copyright

Copyright law was originally developed as a means to prevent others from producing copies of and extracting financial value from a specific material artifact (Petrusson, 2004). As such, copyright can be viewed as a tool that enables transactions of and financial value extraction from artistic work. In order to be granted copyright, the artistic, literary or scientific work must meet the requirement of originality, meaning that the creation must have a certain degree of originality (PRV - Swedish Intellectual Property Office, 2021; European Union, 2021a). Copyright relates to data in two primary ways. Firstly, the source code of a computer program is protected by copyright in the form it was written, thus either preventing unauthorized copying of the code or ensuring availability and distribution of it. Secondly, the structure of a database is protected under the copyright legislation (European Union, 2021b). However, according to Ritter & Mayer (2018), copyright law is inadequate in its application to data because data is vastly different from artistic assets due to its industrial nature. Fundamentally, data is observations and recordings generated by a large network of signals (Mayer & Ritter, 2018). Thus, a complication of applying copyright, and other IPRs for that matter, to data is the question of authorship and inventorship, i.e. whether the creation is considered human or computer made. Furthermore, it can be questioned whether it is the structure of the database, i.e. the arrangement of data, which is the value-driving source for companies or whether it is the volume, characteristics or utilization of the data. Based on the answer, the relevance and importance of the copyright protection for databases can also be questioned.

Sui Generis Right

In the European Union, the sui generis protection grants the creator exclusive rights to the contents of a database (European Union, 2021b). This means that even if the *structure* of the database does not meet the originality criteria of copyright, it is still possible to protect the *content* with the sui generis right. In practice, this means that the creator can prevent others from extracting and/or reusing the entire or parts of the database's content. However, in order to obtain this right the maker must show proof of a substantial financial, material or human investment in obtaining, verifying or presenting the contents of the database.

Trade Secret Management

Trade secret management can be a powerful tool to use either on its own or in combination with other control mechanisms, but it requires investments and management. The requirements and the application to data will be further discussed under "Secrecy".

2.4.2 Market Power

Market power as a control position is fundamentally dependent on how well a firm is able to secure favorable market conditions which result of sustained competitive advantage. To evaluate an organization's market power the following question could be answered: *"to what extent is it possible for the firm to exert control over competitors or other actors in the value chain due to its strong market position?"*. Being a so-called *first mover*, i.e. being the first to establish a certain product or service in an industry, can result in strong market power in the form of *first-mover advantages*, as explained by (Schilling, 2012), see Figure 2.11 for illustration.

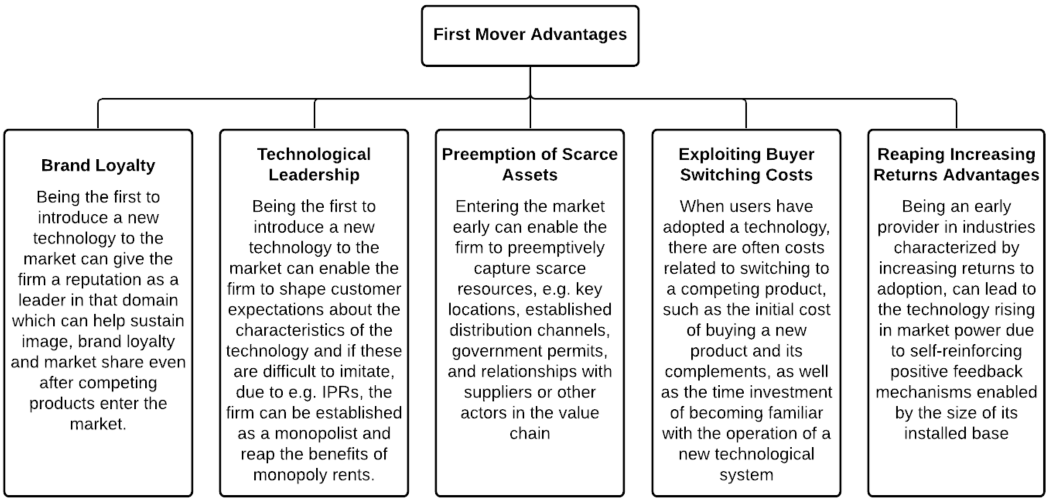


Figure 2.11: First mover advantages as described by Schilling (2012)

The advantages related to increasing returns are, as mentioned, dependent on the size of the installed base, i.e. the number of users (Schilling, 2012). When an industry is characterized by network externalities, the benefits and value of a product increases with the number of users. For example, a user may choose a computer platform based on the number of users rather than the technological benefits, because it increases the ease of exchanging files. For a data-driven business however, this also means that a large installed base gives the firm access to a large amount of data, which can in itself be a strong source of competitiveness as it increases the opportunities for product or service development. Another important aspect of network externalities is complementary goods, which is defined as products or services that enable or increase the value of another product (Schilling, 2012). Some companies produce both an initial product or service and complimentary goods, whilst some rely on other actors to produce goods which make their product functional or desirable. This is strongly related to market power since the larger the installed base is, the more likely it is that companies produce complementary goods, i.e. products or services compatible with that technology. Simultaneously, the availability of complementary goods influences the users' choice and thus, can increase the size of the installed base which creates a self-reinforcing cycle. Furthermore, brand equity can enable a firm to gain market control even if the firm is not an early entrant to the market. Sometimes, the firm that establishes the strongest brand and manages to build trust among the customers can out-compete the actors who were first to market (Schilling, 2012). Positive brand equity can result in numerous benefits for the firm such as the ability to charge higher prices than competitors. In addition, if a firm is perceived as reliable by the general public, chances are that users are more willing to let the firm handle their valuable or sensitive belongs, e.g. banks or firms collecting personal data

2.4.3 Contract Based Property

Contractual agreements present a means for establishing both relations and transactions of commodities between parties (Petrusson, 2004). To elaborate, contracts regulate both the firm's vertical and horizontal relationships, i.e. partnerships and employment. They can therefore be viewed as an agreed system of obligations and demands. A contract as a transaction is, in the legal tradition, often described in the form of offers and acceptance; meaning exchanged promises to sell and buy. However, with the increase of digitalization and an internet-based market, new forms of agreements have surfaced, e.g. content provider and license agreements (Petrusson, 2004). This development shows how contracts are not only a tool for transferring concepts related to the manufacturing of physical products, but also for the transfer of conceptualized digital and intellectual products as well as intellectual elements in all relations. This is especially relevant in relation to data, since the data that flows between entities are not only business essential, but also inevitable in an internet-based market.

As of today, the laws and regulations that have been developed specifically for data are primarily focused on the data privacy and protection aspects (Mayer & Ritter, 2018). However, there is still a lack of legal regulatory frameworks that offer guidance on how data transactions should be constructed. As a consequence, the industry is left to regulate data ownership and rights to utilization, mainly, based on contractual agreements. The exception is in cases where data is considered personal or sensitive, in which data privacy regulations, such as GDPR, apply and regulate the rights and obligations of the various parties (General Data Protection Regulation, 2016). Thus, if data transactions occur between entities, it is of the utmost importance that these are considered and regulated through contracts, e.g. terms and conditions. In the case of open source code however, there are multiple different licenses in use today, with terms that vary in strictness. The most strict open source licenses require all derivative work to be made freely available (Bekkers & Updegrove, 2012). This means that companies must carefully consider which license the open source code is using, before incorporating this in their products or services. Otherwise, they risk having to make their proprietary work publicly available and thus, losing important competitive advantages.

2.4.4 Secrecy

Valuable information on either technology or any other part of the business can be subject to trade secret protection if the following circumstances are true (European Union, 2021d):

- the information is *secret*, i.e. not known by either the public or experts in the field;

- the information is *valuable*, i.e. it is of commercial value to the business;
- *reasonable measures* have been taken to keep the information secret.

In practice, the holder of the trade secret is protected against dishonest behavior, e.g. unauthorized access and disclosure of information. This also applies in the instance of a non-disclosure agreement being breached by the other party. However, trade secrets are not a form of exclusive property right, which means that if others develop the same information through e.g. parallel innovation or reverse engineering, they can use it freely (European Commission, 2021; European Union, 2021d). In the past, theft of trade secrets often implied stealing or copying a physical object (European Commission, 2021). However, in the digital era, this type of unlawful access is increasingly made through breach of cyber security systems and accessing computer networks. This raises particular challenges in collaborative ecosystems. Furthermore, data is often made valuable through analysis of large data volumes, whilst an individual data point might be of little value (Zech, 2016). Simultaneously, companies that manage large volumes of data often collaborate with others or acquire the help of third-party companies in the data processing and analysis activities. Therefore, managing the ownership of data and maintaining processes to keep the data as trade secret can be challenging. This ultimately means that the third requirement for trade secret protection, proof of reasonable measures, can be difficult to meet. Nonetheless, trade secrets are an important control mechanism for data-driven companies and many organizations regard their big data as confidential information (Kaynak & Yin, 2015). However, according to Kaynak & Yin (2015), this can hinder the development of novel approaches by academic researchers. For this reason, there are benefits to an increased collaboration between enterprises and engineers/researchers, where data is made available for improvements of existing techniques and fostering of new ideas.

2.4.5 Technical Control

One way of enforcing technical control is by hindering others, e.g. competitors, from accessing strategically important corporate assets (Petrusson, 2004). Digital Rights Management (DRM) is an established systematic approach, intended to prevent unauthorized redistribution or usage of digital content; hence, protecting the copyright of the media (Liang et al., 2018). DRM solutions can include various tools and other technical protection measures but generally, five key components are required: Security, Access Control, Usage Control, License Management and Payment Management, see Figure 2.12. When it comes to unstructured data, such as raw data, technical protection is especially important since the data is easy to manipulate or damage which could have severe consequences (Liang et al., 2018). Thus, encryption is usually used for this type of data as it is the most secure method for protection. However, technical control also includes barriers intended to retain customers by making it technically difficult to use competing products or services, thus, creating a technical lock-in effect (Cusumano, 2010). Examples of this include geographical blocking of certain technical features, standards or compatibility between products and proprietary software development kits.

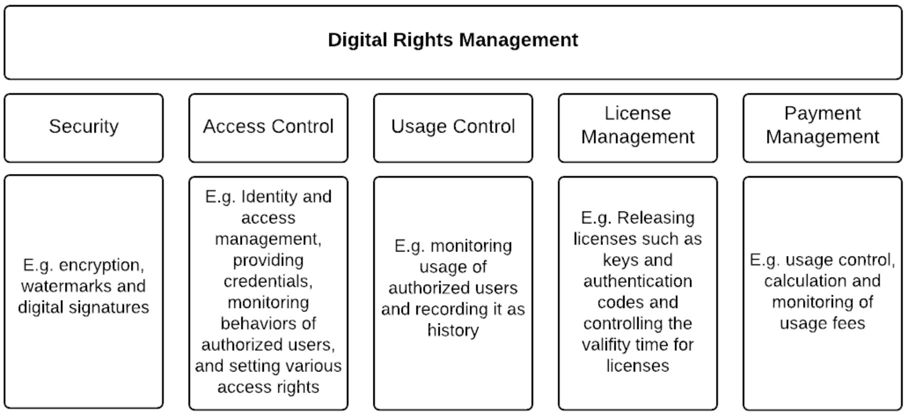


Figure 2.12: Digital Rights Management as described by Liang (2018)

3. Methodology

This chapter will describe the method utilized in the research study. The methodology includes a description and outlining of the main considerations in relation to the chosen research strategy and research design. The Chapter will also include a description of the research method and a discussion on the relevant quality aspects of the research.

3.1 Research Strategy

The research strategy sets the foundation of the nature of the research and its link to theory (Bryman & Bell, 2011). For this study, the research strategy was primarily grounded in the research purpose and the relevant theory, as presented in Chapter 2 Theoretical Framework.

3.1.1 Relationship between Research and Theory

The ultimate goal of this study has been to develop a framework that can be used by companies to better understand their data assets. Therefore, an inductive approach has been dominant in the research strategy, meaning that theory has been generated based on empirical investigations (Bryman & Bell, 2011). However, the complexity of the field has, in some cases, entailed difficulties in obtaining the necessary data to build the theory. For this reason, it was necessary to incorporate a more deductive approach as well, in relation to some aspects of the study. However, due to the unexplored nature of data, the existing theory would not have been sufficiently applicable, which indicates that a strictly deductive approach would have been too restraining for this study. As illustrated in Figure 3.1, it was therefore necessary to alternate between a deductive and inductive research strategy. Initially, existing theory was studied and tested during unstructured interviews, i.e. the "pre-interviewing" phase, in order to provide a foundation for the main interview phase of the study. This revised theory was thereafter tested during semi-structured interviews, i.e. the main interviewing phase, where the empirical observations from the interviews were used to develop new theory. Lastly, the new theory was pressure-tested during semi-structured interviews, called "pressure-test interviews", and was thereafter further revised in order to increase the research validity, reliability and applicability.

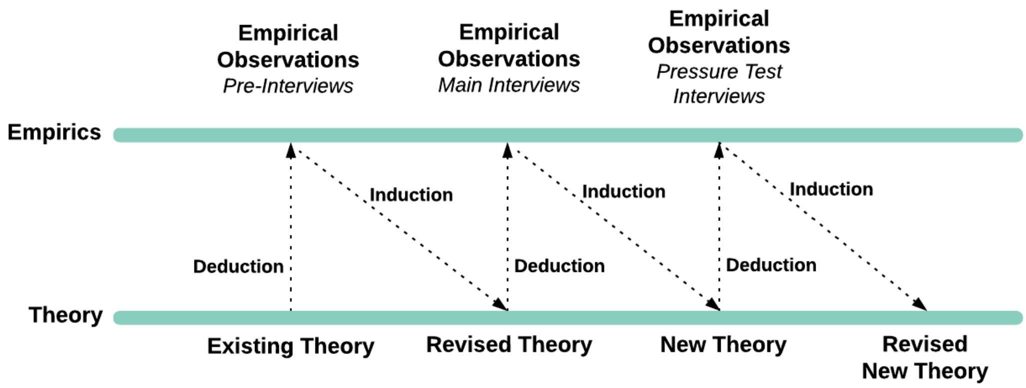


Figure 3.1: Illustration of the research strategy approach

3.1.2 Epistemological & Ontological Positioning

An ontological study regards assumptions based on the nature of reality, whereas an epistemological study regards assumptions about knowledge and what is acceptable, valid or legitimate knowledge (Saunders et al., 2019). In this research study, the key concepts are related to *data*, *sustainable competitive advantage*, *control mechanisms*, and *intellectual asset management*, which all can be described as phenomena whose existence is related to and developed by human creations. These concepts are therefore considered ontologically subjective. Thus, this study uses an ontological position in constructionism, meaning that these concepts are seen as social phenomena which are continually being produced by social actors through social interactions, and are therefore in a constant state of alteration (Bryman & Bell, 2011). In practice, this means that the studied subjects, e.g. data and intellectual asset management, are comprehended and accepted as socially constructed phenomena. Hence, the epistemological position of this study lies in interpretivism, meaning that methods related to natural sciences are not considered appropriate for the study of the social phenomena in this project (Bryman & Bell, 2011). Instead, strategies which respect the differences between people and objects of the natural sciences will be required, which in turn requires an understanding of the subjective meaning of social actions. This means that this study will not attempt to *explain* e.g. data, but rather *understand* its relation to and implication on humans.

3.1.3 Quantitative & Qualitative Research Considerations

Quantitative and qualitative research form two distinct clusters of research strategy, where the differences lie in the study's relationship between theory and research, as well as the epistemological and ontological position (Bryman & Bell, 2011). Based on these aspects of this study, that have been described above, qualitative research methods were deemed most appropriate to use.

3.2 Research Design

Bryman and Bell (2011) describes a research design as a framework and method to collect and analyze the data. The research design utilized for this project is a comparative multiple case study, where data was collected from four different cases, i.e Company A, B, C and D, during a limited period of time, and in relation to the following aspects; existing data resources within the company, data control mechanisms currently in use and value creation areas for existing data. The collected data was then compared in order to enable better understanding of the social phenomena studied, which is an implication of the logic of comparison (Bryman & Bell, 2011). Furthermore, when utilizing a qualitative research strategy, interviewing is a common method to use both in a case study design and a comparative design (Bryman & Bell, 2011). Thus, the data collection in this study was carried out through semi-structured and unstructured interviews, and the examined aspects where then compared between the four cases. However, in order to broaden the perspectives, increase understanding and enhance the quality of the research, the interviews were complemented with an analysis of organizational documents and academic literature. In addition, because the investigated companies were divided to represent two different types of companies, i.e. industrial companies undergoing digital transformation (ICDTs) and digital-born companies (DBCs), the applied research design comprises another comparative aspect as well. The data collected in connection to the different aspects was therefore clustered to enable examination and pattern detection between the two types of companies.

3.3 Research Method

The research method includes various strategies and techniques to collect data (Bryman & Bell, 2011). This section will outline the research method used in this study, including what data was deemed necessary for the research as well as the process for the research study.

3.3.1 Required Data for Research Study

The research study is focused on answering one main research question and four sub-questions. Research Question 1 focuses on investigating what data resources companies have that are important for value creation. In order to answer this question, it was assessed necessary to have a general understanding of what data is and what forms it can take. In addition, theory in relation to data and the RBV and KBV, as described in Chapter 2.2 Sustainable Competitive Advantage, was used to investigate and answer Research Question 1. The second aspect that was important to understand was how data resources are valuable for firms.

Research Question 2 focuses on investigating which control mechanisms are relevant to use in relation to data. The data required for this research question was a general understanding of which control mechanisms there are and which of these are applicable to the identified data assets. In addition, information about how data assets are controlled in the investigated companies was also required in order to answer the Research Question 2.

Research Question 3, regarding how data assets create value for businesses, was focused on identifying key utilization areas for how to create value from data. In order to answer this question, theory in relation to sustainable competitive advantage and data monetization was required. Furthermore, information about how the investigated companies work with creating value from their data was also deemed important. By linking information from theory with the empirical findings from the comparative multiple case study, it was possible to identify which data assets create most value, and this enables a sufficient answer to Research Question 3.

Research Question 4, regarding how value and control of data assets are utilized in DBCs and ICDTs, requires data collected from the interviews in relation to Research Question 1, 2 and 3. Therefore, no "new" data is needed specifically for this research question. This data is necessary in order to enable a comparative analysis between the two types of businesses.

3.3.2 Research Process

The research process includes an explanation and description of the major steps utilized in the research study, see Figure 3.2 for an illustration of the research process. The research process was divided into three main phases (1) defining and scoping of the project, (2) main data gathering and (3) results and conclusions.

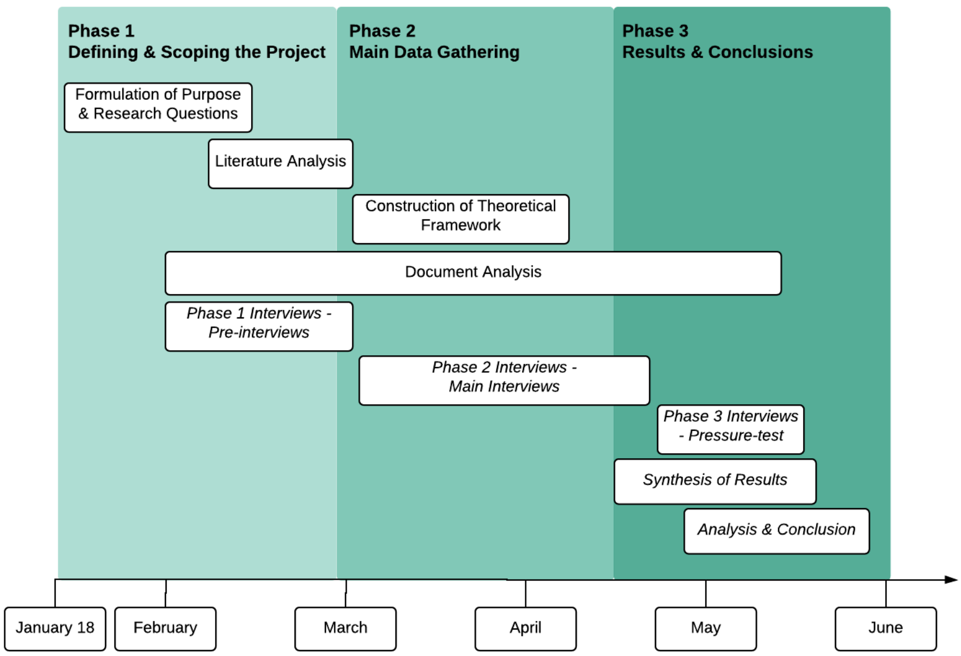


Figure 3.2: Overview of the research process

The first phase included the formulation of purpose and research questions, the literature analysis and the phase 1 pre-interviews. This phase began with the identification of a problem at the two collaborating companies. By reviewing the existing literature landscape and performing a search of prior research, it was validated that the identified problem had not yet been addressed or investigated. From this investigation, the research purpose and research questions were formulated based on the interest of the two collaborating companies, the interest of the researchers and the existing literature landscape.

After determining an initial definition and scope of the project, the literature analysis was conducted. The purpose of the literature analysis was to identify which data resources or assets that exist in firms, and how these are categorized according to the literature, see Appendix A. The key finding in the literature analysis was that there were five different perspectives to categorize data - Data Source, Data Activities, Data Structure, Data Utilization and Data Security. These perspectives were then structured in the form of a data value chain, that represented how data flows in a company. Furthermore, the categories were also used to address relevant utilized data processes within a firm. By utilizing a deductive approach, these findings were used as a basis in the main data gathering phase through the semi-structured interviews. The literature analysis concluded in a data value chain which set the foundation for the phase two interviews, where the researchers asked questions in relation to the data flow thinking, see Figure 3.3 for illustration of the data flow. In parallel, while conducting the literature analysis, the phase 1 interviews were conducted. These interviews were held as pre-interviews with various people in the two mainly investigated companies, where the purpose was to gather information about the companies for the research study as well as finding appropriate interviewees for the main interviews, the phase 2-interviews.

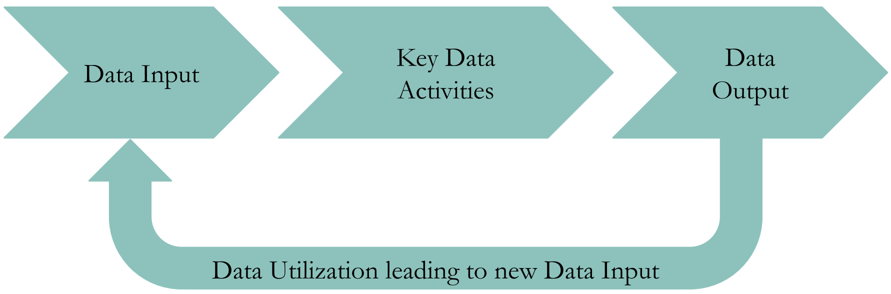


Figure 3.3: Illustration of the data flow used in the company interviews

The second phase constituted of a main data gathering process where the majority of the data collection was conducted. The phase included the construction of the theoretical framework as well as the phase 2 interviews. The theoretical framework was constructed by analyzing relevant theories, models and key concepts and determining which were necessary to use in order to answer the set research questions. Phase 2 was highly iterative in order to adapt and revise to the research purpose and research questions in accordance to the proceedings and findings in the main data gathering. By utilizing the theory found from the literature analysis in the main interviews, a deductive approach was used in order to test the findings from the theory. The interviews utilized the data flow thinking as illustrated in Figure 3.3, where the researchers asked questions in relation to each specific part of the data flow. The purpose was to understand the transformation of data by using the data flow in a company, and to understand how the different companies worked with managing, controlling and creating value from data from each specific part of the data value chain. The phase 2 interviews constituted of the main data gathering interviews which included 22 main interviews, conducted at the four investigated companies. In parallel, a thorough document analysis was performed as part of the comparative multiple case study. This included analysis and assessment of various internal company documents for the purpose of gathering more information and enabling an answer to the research questions. From the phase 2 interviews and the document analysis, a first draft of data assets were identified and captured in the different parts of the data value chain by utilizing Petrusson's IAM Framework logic (Petrusson, 2016). These findings set the foundation for the Data Asset Framework.

The third and final phase constituted of analyzing and synthesizing the results found in the second phase. The data asset findings were tested in pressure-test interviews with six key people, four from companies A and D, and two from companies B and C. The pressure-test interviews with Company B and C were conducted in conjunction with the main interviews, since these were held at the end of phase 2. The four people from Company A and D were identified during the pre-interviews and main interviews, and were assessed to have the knowledge necessary to provide valuable feedback in relation to the identified data assets. Before the pressure-test interviews, the researchers identified and clarified each data asset category. After the pressure-test interviews, the data assets were revised according to the feedback given and according to their applicability at each company. An example of a change was that data derived visualizations was removed as an asset, as it was assessed to be part of the data derived insights category. See Appendix B for a full illustration of how the data assets were adapted, revised and changed. By utilizing the data assets as a basis, findings from both theory and interviews were used to assess how the data assets can create value, as well as how they can be controlled. These findings were then compared in relation to ICDTs and DBCs to enable the comparative analysis. This phase concluded in the Data Asset Framework as well as the Data Pyramid, which are two major findings in relation to this research study. The Data Asset Framework is a framework that is supposed to clarify how data assets can be used in a firm and what is important to take into consideration. The Data Pyramid explains and highlights the findings in relation to how value and control is integrated in relation to data, and how these parameters are important to consider when utilizing data. These findings enabled the formulation of a final conclusion and an answer the corresponding research questions.

3.3.3 Data Collection

By utilizing a comparative multiple case study research design, this research study used three main methods for the data collection. These were the following; (1) literature analysis, (2) interviews and (3) document analysis. This section will in short describe the different methods and relevant considerations in regards to the different methods.

Literature Analysis

The literature analysis was conducted in order to understand and identify which data resources that exist in firms, and also to understand how data is categorized according to the literature. The findings from the literature analysis was used as a basis for the main interviews, to be able to identify and capture the various data assets within firms. Relevant publications were read and assessed by using online literature databases such as Google Scholar, Research Gate and the Chalmers University of Technology database. Due to the time constraints of the research study, a limited collection of 18 articles were gathered and analyzed. Based on parameters such as publication year, citations and references, the identified literature was assessed as highly relevant to use in the research. No blogs or other forums were used as a base for the literature analysis in order to ensure high quality results.

Interviews

Interviews were one of the main data gathering methods for the research study. There are three main interview techniques that can be used to gather qualitative information; structured, semi-structured and unstructured interviews, (Bryman & Bell, 2011). The key difference between the techniques is if the questions are predetermined with either exact wording, i.e. if the interviewer can modify the script or not. The structured interview uses a set script that should not be modified. A semi-structured interview uses a predetermined set of questions, but also enables the interviewee to go off script and make modifications as the interview proceeds. Unstructured interviews do not use a predetermined set of questions, where the interviewer uses a list of topics or issues as a basis for the interview.

The interviews were divided into three main phases, (1) pre-interviews, (2) main interviews and (3) pressure-test interviews. To ensure reliable and valid qualitative data gathering in the research study, pre-interviews were held at each company. When conducting interviews with a qualitative approach, it is important to understand what the interviewee means by asking the right questions (Bryman & Bell, 2011). The pre-interviews were conducted as unstructured interviews and included a short description of the project and a discussion about the interviewee's knowledge about data. The researchers thereafter assessed the interviewee's abilities to provide valuable insights in relation to the research questions.

The main interviews were utilized as the main data gathering for the research study. For these interviews a semi-structured approach was chosen, in order to allow for modification of the script and follow-up questions. The main interviews were conducted at all four companies, where the purpose was to identify how data as an intellectual asset is managed, used and controlled at each company. An interview guide was developed to ensure that the predetermined script was followed but also to ensure that the questions were asked in an appropriate way, see Appendix D. The interview guide was designed to retrieve answers in relation the specified research questions by both asking broad questions related to each research questions, and by incorporating an illustration of the data flow in a company. This interview structure enabled the interviewee to better understand the purpose of the interview, as well as enabled the researchers to adapt the questions according to the interview candidate. The different candidates were asked more specific questions in accordance to their role, for example the Product Managers, Innovation Leaders and Engineers were mainly asked questions related to the technology and data specific questions, whereas a Legal Counsel or Information Security Officer was asked questions in regards to legal compliance.

The pressure-test interviews were conducted as semi-structured interviews as well, and were held in the beginning of phase 3. The pressure-test interviews were held at all four companies, in order to test and validate the results gained from the research, and to be able to revise according to feedback.

There are many various ways to conduct interviews. However, face-to-face interviews is one of the most prominent methods to use that is proven to generate large amounts of data (Bryman & Bell, 2011). In this research study, face-to-face interviews were attempted to the largest extent possible. Thus, nearly all interviews were executed through various video conference programs. The interview candidates were recruited across the various levels of the companies based on their availability, willingness to participate and ability to provide relevant insights. The researchers' goal was to gain perspectives from many various aspects across the companies. The sample of interviewee candidates was comprised of:

- Senior/Director Level Engineers and Product Managers with extensive experience of working with data as an asset in products and projects in various ways, and with a high level of oversight as to what impact data can have on their teams,

- Senior/Director Data Scientists and Data Managers with extensive experience of working with data science and data management related to products and projects in correspondence with the various parts of the organizational teams.

- Product/Innovation Managers or Engineers or Data Scientists, with experience in working with data but not necessarily as a team leader, with a more detailed perspective of the impact of working with data as an asset

- Directors of IP, Product/Innovation Managers with IP experience, Legal Counsels, Information Security Officers, Product Managers with legal and IP knowledge and experience in advising on how to work with data in an organization.

This interview sample was assessed necessary in order to gain the perspectives needed to address and answer the research questions. In total, 44 interviews were held, where 16 were pre-interviews, 22 were main interviews and 6 were pressure-test interviews, see Appendix C for the complete list.

Document Analysis

Document analysis was used as a method for data collection in the research study, where the researchers analyzed internal company documents. Collecting data through documentation can be assessed as a systematic method to review and evaluate documents to deepen the knowledge that is utilized in a qualitative research study (Bowen, 2009). The documents that were analyzed in this research consisted of internal company material such as PowerPoint presentations, intranet web-pages and organizational reports. These documentations provided additional data for the research and were utilized to increase the understanding of the investigated companies.

3.4 Research Quality

When conducting business research, it is important to address the quality of the research in relation to reliability and validity. This section includes an assessment of the four main quality criteria in relation to a qualitative research study, as proposed by Bryman & Bell (2011).

3.4.1 Credibility

Credibility refers to the ability to address if the research results are trustworthy or seen as consistent with reality (Bryman & Bell, 2011). In order to maintain and establish credible research results, the research needs to be conducted with good practice and be validated in accordance to social reality. In this research study, credibility was achieved by using a comparative multiple case study research design and triangulation. According to Bryman & Bell (2011), triangulation is when several sources and methods are used to collect data on the same topic or research field in order to assure valid research findings. In this research study, the comparative multiple case design enabled collection of data both from literature, interviews and organizational documents. The research design also included an investigation of four companies, which also enabled the quality of the research to become more credible. For the interviews, it was assessed important to interview many employees within the organization, both with different roles and hierarchical positions, in order to capture a broad perspective of the research findings. In addition, the interviews were conducted in different phases, i.e. pre-interviews, main and pressure-test interviews, which also enabled higher verification and validity assurance of the research findings. An important consideration to have in mind in relation to the credibility of the research findings, is that the interviewees at the companies have represented a relatively small sample in relation to the size of entire company. This is especially true for the additional two companies that have not been as thoroughly investigated. Therefore, some of the empirical findings from the interviews might not be representative of the company as a whole but do, however, represent smaller company segments.

3.4.2 Transferability

Transferability refers to the ability to transfer the research findings and make them applicable to other contexts (Bryman & Bell, 2011). In this research study, the research results have been developed with the purpose of making them as applicable as possible for all the investigated companies. By providing descriptions and illustrations of the given results, the transferability is assessed as high, since this enables companies to read and thereafter implement the research results. However, it is difficult to assess if the transferability is high or low, since the resulting frameworks have not been implemented at the companies and this field is yet to be further investigated throughout the companies. However, in relation to the transferability quality critera, the research study is assessed as reliable and valid.

3.4.3 Dependability

Dependability refers to the ability to repeat the created results and findings (Bryman & Bell, 2011). In this research study, the research findings are based on both theory and an extensive amount of interviews. Therefore, they are estimated to be difficult to replicate to the same extent. However, utilization of the described research process, could enable abilities to arrive at similar results.

3.4.4 Confirmability

Confirmability refers to assurance of complete objectivity and assurance that the researchers have acted in good faith (Bryman & Bell, 2011). Prevention and management of biases has been one of the main methods used to ensure strong objectivity in this research study. In the interviews, biases were minimized by interviewing people in many various roles and hierarchical positions. This was assessed to minimize biases as it enabled an overview of the current company landscape. Biases were also reduced by investigating and interviewing four companies, which enables a broader perspective of the two types of businesses (DBCs and ICDTs). The research might have been even better if more companies were investigated, as two companies within each industry might not enable complete objectivity. However, as time constraint was a large limitation, this was assessed enough to minimize biases towards the companies and ensure reliable and valid research results.

4. Empirical Results

This chapter presents the results collected from the data gathering and is structured into four main sections. The first section includes the results and key findings from the investigation of the literature analysis. The second section includes the results and key findings from the investigation of data resources from the main interviews. The third section includes the results and key findings in relation to the investigation of control mechanisms used for data assets, from both the theory and main interviews. The fourth and final section includes the results from the investigation of how data assets can create value from both the theory and main interviews.

4.1 Investigation of Data Categorizations

This section presents the results from the conducted literature analysis. The purpose of the literature analysis was to examine which data resources and/or data assets that exist within companies and how they, theoretically, should be categorized in order to provide business value. The key finding from this analysis was that there are numerous ways to categorize data depending on which perspective the authors take, and what the purpose of the categorization is. Five overarching perspectives of data categorizations have been identified from the analyzed literature, see Table 4.1, and have been defined as follows:

- *Data Source* refers to data being categorized based on the origin of the data, i.e. how or from where data is generated, for instance an action or an actor;
- *Data Structure* refers to data being categorized based on the technical structure of data, e.g. degree of processing or filtering;
- *Data Activities* refers to data being categorized based on what is being done with the data, i.e. various activities which are performed in relation to data;
- *Data Utilization* refers to data being categorized based on how data is used, what value it creates or what purpose it serves;
- *Data Security* refers to data being categorized based on a security and/or sensitivity perspective.

Table 4.1 presents the analyzed literature and the various perspective that each article uses. As shown, some articles and books were found to include several perspectives. For instance, Rizk et al. (2018), Gimpel et al. (2018) and Möller et al. (2020) include three different perspectives of data categorization. Gimpel et al. (2018) and Möller et al. (2020) both cover the same three perspectives; Data Source, Data Structure and Data Utilization, which are the three most commonly used perspectives overall in this literature sample. A more detailed description of the investigated literature and the data categories included can be found in Appendix A. The following sections will present the findings in relation to each data categorization.

Table 4.1: Representation of the data categorization perspectives included in the analyzed literature

	Data Source	Data Activities	Data Structure	Data Utilization	Data Security
Xie et al. (2016)	x				
Hartmann et al. (2016)	x		x		
Erevelles et al. (2015)			x		
Hannila et al. (2019)	x		x		
Rizk et al. (2018)	x	x		x	
Hunke et al. (2019)	x			x	
Gimpel et al. (2017)	x		x	x	
Gregg (2006)					x
Allen & Cervo (2015)	x		x		
Zhang et al. (2017)			x		
Möller et al. (2020)	x		x	x	
Püschel et al. (2016)	x			x	
Möller et al. (2019)	x			x	
Siddiqa et al. (2016)		x			x
Bock & Wiener (2017)	x			x	
Chen et al. (2012)	x		x		
Weibel et al. (1998)	x				
Cavanillas et al. (2015)		x	x	x	
Total No.	***13***	***3***	***9***	***8***	***2***

Data Source

The Data Source perspective was the most common to use and includes categories such as transactional data and interaction data which refers to data created through these actions (Xie et al., 2016; Hannila et al., 2019; Gimpel et al., 2018). Rizk et al. (2018) presents data categories based on acquisition mechanisms such as tracking/sensors, crowd-sourcing and secondary devices which highlights where the data is generated. Similarly Hannila et al. (2019) and Hunke et al. (2019) define data which originates from customers, non-customers/suppliers and products/objects. This highlights an important finding from the literature analysis, which showed that it is common to categorize data based on the actors involved. In these cases, customer-related data is often mentioned (Hunke et al., 2019; Hannila et al., 2019; Möller et al., 2019, 2020; Bock & Wiener, 2017). Hartmann et al. (2016) was identified as a key piece of literature in relation to the Data Source perspective, since it summarizes where companies gather data from an actor-based point of view. This includes whether the data is generated internally, i.e. within the firm, or externally, i.e. outside the firm, and the nature of the data transaction. Internal data includes *self-generated data* and *existing data* (Hartmann et al., 2016). Self-generated data is generated for a specific purpose through tracking, e.g. sensors, or crowd-sourcing, i.e. data created by a large set of contributors, and *existing data* includes all data which can be drawn from IT systems within the company but that is currently unused, e.g. ERP data. External data includes (1) *acquired data*, e.g. purchased from data providers, (2) *provided data*, which can come from customers or business partners and is generally not available, and (3) *freely available data*, i.e. publicly available at no direct cost. These categories both highlight the different actors which are involved in the data generation or transaction, as well as the nature of the relationship and transaction. Furthermore, two different ways of categorizing metadata emerged from the literature analysis. Allen & Cervo (2015) divides metadata into business, technical and process or operational. This perspective provides a broad understanding of all information that exists within a firm. Weibel et al. (1998) on the other hand, divides metadata into the categories of content, intellectual property and instantiation. This categorization refers more to the creation and ownership of data, making it more control-based, and has nowadays been formally recognized as part of the ISO 15836 standard. Thus, Weibel et al. (1998) has been identified as another piece of key literature.

The Data Source perspective is an important categorization, as it highlights and helps the firm understand where the data comes from, i.e. what essential data generators should be protected and controlled. In order to secure the access and control of data, the firm must first understand where and how it is generated. In this way, significant competitive advantages can be identified and secured. Additionally, the Data Source perspective provides insights about which actors that are involved in the data generation, which is another important aspect in relation to control and ownership of data which can be essential to define in contractual negotiations. Figure 4.1 is an adapted version of the categories in the Data Source perspective, primarily as presented by Hartmann et al. (2016). In the adapted version, data generated from products and services has been categorized as Primary Data Input, and provided, acquired and freely available data has been referred to as Secondary Data Input.

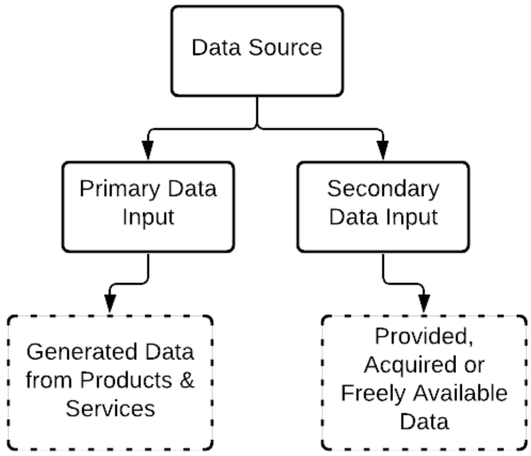


Figure 4.1: Adapted version of a categorizatioon of data by Hartmann et al. (2016)

Data Activities

The Data Activities perspective was only included in three of the examined literature pieces. Siddiqa et al. (2016) mentions *storage*, *pre-processing* and *processing* as main categories of key data activities that are performed in order to transform data and make it usable. Similarly, Rizk et al. (2018) presents the category of *information processing*, which contains activities such as cleaning, aggregation and synthesizing, as well as *advanced analytics* containing data mining, machine learning and visual analytics. In addition, Cavanillas et al. (2015) presents annotation, information extraction, and machine learning algorithms as key activities needed in the big data value chain. Data activities are significant since they, through processing and analytics, enable value creation from data (Rizk et al., 2018). Different types of activities have different implications and thus create value in various ways. This is an important perspective in relation to data categorization since it enhances the understanding of how data is transformed and what needs to be done to add and extract value from it. This perspective is therefore closely related to the Data Structure categorization.

Data Structure

The Data Structure perspective is the second most used perspective among the analyzed literature. Several articles mention structured, semi-structured and unstructured data as different categories in this perspective (Erevelles et al., 2015; Hannila et al., 2019; Allen & Cervo, 2015; Zhang et al., 2017; Gimpel et al., 2018; Cavanillas et al., 2015). The definitions of these categories are fairly harmonized in the analyzed literature and relate to the accessibility of the data and how it is stored. Structured data typically has well-defined relationships and is stored as Tables in a database (Hannila et al., 2019; Gimpel et al., 2018). Semi-structured data on the other hand, lacks the strict data model structure, and unstructured data often comprise full-text documents or multimedia content that lack the traditional row-column storage structure. These categories have been identified as key aspects in relation to the understanding of how data is transformed within the company, i.e. how it changes structure based on the activities that are performed. This is illustrated in Figure 4.2 below, where the *Company Data Flow* represents data coming in to the company and undergoing transformations thorough various processing activities.

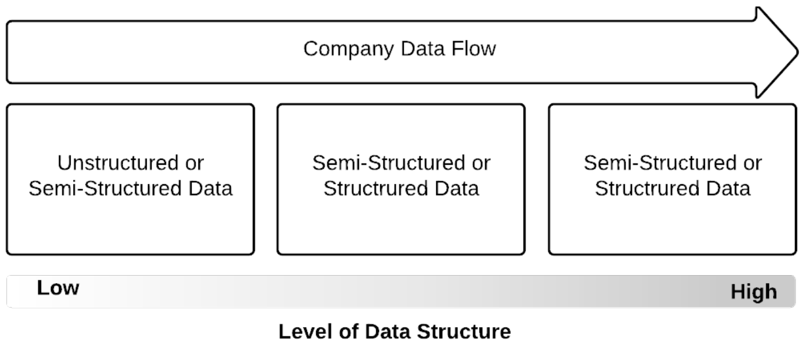


Figure 4.2: Adapted version of data categorization according to the analyzed literature in the Data Structure perspective

Data Utilization

Rizk et al. (2018) categorizes data based on the derived insights, i.e. information and knowledge, are utilized within the provided service. These utilization categories include visualizations of data for the purpose of facilitating decision making, as well as autonomous decision-making which allows for continuous development of the systems. Rizk et al. (2018) also highlights that insights can be operationalized to enable a new service feature or recommendations, most commonly used to guide the users-service interface in interactive or user-centered environments. Möller et al. (2020) & Hunke et al. (2019) present the following categories of data analytics: descriptive, diagnostic, predictive and prescriptive. These refer to data being utilized to explain what happened, why it happened, what will happen and what should be done. Thus, there is a a distinct time aspect related to the utilization of data. Gimpel et al. (2018) presents this as categories of time horizons, which include historic, current and predictive. Furthermore, Hunke et al. (2019) mentions the integration of data utilization into the existing portfolio of services and products as its own data category. This includes a stand-alone solution, meaning that the new service is offered separately and does not require an existing service/product, as opposed to wrapped around product and wrapped around service which refers to the data utilization being used to enrich already existing products/services. The Data Utilization category strongly relates to the various value creation opportunities that data can enable, and does thus not include specific data resources but rather ways in which data in general can be utilized. However, for the purpose of this study, Baecker et al. (2020), as described in section 2.1.3 Data Monetization, was chosen as a key piece of literature to represent the Data Utilization perspective, as it provides a holistic overview of data value creation and the related economic benefits possible for companies.

Data Security

Data Security is the last, but equally important, perspective in relation to data categorization. Only two pieces of the analyzed literature mention the Data Security perspective. However, this does not speak for the importance of this perspective. Siddiqa et al. (2016) presents privacy, integrity, confidentiality and availability as different components needed to assure big data security management. Gregg (2006), on the other hand, offers a categorization of data based on the level of availability that the data should have depending on its sensitivity. The levels are the following:

- Public - Anyone can obtain the data;
- Internal - The data is not available outside the company;
- Limited Distribution - The data is only available to a certain number of individuals;
- Personal - Data about an employee's individual status.

The importance of this type of categorization has increased with the implementation of GDPR, as it strongly relates to the management of personal data. The fact that the scope of this paper has been limited to not include a thorough investigation of personal data, and the implications that the relevant regulations entail, has likely limited the amount of literature found which includes the Data Security perspective. Nonetheless, this type of categorization plays a vital role in ensuring suitable data management processes that both protects the data and ensures compliance with data regulations.

Key Findings in relation to Data Categorization

As illustrated in Figure 4.3, the identified data category perspectives complement each other and can be presented as different steps in a data value chain. Based on this thinking, data can be said to originate from a certain source and in a certain structure, where-after key data activities are performed which transform the data into a new structure and in turn enables data utilization in various ways. Throughout the process, data security is an important aspect in relation to control and protection of data assets. This logic presents the foundation of the comparative multiple case study, including the main interviews.

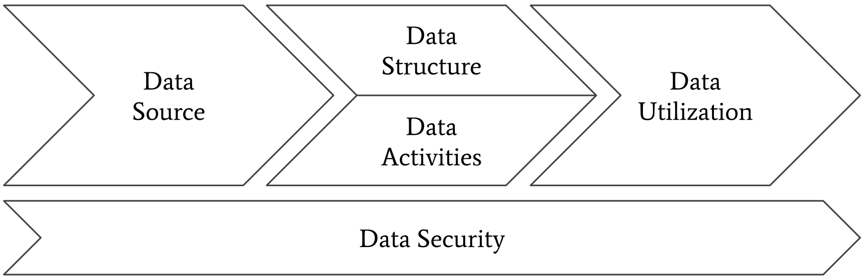


Figure 4.3: The Data Value Chain

4.2 Investigation of Data Resources

This section will present the key findings from the main interviews in relation to research question one, regarding what data resources are important for companies' value creation.

The main interviews resulted in various exemplified data resources from each company representative. Table 4.2 presents the most prominent data resources identified in the companies through the main semi structured interviews. There is no correlation between the order of how the resources are presented in the Table. Some resources have been grouped into one category and some are standalone resources. An example of a grouped category is *Technical Data*, which in the interviews had many various wordings such as vibration data, telemetric data, sensor data or technical performance data. A full description of all identified data resources can be found in Table 4.3.

Table 4.2: Data Resources within the investigated companies

Identified Data Resources	ICDT		DBC	
	Company A	Company B	Company C	Company D
User Data	x	x	x	x
Sensor Data	x	x		
Technical Data	x	x		
User Behavior Data	x	x	x	x
Raw Data	x	x	x	x
Statistical Data	x	x	x	x
User Feedback Data	x	x	x	x
ID Data	x	x		x
Personal Data	x	x		x
Event Data	x	x	x	x
Location Data	x	x	x	x
AI Generated Data			x	x
Third Party Data	x	x	x	x
Business Sensitive Data	x	x	x	x
Temperature Data	x			
Outdoor Climate Data	x	x		
Metadata			x	x
Refined Datasets	x	x	x	x
Client Performance Data			x	x
Power Consumption Data	x	x		

Table 4.3: Definitions of Data Resources

Identified Data Resources	Definitions
User Data	Data related to the user of the product or service such as ordering, invoice, user preferences, interactions, etc.
Sensor Data	Data generated from control systems such as motion detectors, power consumption detectors, temperature sensors, etc.
Technical Data	Data related to the hardware or software technical aspects of a product or service, such as telemetric data, sensor data, etc.
User Behavior Data	Data related to the user's behavior that is generated through analysis of user data, either using AI or the human intellect, with the purpose to increase understanding of how the user interacts with the product or service.
Raw Data	Data collected from all types of sources that has not yet been processed in any way and is therefore unstructured and unprocessed.
Statistical Data	Data generated from the utilization of the product/service or acquired statistics from third-party actors that can be used for both diagnostics, calculations and or predict new market trends and business opportunities.
User Feedback Data	Data generated from users' utilization of the product/service in the form of feedback which can be used to improve the products/services or find new business opportunties.
ID Data	Data that has been converted from personal data into an ID number in order to encrypt the personal data and enable usage in diagnostics/calculations.
Personal Data	Data which can be used to identify a specific individual and is therefore considered personal and highly sensitive.
Event Data	Data about all events happening to the product or service, collected through either sensors or user interactions.
Location Data	Data related to the geographical location of a data point or a user.
AI Generated Data	Data generated through AI algorithms, such as ML recommendation systems that enable e.g. new data input.
Third-Party Data	Data that is gathered through external sources, such as collaborators or partners, used to integrate and improve features in the existing product or service offerings.
Business Sensitive Data	Data that can be classified as sensitive data to the company and is therefore important to protect, such as price settings, predicted market trends, etc.
Temperature Data	Data gathered from external parties or internal temperature sensors about the temperature of a product when utilizing the product or service.
Outdoor Climate Data	Data gathered from external parties regarding the outdoor climate such as air pressure, air quality, etc.
Metadata	Data about data; raw or processed information about the data itself which facilitates structuring of data.
Refined Datasets	Data collected and gathered from various sources that has been processed and structured as a dataset in order to be used for diagnostic, analytical or calculation purposes.
Client Performance Data	Data related to the technical performance of a client.
Power Consumption Data	Data related to the technical performance of the product or service that can be used to enable diagnostics/calculations.

Table 4.2 illustrates that there are many various terms and wordings used for data resources in companies, and that all investigated companies collect a lot of data in one way or another. There are many similarities found when comparing how the resources are used. Examples of these are User Data, Third Party Data, Statistical Data and Location Data. User Data was a common data resource that was exemplified in the interviews. This resource includes all types of data that can be related to the user, which is important to gather in order for a business to create value from the products and services. Third Party Data was also a common resource, where many companies had various partnerships or collaborations through which data was gathered or shared. Furthermore, Location Data was used differently in the investigated companies, both from a utilization and origin perspective. Examples of where Location Data was retrieved from was through GPS signals, global access points or internet protocol addresses.

A major difference in relation to data resources was how much of the data actually was used to create value within the company. Some of the data resources are more commonly collected in the investigated companies, such as User Data, User Feedback Data and User Behavior Data. However, Company C and D gather larger volumes of these types of data and find more value creating opportunities for it. Both AI Generated Data and User Behavior Data is collected and used to a large extent at Company C and D, which are very far ahead in utilizing these types of data resources. AI Generated Data has not yet been implemented to a large extent in Company A and B, which is the reason for them not being marked in the Table. However, Company A and B are starting to become more knowledgeable in this area and will likely, in the near future, use this as a resource as well.

Another interesting difference that was found, was how the data resources could be directly linked to the product offering of the company. Since Company A and B, which are ICDTs, have hardware products, hardware related data resources were found within the companies. An example of this was Sensor Data and Power Consumption Data. By utilizing hardware products with sensors, ICDTs have been able to collect large amounts of data for a long period of time. However a large obstacle has been that many parts of the companies have been unaware of what they can do with the data, which has led them to not utilize the data nor gather all the data possible. This, however, indicates that there are different data vocabularies within the companies which relate to the product offering of the firm. Furthermore, despite being a digital-born company, Company D recently launched a hardware product. One of the stated aims of this launch by Company D was to gather related data resources which had, so far, not been easily accessible to the company. An interesting finding was that, despite being digital-born and relatively data savvy, Company D had not developed a sophisticated vocabulary about hardware-related data. This further points to the fact that the product offering of the company affects the data vocabulary.

All the companies have discussed Personal or Business Sensitive Data, which indicates that there is a need to implement and maintain internal security policies regarding how to manage these resources. Personal Data was heavily emphasized as important in the interviews, due to its high relevance for value creation but also due to the difficulty to use because of strict personal data regulations, such as GDPR. Personal Data was assessed as especially important in the companies working close to the users. Company C has not been marked in the Table, since this company has chosen to not utilize personal data in their current business model. Business Sensitive Data was also discussed, however not to the same degree. Nonetheless, it emerged as an important aspect to consider when working with data. Many interviewees mentioned that when working with data, it is important to regulate what data is used, who uses it and how it is used. For instance, business Sensitive Data does not necessarily contain personal data, but could be sensitive for the business if it it ended up in the wrong hands. This implies that it is important for all companies to assess data resources according to Personal Data and Business Sensitive Data.

Another interesting finding was that data resources had many different technical structures and that the companies lacked a common data vocabulary. This finding indicates that there are many different perceptions in how to define a data resource, since the wording can be interpreted in various ways. The lack of a common vocabulary was identified in the interviews, but can also be seen in Table 4.2, where the various data resources are both broad and narrow. As mentioned, this was also identified in the main interviews where the answers to the questions was dependent on the interviewee's role and position in the company. It was clear that various roles highlighted the value of data resources in different ways. When speaking to an engineer, the data resources discussed were for example Raw Data, AI Modeling Data, Technical Performance Data and Mathematical or Analytical Data. However, when asking the same questions to a product or program manager, with a higher level of oversight, the resources discusses were e.g. User Data, Business Critical Data, Financial Data, Sales and Market Data. In the discussion with legal representatives, data resource perspectives such as e.g. Personal Data and Sensitive Data were most frequently discussed. A common statement by interviewees that were engineers was that a database is a way to store data and is not necessarily a resource nor an asset. However, when speaking to a legal representative, a database has a function of being legally protected by the database protection, sui generis. This finding indicates that there is a clear need to implement a more common vocabulary, to minimize confusion and to create better communication regarding data within a firm.

4.3 Investigation of Control Mechanisms

This section will present the findings from the main interviews related to research question two regarding which control mechanisms can be relevant to use in relation to data assets and why. The identified control mechanisms used in practice within the investigated companies are presented in Table 4.4. For a full description of each identified control mechanism, see Table 4.5.

Table 4.4: Identified Data Control Mechanisms

Identified Data Control Mechanisms	**ICDT**		**DBC**	
	Company A	**Company B**	**Company C**	**Company D**
Key Encryption & Pseudonymization of Data	x	x		x
Access & Utilization Tracking Tools				x
Access Management Systems	x	x	x	x
Data Lineage Tools				x
Data Quality Control Systems	x	x	x	x
Geographical Lock-In System				x
Secure Network and Servers	x	x	x	x
Data Separation Processes				x
Terms & Conditions Agreements	x	x	x	x
Company Guidelines Regarding Data Management	x	x	x	x
Contractual & Licensing Agreements	x	x	x	x
Trade Secret Management	x	x		x
Brand Equity	x	x		

Table 4.5: Definitions of Data Control Mechanisms

Identified Data Control Mechanisms	Definitions
Key Encryption & Pseudonymization of Data	Various methods to encrypt and pseudonmize personal or business sensitive data.
Access & Utilization Tracking Tools	Processes and systems used to track who has accessed or utilized data.
Access Management Systems	Systems which enables management of access to data, such as giving or taking away access.
Data Lineage Tools	Tools used to track where data comes from, how it has changed and who is responsible for managing it within the company.
Data Quality Control Systems	Systems, processes or tools used to control the quality of data, either within the organization or that is provided to/shared with other actors.
Geographical Lock-In System	Systems used to enable access to certain type of data based on the user's geographical location.
Secure Network and Servers	Ensuring that the networks and servers that are used by the organization are secure, for the purpose of preventing unauthorized access and reducing cyber threats.
Data Separation Processes	Methods used to separate data, can both used to remove and separate personal data or business sensitive data from other data to ensure compliance with personal data or internal business policies.
Terms & Conditions Agreements	Contractual agreement with the primary purpose of clarifying the ownership and utilization rights to data between the user and provider of products/services.
Company Guidelines Regarding Data Management	Internal policies and guidelines which state how data should be managed by the employees of the organization.
Contractual & Licensing Agreements	Contractual agreements with the purpose of determining the ownership and utilization rights to data between the company and various actors, such as collaborators and partners.
Trade Secret Management	Internal measures used to control and prevent undesirable sharing of data assets.
Brand Equity	The effects that the brand of a company has on other actors, such as users or partners, that affects the company's access to and control of data.

Based on the findings from the interviews, it is clear that Company D, which is a DBC, has come much further in the implementation and development of data control mechanisms compared to the other companies. Another interesting finding is that Company A and B, which are ICDTs, use the exact same control mechanisms. This indicates that ICDT work quite similarly and have started implementing control of data to a large extent.

During the interviews, Key Encryption & Pseudonymization of Data was mentioned as an important control mechanism in relation to raw data and was strongly related to the management of personal data, meaning that it is used in order to ensure compliance and facilitate the utilization of personal data. Company C, which is a DBC, does not collect or process personal data which is the reason for it being the only company that does not use this type of control mechanism. Access Management Systems, Data Quality Systems and Secure Networks & Servers were also control mechanisms used by all four investigated companies, primarily for the purpose of preventing unauthorized access to data and minimize the risk of cyberattacks. In addition, Company D works a lot with various control mechanisms intended to track changes of the data, what the data is being used for, who is using it and who is in charge of it, see Table 4.4. This is done in order to increase the transparency around and control of data within the organization.

In regards to data ownership, the different companies work in various ways. All companies emphasized that this is an important issue to consider and had various methods to ensure ownership of or utilization rights to data. All four investigated companies utilize Terms & Conditions Agreements, Contractual & Licensing Agreements and Company Guidelines Regarding Data Management to control their data. These control mechanisms were stated as especially relevant in relation to regulation of data ownership during transactions. Terms & Conditions Agreements primarily relates to ensuring consent from the users when collecting and utilizing personal data. However, when it comes to regulating ownership for non-personal data, other contractual agreements are usually needed. This is represented by Contractual & Licensing Agreements in Table 4.4, which is a control mechanism used by all investigated companies in relation transactions of data during collaborations and partnerships. Lastly, Company Guidelines Regarding Data Management is also used by all companies which is not surprising as this sets the foundation for how data should be managed within the organization. Trade Secret Management is another control mechanism used by three of the investigated companies, mainly for the purpose of ensuring that business critical information does not end up in the wrong hands. Similarly to others, this control mechanism is implemented to various degrees within the companies.

Another interesting finding from the interviews relate to Company A and B having brand attributes that strongly correlate to words such as safety and security which, according to the interviewees, incites trust. The brand equity of these companies therefore affect the willingness of, for example, users to give away data, as they perceive it to be in safe hands. Similarly, collaborators, suppliers and partners have more trust in the data quality because of the companies' brand equity. Thus, Brand Equity was recognized as an important control mechanism.

Lastly, none of the four investigated companies mentioned Right Based Property as a control mechanism in relation to data. Although IPRs have been mentioned as important aspects, their direct application and relevance to data has shown to be ambiguous. The reasons for and implications of this finding are further discussed in Chapter 5.2 Control of Data as an Intellectual Asset.

4.4 Investigation of Value Creation

This section presents the findings related to research question three regarding how data assets can create value for companies. The section will be presented based on theory and the findings from the main interviews on how the companies work with creating value from their data.

Table 4.6 presents the key utilization areas identified in relation to how the companies utilize and create value from their data. These results were captured by utilizing theory regarding data monetization models as provided by Baecker et. al (2020), and described in section 2.1.3 Data Monetization. This was accomplished by synthesizing and converting the interviewees' responses in relation to the twelve presented utilization areas.

Table 4.6: Identified value creation from the investigated companies

Identified Utilization Areas	ICDT		DBC	
	Company A	Company B	Company C	Company D
Asset Sale			x	x
Sales of Data Insights			x	x
Business Process Improvement	x	x	x	x
Data Enrichment			x	x
Product or Service Innovation	x	x	x	x
Product or Service Optimization			x	x
Contextualization		x	x	x
Individualization			x	x
Strengthen and Build Customer Relationships	x	x	x	x
Data privacy and Control Guarantee				x
Strategically Opening Data			x	x
Data Bartering				x

When contextualizing the results presented in Table 4.6, a main finding was the difference in utilization and value creation of data between the DBCs and ICDTs. Company C and D, the DBCs, have come much further in their process of utilizing and creating value of their collected data in comparison to Company A and B, the ICDTs. However, Company C and D implement the different utilization areas to various extent; some are implemented to a large extent while some to a much lesser extent. Many of the utilization areas where Company C and D have been marked in the Table are areas where company A and B are on the verge of creating capabilities to leverage data. An example of this is Individualization, Contextualization, Data Enrichment and Asset Sales. The main reason for company A and B not being marked in the Table is that these were not identified during the interviews. However, capabilities have been found which can lead to the utilization areas being implemented in a near future.

An example of a highly implemented utilization area is Contextualization and Individualization, where Company C and D create business opportunities of their main product and service offering. The companies' core business revolves around an individual that uses their product or service. The companies then capture the individual's data and find ways to leverage it through value creation such as personalized experiences or recommendations based on preferences. Sales of Data Insights was another utilization area that was used to various extents. Company C and D use this more as part of their business model, where they sell or distribute insights to other parts of their value chain, e.g. suppliers. However, the companies distribute or sell the data by using licensing models rather than one-time costs. Company A and B have not yet developed the capability to distribute or sell data insights to the extent that Company C and D have. This is a valuable utilization area to implement, as it can help the organization work more united and utilize the data in the company better. A utilization area that is not utilized to a large extent is Data Bartering, which occurs when companies exchange data in return for other valuable assets. Company D is the only company that utilizes this to a certain extent, where suppliers are allowed access to aggregated and visualized user data about their specific product offering. Through this, the suppliers can gain insights about the popularity of their products which can help them make decisions to correlate to the user needs. This makes Company D's value offering more valuable which is a benefit for the company. In addition, Company D can gain leverage over their suppliers in other negotiations. Conclusively, Data Bartering is an important utilization area, which both the DBCs and ICDTs could benefit from implementing more.

The main similarities between the investigated companies are that the most used methods to create value from data is through Business Process Improvements, Innovating New Products and Services or Building and Strengthening Customer Relationships. These methods are utilized to a large extent in Company C and D as they incorporate R&D data into their decision making processes and utilization activities. Company A and B do, however, also utilize these areas, but not to the same extent. An example of how Company A and B are utilizing these areas is how they have started to collect data from their products and services in order to predict when they will need maintenance. This will be of very high value when implemented in a larger scale. All of the investigated companies utilize the opportunity to create and strengthen their relationships with customers in various ways, which was addressed as an important utilization area that creates value for companies. This was implemented in various ways. For instance, the companies gather data about customer needs and behaviors in order to find new ways to leverage data.

An interesting finding in relation to the capability to utilize data and to Innovate New Products or Services, was the difference in sharing data throughout the companies. The DBCs invite the risk of having a lower control of their data in return for the increasing opportunities to innovate that arise from data sharing within the company. In comparison, the ICDTs choose to not take the risk of lower control, but might instead hinder innovation within the company. This finding indicates that there is a large difference between DBCs and ICDTs in the reasoning of data sharing. Another interesting finding in relation to this was the difference in what type of data was used to innovate. The ICDTs used more internally generated test data from the products and services, whilst the DBCs to a higher degree rely on data generated from the user utilizing the product.

5. Analysis

This chapter aims to discuss and analyze the presented results from Chapter 4 Empirical Results in relation to the theoretical framework. The analysis is divided into seven main sections. The first five sections aim to connect and analyze the empirical findings in relation to the research questions. The sixth section presents the Data Asset Framework. The seventh section aims to answer the four sub research questions.

5.1 Converting Data Resources to Data Assets

Capturing and collecting data resources is nowadays an essential task for companies to perform in order to be part of the digital transformation in society. However, to be able to create value from data, it is not enough to only collect a large amount of data resources. It is also important to convert the resources to assets in order for them to create value for a business. The main difference between a resource and an asset is that the resource is the antecedent form of the asset, and assets are defined as something of value and is owned and controlled by someone. This means that when there is a strategy and purpose for the data resource regarding control and value creation, it can be considered an asset. Through this logic, an asset can in turn be leveraged for monetary or economic value, as it is seen as something of value.

Data resources have an important role within companies' value creation processes in many ways as it sets the foundation for the value creating opportunities. For an organization to create value from data, the company needs to, first of all, have the internal capabilities to collect data. Secondly, the company needs to understand how to manage data as well as understand how to communicate about data, which was confirmed in the interviews as being a large obstacle. Thirdly, it is important to understand how to utilize and leverage the data in order to create economic benefits and abilities to monetize upon it. These fundamental reasons on why data resources are valuable for companies set the foundation for the importance of converting data resources into data assets.

5.1.1 The Concept of Data Resources in Companies

The company investigation demonstrated that all companies collect various data resources and that they all collect large amounts of data. The investigation also showed that there was a large difference in how DBCs and ICDTs communicate about data, for instance, there were many various perspectives to take into consideration. In order to convert data resources into data assets, it was assessed important that the assets need to be understood by all the various perspectives identified in a firm.

From an IP perspective, which aligns business, technical and legal aspects, it is important to identify what is valuable and capture this into an asset in order for it to create value. An interesting finding from the comparative multiple case study was that it does not really matter how much data is collected or generated in a company if there is not a corresponding strategy to create value from the data. The various amount of data resources were found to not necessarily correlate with the value driving factors of data resources. It was assessed to be more important to find and implement strategies to utilize and create value from the data, and focus on transforming resources to assets that enable value creation. By utilizing the IAM Framework logic, seven main data assets were identified at the investigated companies. These were found by analyzing the empirical results from the main interviews and using theory regarding how to capture intellectual assets in firms. A comprehensive analysis of the findings of the captured data assets will be presented in this section.

5.1.2 The Main Data Assets

Figure 5.1 presents the seven data asset categories, including sub-categories, that were identified and captured in the investigated companies. These seven main assets were captured by clustering together the data resources found in the interviews and the key aspects in literature, where-after common denominators were found within the categories. When presenting the main categories in the pressure-test interviews, a finding was that the Raw Data and Metadata categories were very broad and were difficult to understand. Therefore, both of these categories were divided into sub-categories to make the framework more useful and easy to understand. The remaining categories were deemed more valuable without set sub-categories in order to avoid limiting the data assets. The presented main categories and sub-categories have been defined by the researchers, from findings in the literature, the comparative multiple case study and the interviews. All recommendations are based on the researchers' findings in relation to the research study. The following text will explain the data assets and provide generic examples of what the different categories represent.

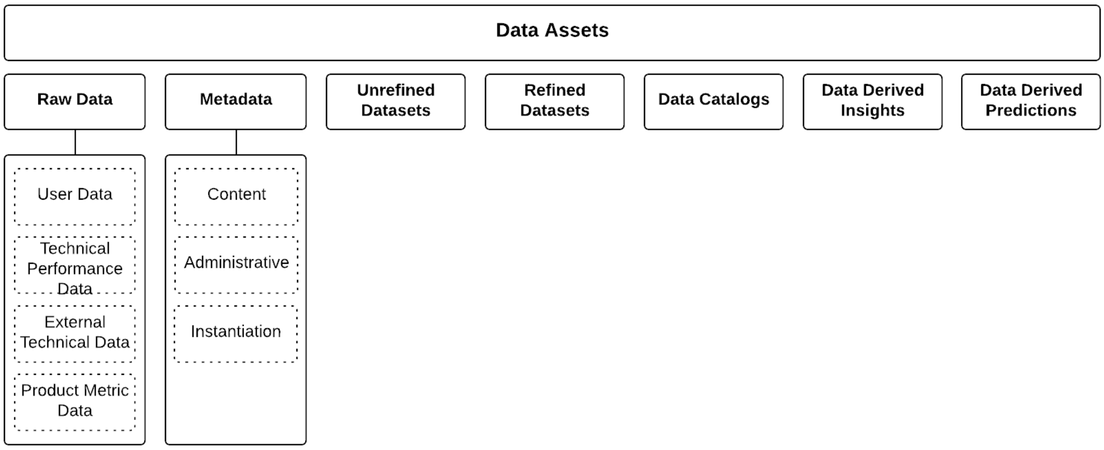


Figure 5.1: The Data Asset Categorization

Raw Data

This category is defined as *"unstructured and unprocessed data originating from either primary or secondary data sources"*. Raw Data refers to all types of data that is either generated, acquired, provided or freely available that can be used by the company. The category includes large volumes of data with many varieties of data types and data that is retrieved with a high velocity to the company. These characteristics can be directly linked to the 3 V's that describe big data, where the two remaining V's, Value and Veracity, can also be linked to the essence of the raw data category. To make the user of the framework understand what types of data that can be assessed as raw data, the raw data category includes a set of sub-categories. The four generic sub-categories are User Data, Technical Performance Data, External Technical Data and Product Metric Data, see Table 5.2 for a full explanation. It is, however, important to note that there might be more sub-categories in relation to raw data. For the purpose of this research study, the presented sub-categories in the Table were the identified sub-categories that were most applicable for value creation at ICDTs and DBCs.

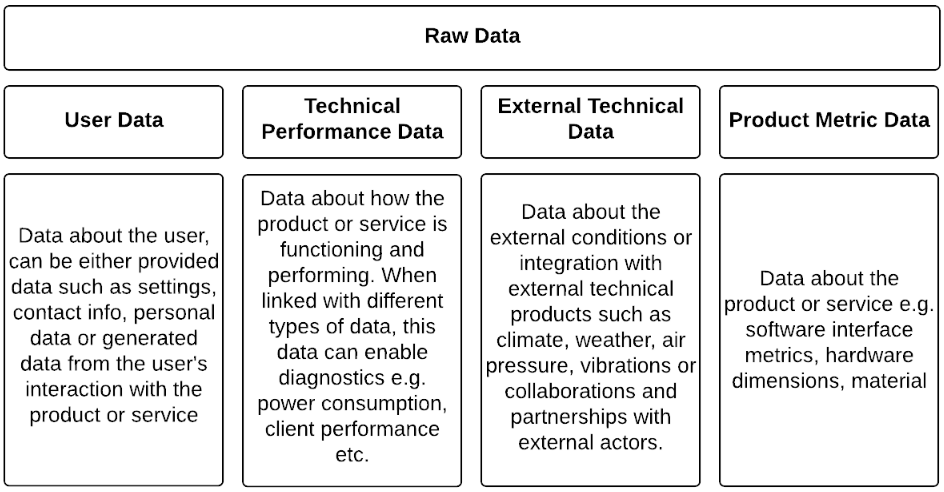


Figure 5.2: Definitions of Raw Data sub-categories

Raw data is assessed as an important data asset as it captures all data that flows into a company from either products and services, partnerships, collaborations and more. It is one of the first types of data that flows into a company and is important to collect and store, in order for the company to become more data-driven. From an intellectual asset perspective, this is an important asset to capture as it is critical to ensure where the data comes from, what quality it has, as well as what it can be used for to create value.

In the pressure-tests and interviews, an interesting finding was the discussions about personal data and where personal data would be included in the framework. The researchers concluded that personal data is a type of user data that is collected as raw data. Personal data is one of the most important resources to address in terms of control and value, since it requires processing in order to be leveraged in a way that ensures compliance with the laws and regulations. One could assume that personal data would be included in more data assets, however the researchers believe that it is most important to highlight the personal data when it is raw and unencrypted, since it is difficult to leverage it in this form without first processing it. After encryption and pseudonymization however, the personal data can be leveraged in a completely different way and the opportunities for value creation increase tremendously.

Metadata

This category is defined as *"unstructured or semi-structured data and or processed data, originating from either raw data, primary or secondary data sources"*. Metadata includes three generic sub-categories; content, administrative and instantiation, see Figure 5.3 for the full explanation. The term *creator* is used in both the content and administrative categories, and in both cases it refers to the creator of the content, but for different purposes. For the content category, this specifically refers to what the creator has developed and who is responsible for it, whilst in the administrative category the purpose is to determine the rights and ownership of the created content. This means that the creator and the right-holder are not always the same person which is an important aspect to distinguish. Similarly as described in the raw data category, there might be more generic sub-categories that describe metadata. However, this research study, the proposed sub-categories were developed for the purpose of understanding the metadata category and how metadata is important for value creation for DBCs and ICDTs.

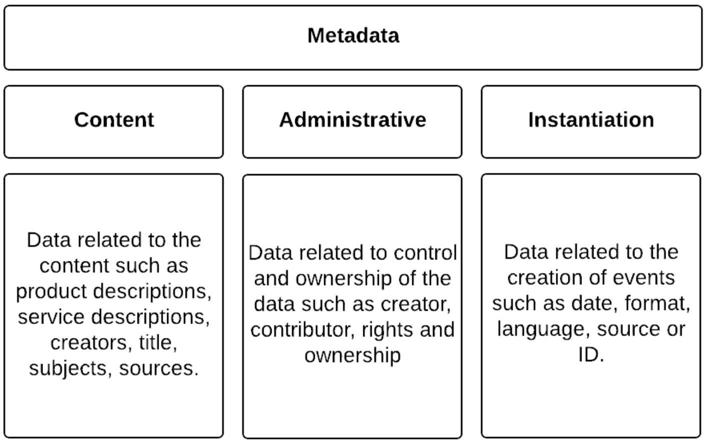


Figure 5.3: Definitions of Metadata sub-categories

Metadata is defined as a significant asset in a company since it defines the coherency and structure of data within the firm. It is a data type that flows into a company, most often collected from either the companies' own products or services, or externally acquired from other companies. Metadata has shown to be a critical asset for data-driven companies and companies that are striving to become more digitalized. From an internal perspective, metadata plays a vital role since well-structured metadata enables higher coherency and usability of all other the existing data within the company. Thus, from an intellectual asset perspective, metadata is an important asset to capture since it enables value creating opportunities within the firm.

Unrefined Datasets

This category is defined as *"semi-structured data originating from raw data or metadata, where the data is either separated or merged into a dataset, which gives some structure to the data. The data is however not yet ready for diagnostic and calculation purposes"*. The unrefined datasets were identified as the first step in the data flow where activities in relation to data creates value for a firm. The unrefined datasets captures the transformation of making raw data and metadata more valuable. These activities can for instance be pre-processing activities such as encryption of data, filtering of data, annotation and cleaning of data, to make it more valuable and useful for diagnostic and calculation purposes. The datasets do not have the highest quality, might not be available for all employees and might not be fully compliant with the internal data policies. For these reasons, unrefined datasets are seldom useful for diagnostic or calculation purposes. However, the unrefined datasets are defined as an important asset since they can be described as the first process to create value from data. In this process, value is created to the degree that the datasets could be sold in their given structure for direct economic benefits.

Refined Datasets

This category is defined as *"structured or processed data originating from unrefined datasets, which has undergone a transformation that makes it ready to use for diagnostic and calculation purposes"*. The refined datasets capture the transformation of when the unrefined datasets are turned into refined datasets, which can be used to generate data derived insights and predictions. The difference between unrefined and refined datasets is that the data is more structured and incorporates more correlations, which make datasets ready for analytical purposes. These datasets have a large value to the company since they hold a high quality standard, are available to anyone within the company and are fully compliant with internal data policies. A recommendation is to appoint a team that manages the refined datasets in order to maintain the quality and policy compliance. These datasets are deemed important assets for a company, as they enable a firm to leverage and create value from the collected data in the firm. This is primarily done through analysis and statistical observations. These activities in turn can lead to the firm creating new business opportunities and methods to leverage data.

Data Catalogs

This category is defined as *"structured data or refined datasets stored in a searchable database, making it possible to search and find available data within the organization"*. An example of this is a type of digital dictionary that comprises searchable data and thus allows employees to find the datasets needed. Data Catalogs is an asset that captures all the data types in a firm and converts these into available and accessible data throughout the firm. This asset creates more transparency within the firm and enables the firm to ensure high quality data. If any employee can search and find data, anyone can note if there is something wrong with the data which in turn can lead to a high quality data catalog. A recommendation would be to have a team responsible for maintaining the data catalog, in order to ensure high quality, and to make sure that not everyone has access to make changes since that could be problematic. Data catalogs was one of the categories that was not as implemented at all of the investigated companies, but was still addressed as a valuable asset in a firm that can be highly recommended to implement. From an intellectual asset perspective, data catalogs are important to capture as it enables a company to work more united and increases transferability and availability of the existing data.

Data Derived Insights

This category is defined as *"a new type of correlation that has been drawn from data related to the cause and effect of e.g. historical actions, events or decisions, commonly presented in a graphical form such as a report, presentation or visualization, and originating from refined datasets and data calculations or diagnostics."* The data derived insights capture the transformation of creating value from the refined datasets by converting these into insights that can be used as a basis for decision making in a firm. There are many examples of this, however, if set into an R&D data context, the most prominent would be insights regarding the performance of a product or service, a user's interaction with a device or the utilization of a product or service. The insights are either generated by a human or an AI through a computer program, where various refined datasets are analyzed and transformed into insights. These types of insights can enable the firm to identify and create new value creating opportunities. From an intellectual asset perspective, data derived insights enable large value as they allow a firm to find value creating opportunities such as new business opportunities or improvements of the existing products and services.

Data Derived Predictions

This category is defined as *"futuristic forecasting of e.g. events, trends or decisions based on historical events, decisions or correlations and originating from refined datasets and calculations or analytical based on the datasets."* Data derived predictions are very similar to data derived insights, where the main difference is that this category is a futuristic prediction in comparison to insights, which are focused on historical or present events. The asset captures the same type of transformation from the refined datasets, but converts these into predictions that can be used as a basis for decision making in a firm. Examples of data derived predictions could be predictions about the maintenance of products, e.g. "this product has been used for an X amount of times and is predicted to be in need of reparation in X days". Another prominent example is predictive recommendations to consume new products or services, based on a user's purchasing behaviors, to accommodate their future needs. Similarly to data derived insights, data derived predictions enable large value as it allows a firm to find value creating opportunities such as predicting new business opportunities, improvements of the existing products and services or new product markets to enter.

5.2 Control of Data as an Intellectual Asset

Control of data can essentially be used for the following four purposes:

- To regulate ownership and rights to utilize data;
- To ensure compliance with relevant data regulations;
- To protect data assets from unauthorized access;
- To ensure the continuous access to data.

This indicates that control of data can have many meanings and implications. All of the above mentioned purposes emerged during the main interviews, and what became obvious was the need to combine several different control mechanisms in order to achieve all four purposes. Table 5.1 presents the identified control mechanisms in relation to the theoretical control mechanisms described in Chapter 2.4, Control of Data. As illustrated by the red marking in Table 5.1, four additional control mechanisms related to Market Power were identified, which were not mentioned during the interviews. This section will provide a more detailed analysis of the control mechanisms related to theory and to the identified data assets.

Table 5.1: Control Mechanisms in DBCs and ICDTs

Identified Data Control Mechanisms	ICDT		DBC		Type of Control
	Company A	Company B	Company C	Company D	
Key Encryption & Pseudonymization of Data	x	x		x	Technical Control
Access & Utilization Tracking Tools				x	Technical Control
Access Management Systems	x	x	x	x	Technical Control
Data Lineage Tools				x	Technical Control
Data Quality Control Systems	x	x	x	x	Technical Control
Geographical Lock-In System				x	Technical Control
Secure Network and Servers	x	x	x	x	Technical Control
Data Separation Processes				x	Technical Control
Terms & Conditions Agreements	x	x	x	x	Contract Based Property
Company Guidelines Regarding Data Management	x	x	x	x	Contract Based Property
Contractual & Licensing Agreements	x	x	x	x	Contract Based Property
Trade Secret Management	x	x		x	Secrecy
Brand Equity	x	x			Market Power
Installed Base	x	x	x	x	Market Power
Network Externalities			x	x	Market Power
Technological Leadership			x	x	Market Power
Switching Costs	x	x	x	x	Market Power

5.2.1 Control Mechanisms in Relation to Theory

This section will investigate and analyze the correlations between the control mechanisms found in the comparative multiple case study and the control mechanisms described in the Theoretical Framework, Chapter 2.4 Control of Data.

Technical Control

As illustrated in Table 5.1, it is clear that Technical Control is the most widely used control mechanism in relation to data. Access Management Systems refers to control of who has access to the data and, thus, prevention of unauthorized access. This correlates to one of the key DRM components, *Access Control*, as discussed in Chapter 2.4 Control of Data, and is a mechanism used by all investigated companies. This finding suggests that it is an important and fundamental type of control mechanism in relation to data. Similarly, Data Quality Control Systems, which relates to processes and tools used to ensure the quality of the data, and Secure Network & Servers, primarily used to prevent unauthorized access and cyber threats, are also used by all four companies and can therefore as well be viewed as fundamental control mechanisms. Furthermore, Key Encryption & Pseudonymization of Data correlates to the key DRM component *Security* and is especially important in relation to raw data. In the investigation, this control mechanism was strongly related to the management of personal data, meaning that it is used in order to ensure compliance and facilitate the utilization of personal data. Company C, which is a DBC, does not collect nor process personal data which is why it is the only company that does not use this type of control mechanism. In addition, Data Separation Processes is only used by Company D, primarily in order to facilitate utilization of personal data within the organization. This indicates that Company D has come further in the development of processes for management of personal data. Company D also stands out in their utilization of various tracking tools, which are used to track who accesses or uses data, what data is utilized for and who is responsible for managing certain data within the company. In relation to the key DRM components, these mechanisms relate to both *Access Control* and *Usage Control*, which implies that Company D has implemented technical control mechanisms with the purpose of increasing the knowledge of what data exists within the company and how it is being used.

Contract Based Property

All four investigated companies utilize three different contract based control mechanisms, namely Terms & Conditions Agreements, Contractual & Licensing Agreements and Company Guidelines Regarding Data Management, as illustrated in Table 4.4. As mentioned in Chapter 2.4 Control of Data, there are more established regulations regarding ownership regarding personal data. Stating the purpose and reason for gathering personal data is one of the many requirements set out by GDPR (General Data Protection Regulation, 2016). Therefore, it is easy to understand why Company A, B and D have implemented Terms & Conditions Agreements in relation to data gathering from their products and services. However, when it comes to regulation of ownership for non-personal data, the existing legal frameworks have proven non-sufficient, see Chapter 2.4 Control of Data . Therefore, the need for contractual agreements other than Terms & Conditions Agreements is evident. This is as well the reason for why all investigated companies use Contractual & Licensing Agreements, see Table 5.1, in relation to transactions of data during collaborations and partnerships. However, it was confirmed during the main interviews that the laws regulating data during transactions are still not extensive enough and have not been tested to the extent that actors know how this should be applied or regulated in practice. Therefore, it can be concluded that there is still ambiguity and uncertainty when it comes to data transactions or data sharing between different actors in the industry.

Another important perspective in relation to ownership and utilization rights regarding data is the Data Source perspective. As soon as the company has an inflow of data, a transaction is taking place which is however dependent on where data is gathered from, i.e. the nature of the transaction. In relation to this, there are different aspects that need to be considered. As shown by Figure 4.1, in Chapter 4.1 Investigation of Data Categorizations, the data input of a company can either be primary or secondary. If it is primary, i.e. generated from the company's products and services, the ownership and utilization rights can effectively be managed through a terms and conditions agreement, which must be agreed to in order to use the product or service. However, if the data input is secondary, i.e. provided, acquired or freely available, it can become more complicated. Freely available data can often include licensing agreements, such as open source licenses, with various terms that can have detrimental effects on the company if not considered carefully before implementation. The term *acquired data* implies an exchange, meaning that the data is acquired by offering the data provider something in return. This often occurs during collaborations or partnerships with third-parties. For this reason, determining the terms of the transaction through contractual agreements is crucial in order to clarify e.g. who owns the data, who can utilize it and how it can be utilized. Lastly, an example of a transaction containing provided data is, when a user enters e.g. personal data, such as name, birth and email address, or customizes the settings. In some cases, this is regulated through accepted terms and conditions of the product or service but sometimes there is a need to explicitly state the terms of the transactions.

Secrecy

Trade Secret Management is a control mechanism used by three of the investigated companies, and the only identified mechanism related to Secrecy. As discussed in Chapter 2.4 Control of Data, Trade Secret Management does not grant any exclusive rights, but can still be an effective method to use when the firm does not want to disclose certain information. Therefore, this control mechanism can be especially relevant when the firm has performed analytical activities and developed business-critical knowledge from their data. However, to maintain and uphold this type of control the company must invest in strategies to manage the necessary processes. As discussed previously, the company can promote innovation within the company by enabling sharing of data and making data more available. However, the more freely shared data, the more difficult it is to ensure that business-critical information is not shared outside the company.

Market Power

Market Power includes mechanisms related to the firm that provide abilities to exert control over other actors in the value chain. In the company investigation, this control mechanism primarily appeared as method to secure continuous access to data, which in turn often led to that the data increased the company's control position. As indicated in Table 5.1, four additional control mechanisms related to Market Power were identified, although they were not mentioned during the interviews. The definitions of these can be seen in Table 5.2.

Table 5.2: Additional definitions of identified Control Mechanisms

Identified Data Control Mechanisms	Definitions
Technological Leadership	Establishment of a technological value offering which is not matched by any other on the market and thus ensures exclusive access to and control of data.
Installed Base	The number of users or operating physical products that can ensure access to data.
Network Externalities	The effects that arise from the size of the installed base of a product or service which leads to the installed base increasing further, and thus ensures control of and access to data.
Switching Costs	The cost of switching to a competing product or service after its intital adoption by the user.

All the companies included in this study have been assessed to have strong established brands, either by themselves or through their holding company. However, in relation to control of data, the investigation showed that Company C and D's brand does not offer any control of data. Company A and B on the other hand, have brand attributes that strongly correlate to words such as safety and security. This in turn affects the willingness of users to give away data and the trust that collaborators and partners have in the data. In the long run, the Brand Equity control mechanism could lead to users choosing products or services based on the company's brand and how trustworthy it is in relation to data.

Company C and D, however, have a similar type of control that originates from a different mechanisms which is recognized as Technological Leadership. This means that Company C and D have established a control through their value offering, i.e. technology, which is derived from the data and is distinct from any other technology on the market. Similarly to Brand Equity, this provides the companies with the opportunity to gather and process data, meaning continuous access to data, not due to trustworthiness but purely due to technological superiority. Furthermore, all four investigated companies have a large established installed base; whether it is physical products or users. This, in itself, provides a strong control mechanism as it grants market power and, for data-driven companies, ensures access to data from the installed base. This is true for all four companies as they all gather data from their respective installed base. However, only company C and D, the DBCs, have managed to utilize Network Externalities to a large extent based on their installed base. This indicates that the value and benefit of the products and services offered by these two companies increase as the installed base grows, which in turn incites more users. This loop ensures exclusive access to and control of new and more data, which in turn strengthens the company's market position further.

Another important control mechanism that did not emerge during the main interviews, but that was identified as critical for all four companies, is Switching Costs. For Company A and B, the users can experience a cost of switching to a competing product or service due to the investments made in an expensive physical product. Besides the time investment of getting familiar with the technology, other necessary investments include for example money and physical adaptions. Company C and D have managed to achieve the same effect, i.e. the control of Switching Costs, without the user having invested a significant amount of money or physical adaptions. Instead, these companies have increased the Switching Costs for the users thorough personalization of the product and service. This means that by utilizing their user data, Company C and D have made it difficult for users to switch to a competing service because it has been strongly adapted to the user, which has immensely increased the value of the service. By utilizing data, Company A and B could extend the Switching Costs in the same way, in order to increase their control position and not only rely on the monetary or physical Switching Cost of their products. In this way, the users will have made various types of investments that prevent them from switching to a competing product or service, no matter their monetary capabilities.

Right Based Property

Right Based Property was not mentioned as a distinct control mechanism for data assets during the interviews, neither was it identified during the investigation of the companies. The most plausible reason for this is, as mentioned in Chapter 2.4 Control of Data, that IPR regulations have not been adapted to data and can therefore not provide adequate protection for or direct control of data assets. Copyright could be used to protect specific program code, but in most cases during this study, it has been the volume, flow or utilization of data which has been the source of value and not the specific code. It should also be mentioned that copyright provides a relatively weak protection since it is fairly easy to create a different source code that achieves the same purpose as the original, which the copyright protection does not prevent. Similarly, patent protection provides strong protection for technical inventions that utilize data, such as systems or processes. However, this does not provide direct protection of data. Furthermore, the applicability of IPR claims is affected by the level of data structure. The Sui Generis Right can be valuable as it can protect the content of a database. However, it requires proof of a substantial financial, material or human investment in the creation of the database, which essentially means that the data must have been processed and structured to a great amount. This was, however, not something that emerged during investigation. In addition, it is unclear how control mechanisms related to Right Based Property is affected by the implementation of AI and AI generated data. This since the property is not directly created by a human, which is one of the main requirements for retrieving Rights Based Property. However, one could suppose that the AI is implemented by a human, and therefore can be used for Right Based Property control. This is an interesting finding, as there have been some cases where the Right Based Property has been discussed as not feasible protection, due it not being human created. However, for the purpose of this project, IPR protection should conclusively not be ignored as a means to control and protect data, but it should not be assumed to provide enough protection on its own.

5.2.2 Control Mechanisms in Relation to Data Assets

The importance and relevance of the control mechanisms change throughout the data value chain, meaning that a certain type of control might be more relevant for a certain type of data asset. This is illustrated by Figure 5.4, where Technical Control, Contract Based Property and Secrecy are considered the most relevant control mechanisms to use in relation to the identified data assets.

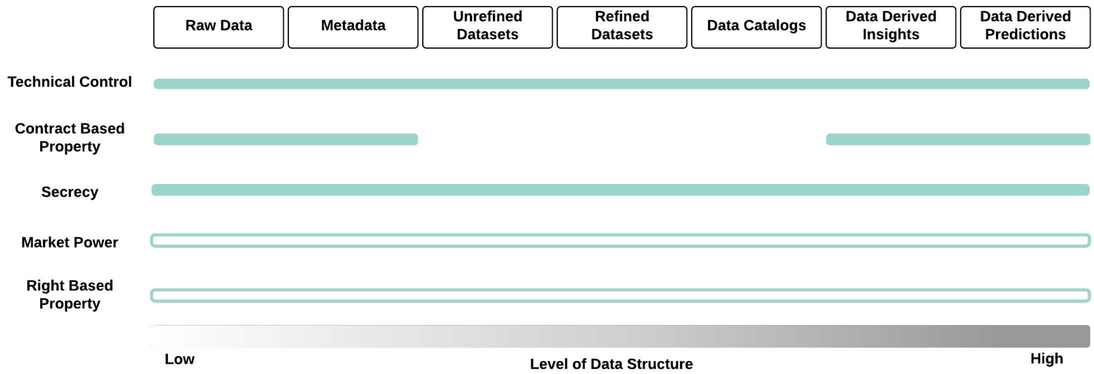


Figure 5.4: Analysis of the most relevant control mechanisms in relation to the Data Assets

As presented in the Table, Technical Control is an important control mechanism to use in relation to all seven data asset categories. For instance, encryption is important for Raw Data and Metadata since these assets have data that is unstructured or semi-structured and therefore easily manipulated. In addition, encryption is crucial in order to enable utilization of personal data. Another important control mechanisms for all data assets is Access Management, which can be used to prevent unauthorized access. Tracking of data is especially important during the transformation from Raw Data to Unrefined and Refined Datasets in order to increase transparency and facilitate troubleshooting if, for instance, data has been misinterpreted. Control mechanisms related to Contract Based Property are, as previously mentioned, necessary when data transactions occur. Therefore, this is assumed to be most relevant for the control of Raw Data, Metadata, Data Derived Insights and Predictions, since these assets are the most commonly transferred.

Secrecy has been acknowledged as a crucial type of control mechanism for all seven data assets, but especially when the firm develops valuable knowledge which can grant strong competitive advantages, for instance regarding user behavior or future trends. Thus, the importance of Secrecy increases with the amount of competence, analysis and processing that has been applied to the data. In addition, Secrecy is crucial in relation to collection and management of personal data as well in order to stay compliant with data privacy and integrity regulations. Since Right Based Property can be difficult to apply to data assets, Secrecy becomes an even more imperative control mechanism. Although this mechanism does not grant exclusive rights to data, it is an effective method of ensuring that valuable information is not shared with competitors or other actors and that the firm can maintain their competitive advantages.

Market Power is a significant type of control mechanism in relation to all seven data assets since this type of control is primarily used to secure access to data. The value is in establishing a control position in the market which enables one to have constant flow of data and an established process for how to manage it. However, this is considered less important in the Figure compared to the other control mechanisms as it does not offer direct protection or control over data assets in the same way as the other mechanisms do. Another aspect to consider is that Market Power can be a rather unstable type of control, as it is dependent on market structures, other actors and public opinions. This is particularly true in the current digital climate with ongoing discussions regarding privacy and integrity in relation to data, which can cause rapid changes.

Lastly, Right Based Property has been established as a type of control mechanism that can be useful in combination with other types of control. Right Based Property is particularly useful when interactions of intellectual assets or property occur since it provides a mean of establishing ownership, which is essential as it enables the transaction. However, in relation to data this type of control is considered less relevant compared to the other types. This is especially true in relation to Secrecy and Contract Based Property, which are both control mechanisms anchored in legal systems, but estimated to provide greater control and protection of data. Patent Right is not applicable to the data itself and is considered non-relevant, whilst Copyright and the Sui Generis Right could be used to establish control over databases. As all data assets are essentially stored as databases, the Sui Generis Right and Copyright can possibly be used to protect all seven data assets. The Sui Generis Right is believed to offer the most relevant protection as it relates to the content of the database. However, it is unclear what constitutes as a "substantial investment" in relation to the creation of a database, and therefore the applicability of the Sui Generis Right cannot nor will not be established in this report.

5.3 Value Creation from Data Assets

The abilities to create value from data has been one of the most discussed perspectives in the interviews. The interviews indicated that there are several ways to drive value from data and that the investigated companies work with this in different ways.

For a business, it is important to understand how data can create value since the value is not necessarily restricted to economic or monetary value. For instance, by understanding how data can be used to create customer value, the company can improve the customer experiences which leads to more satisfied customers and, in turn, ensures that the customers will continue paying for the service. Another example of this is sharing data across a large corporation which can lead to increased transferability and availability of the data. This can, in turn, increase the product and service development processes, which leads to improved internal business processes and a more transparent and cooperative company culture. By utilizing this kind of thinking, a company can create more than only economic and monetary benefits from data. The identified utilization areas that create value for the investigated companies are the following five main categories: Sales of Data, Internal Process Optimization, Product & Service Development, Maintaining Sustainable Relationships and Sharing & Exchange of Data. These were captured and explained in the Theoretical Framework, Section 2.1.3 Data Monetization.

5.3.1 Value Analysis in relation to Data Assets

Figure 5.5 illustrates to which extent the various data assets can be relevant to use in relation to the utilization areas. These findings were captured through the comparative multiple case study as well as the empirical findings in regards to the main data assets used in companies.

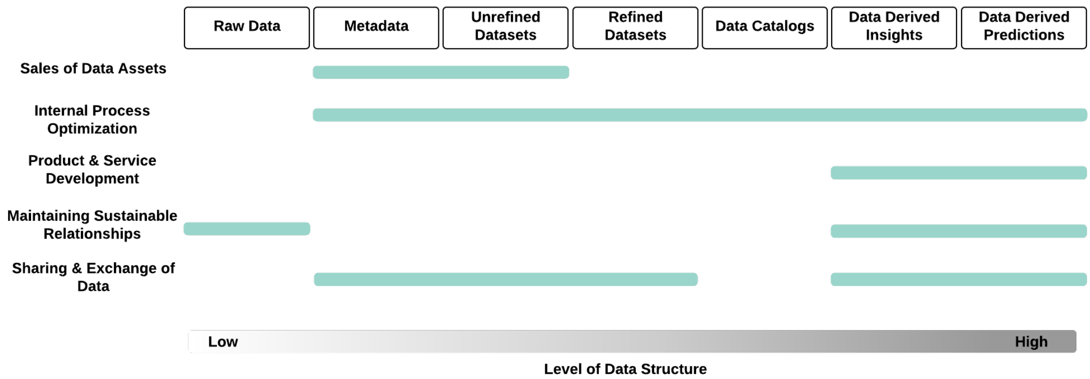


Figure 5.5: Identified value creation opportunities in the investigated companies in relation to the Data Assets

Sales of Data Assets

This utilization area includes utilization of data through either sales of assets or data insights. This was one of the few utilization areas where the investigated companies had not focused on creating value from data. Company C was the only company that had mentioned that they utilize this possibility, as the company sells or distributes Metadata or Unrefined datasets externally and internally. There are many strategies where sales of data assets can be used to create value in a business. However, these strategies have not been identified within the investigated companies. One strategy that could be used, is to sell data derived insights and predictions to other collaborating companies in order to enable more value co-creation between the actors. This can enable companies to utilize their assets in new ways and co-create new types of assets, which could be an interesting strategy to use depending on the company business model. Another strategy could be to sell or license Raw Data or Metadata to other companies and to create value by providing data to other companies that need it and through that generate royalties. However, in order to create value in this way, the data needs to be unique and not freely available. By incorporating licenses or sales of Raw Data or Metadata, the company can create more business areas and new business capabilities. Sales of Data Assets can be applied to any of the given data assets depending on the company business model. The utilization area can enable value through many various ways such as new business areas, new monetary streams, extension of the customer base and additional sales or provisions of data. It can also enable value in the form of control such as a stronger market position, e.g. by setting a new data standard within the given industry area. This can be either through a technological perspective or a data perspective, where the company has the ability to set an industry standard in how the data should be utilized and managed.

Internal Process Optimization

This utilization strategy includes utilization areas such as data enrichment and improving business processes. Data is utilized to optimize the business processes, to ensure high quality data and to create value by internally aggregating the data to enable a consolidated data landscape throughout the company. The data assets that can be relevant to use with this utilization strategy are all except for Raw Data. Raw Data could be used to optimize business processes in general, but was not identified as a utilized nor relevant asset for this strategy in the company investigation. The remaining data assets were found to be used within the companies to different degrees and purposes. Company C and D, the DBCs, worked a lot in this field by creating value and optimizing their products with all identified data assets. This was done to optimize the flow of data in the company, to ensure high quality data and to create a more transparent way of working with data within the company. Company A and B, the ICDTs, work with optimizing internal business processes through data derived insights and predictions, but have however not utilized this to a large extent compared to the DBCs. The ICDTs are focusing more on ensuring high quality data within the company, but have not yet implemented optimizations of data flow and transparency. Company D was the only company that utilized data catalogs through this utilization area, by creating a larger data transferability and transparency within the company. This asset is estimated to become an valuable asset to implement as ICDTs and DBCs become even more data reliant, where the ability to share data across the company will be essential in order to drive value creation from it. The remaining addressed assets within this utilization field have large capabilities to be valuable within the context of creating higher transparency and data quality to optimize internal business processes. An example of this is how refined datasets can be used to set an internal standard across the company regarding how the datasets should be utilized and shared across the company. This in turn can lead to data being more transferable and can help minimize duplicate datasets and thereby ensure higher data quality.

There are many ways that internal process optimization can create value for the various identified assets. This means that the utilization area can create value in ways such as improved internal company value creation, larger availability of data across the company, increased data quality, reduced costs, increased sales, increased internal productivity, improved systems for decision support and improved detection systems for inconsistencies and fraud.

Product & Service Development

This utilization field includes utilizing the data for product and service optimization, product or service innovation, contextualization or individualization. All of the above mentioned areas relate to optimization of the existing products or services and utilization of data to either make the product or service better, or to create better experiences for the users. This utilization area requires both investments of competence and structured data, which is the reason for Data Derived Insights and Data Derived Predictions being the only relevant assets to use. This utilization area is implemented in all companies, however, to very varied degrees. Company C and D incorporates this in their current business the most by using data to improve the products and services and also to create recommendation features and personalized recommendations to the users. Company A and B have implemented some of these strategies to various extents, e.g. optimization of the existing products and services or innovating around new business areas. The ICDTs have not yet focused on implementing contextualization or individualization, which can be because their businesses do not necessarily revolve around a user utilizing the product or service. In comparison, the business of the DBCs revolve around a user utilizing a product making it easier to create individual recommendations. However, some of the interviewees mentioned that the ICDTs have started to utilize predictive maintenance, which can be addressed as a personalized experience, where a user or a company can get predictions in how long their products or services will last until maintenance is needed. The ICDTs are assumed to, in a near future, also have implemented the full capabilities in order to utilize their data and enable fully optimized product and service developments.

In order to utilize these assets in a firm, it is firstly very important that the Raw Data and Metadata maintain the structure and quality standards needed. This is because these assets are used as a basis for the Data Derived Insights and Predictions. It is also assessed that the employees need to have large analytical capabilities and competence in order to understand the customer demand and need. This is a valuable intellectual asset within the firm, which is captured through the two data assets, Data Derived Insights and Predictions. It is important that the data is well structured and has high quality, since without these standards, it will be difficult to create value through the product and service development utilization area. These parameters are very important in all of the given utilization areas, but they are even more important in this case where it sets the foundation for the results retrieved from the Data Derived Insights and Predictions. This utilization area can create value such as improved products and services, new products and services, new business segments, improved customer experiences, personalized or customized recommendations, increase of sales and price optimization.

Maintaining Sustainable Relationships

This utilization strategy includes data privacy and control guarantee, as well as strengthening and building of customer relationships. Data is utilized by leveraging upon the creation and maintenance of sustainable relationships with the customer in various ways. This utilization area requires that the data asset leverages and utilizes customer data in one way or another. The most relevant data assets to utilize in this field is Raw Data, Data Derived Insights and Data Derived Predictions, as these are assets all related to customer data. These assets have also been identified in the investigated companies. However, they are used to various extent for the utilization area of maintaining sustainable relationships. The Raw Data asset category was utilized as it includes either internally or externally collected customer data. The companies collecting raw customer data need to be compliant with GDPR regulations, where the customer for example should be able to request deletion of their data and the company is required to have consent from the customer to use the data (General Data Protection Regulation, 2016). By doing so, the customers gains trust and loyalty towards the company, and gives the company control of their data, which enhances their relationship. This was found to be done at all companies that utilize customer data, which can be addressed as a way to maintain sustainable relationships. In conclusion, besides being compliant with GDPR there is also business value to gain, as it leads to an increase of the company's trustworthiness.

The assets Data Derived Insights and Predictions utilize customer data as they can be used to create insights and predictions about the customers' needs and demands. This can lead to improved and maintained sustainable relationships as the customer satisfaction increases, which in turn leads to customer retention. This utilization field is used to a various extent within the investigated companies where the DBCs have come further in implementing AI data-driven insights and predictions. The ICDTs worked with this type of utilization as well, but their focus is still on manual data analysis, rather than AI driven methods. However, as the ICDTs are becoming more data-driven, this will create a need to create more automatic processes and competence in utilizing these types of methods, both to be in line with competitors but also to provide as enhanced experiences as the DBCs. Maintaining strong and sustainable relationships can create value such as customer retention, as well as increased customer loyalty, satisfaction and trust. In addition, this can also lead to increased recurring revenue, increased market share and improved brand reputation.

Sharing & Exchange of Data

This utilization strategy includes data bartering and strategically opening data, where data is utilized either by being shared or exchanged internally or externally in return for other assets. The assets that were found to be relevant to use from the interviews, were assets that were either semi-structured or structured. Unstructured data can be valuable to share or exchange, but was not identified as an asset that was used for this purpose. The data assets identified being shared or exchanged within the companies were Metadata, Unrefined Datasets, Refined Datasets, Data Derived Insights and Predictions.

The DBCs worked with sharing and exchange of data through various parts of their company value chains or internally. The interviews confirmed that DBCs have come much further in sharing the data across the company, and they had identified that this creates high business value. However, the different DBCs shared and exchanged different types of data. Company D works to a larger extent with sharing Metadata, Unrefined and Refined Datasets, as it is a part of their business model. Company D, however, shares and distributes all above mentioned assets internally across the company, in order to create an environment that fosters innovation processes. The company also shares and exchanges data with their suppliers, however, not to a large extent with external collaborators. Company C works to a larger extent with sharing and exchanging data both internally and externally. From the interviews, only Company C worked with strategically exchanging data with third parties for other valuable assets. This can be an interesting way of leveraging data as it can enable new partnerships and value co-creation between third party collaborations. It will, however, be important to regulate the access and rights to the data, as this becomes more important when more parties are involved. It will be important to discuss the terms and conditions of the data in such collaborations, and to ensure that both parties are on the same page so that the collaboration enables more value creation. Sharing and exchange of data can create value such as new data insights, tools or services that are captured through the exchange of data. It can also create value through new partnerships, increased company visibility, collaborative co-creation, abilities to share costs between partners sharing data. The ICDTs have, in general, not yet implemented this utilization strategy to a large extent. However, as described above, strategically sharing data across the companies can be of high business value as it enables the company to work more united and share insights and knowledge.

5.4 The Correlation between Value and Control

There is a strong correlation between the value creation and control mechanisms when discussing data assets. The value and control affect and inflict each other in many ways. From Figures 5.4 and 5.5 above, it is possible to detect that various utilization approaches require different control strategies. Product & Service Development for instance, utilizes Data Derived Insights and Predictions with the purpose of providing the user with better products or services and improving the experience. All five types of control mechanisms are applicable to these two data assets, but for this specific utilization area, some are more crucial. In cases where Data Derived Insights and Predictions are used to provide e.g. individualization, there is a transfer of valuable information occurring. Thus, there is a need to establish the ownership of the data assets in use through contractual agreements, to restrict the access using access managements systems and to ensure that the valuable data assets used are not shared outside the company through trade secret management. Similarly, Sharing & Exchange of Data entails transferring of data assets, either internally or externally, and thus requires explicit legal anchoring to enable the transactions. Therefore, control mechanisms related to Contract Based Property and Right Based Property are valuable to use in a control strategy. These control mechanisms enable contextualization of the data assets in question, which makes them manageable according to intellectual property and contractual legal frameworks. Furthermore, Internal Process Optimization utilizes Metadata, Unrefined and Refined Datasets, Data Catalogs and Data Derived Insights and Predictions. This utilization area can enable unique and valuable processes for data management, which can be the source of strong competitive advantages. Therefore, this type of utilization requires to apply Secrecy to the relevant data assets in order to keep valuable information within the organization and maintain the firm's competitive advantage.

The identified seven data assets can be constructed in the form of a pyramid where the value of each asset increases the higher up it is, see illustration in Figure 5.6. One reason for this, is the simultaneous increase of *competence intensiveness*, meaning that the higher up the asset, the more competence is required for the creation of that asset. This correlates to the definition of data, as discussed in section 2.1.2 The Definition of Data, where data described as a transformation in various steps before finally becoming wisdom. This transformation occurs through various degrees of processing, analysis and internalization, meaning that the more data is comprehended and contextualized, the more it evolves and changes. Therefore, it is clear that the value of the data assets increases the more competence that has been applied, since the assets become more unique and contextualized.

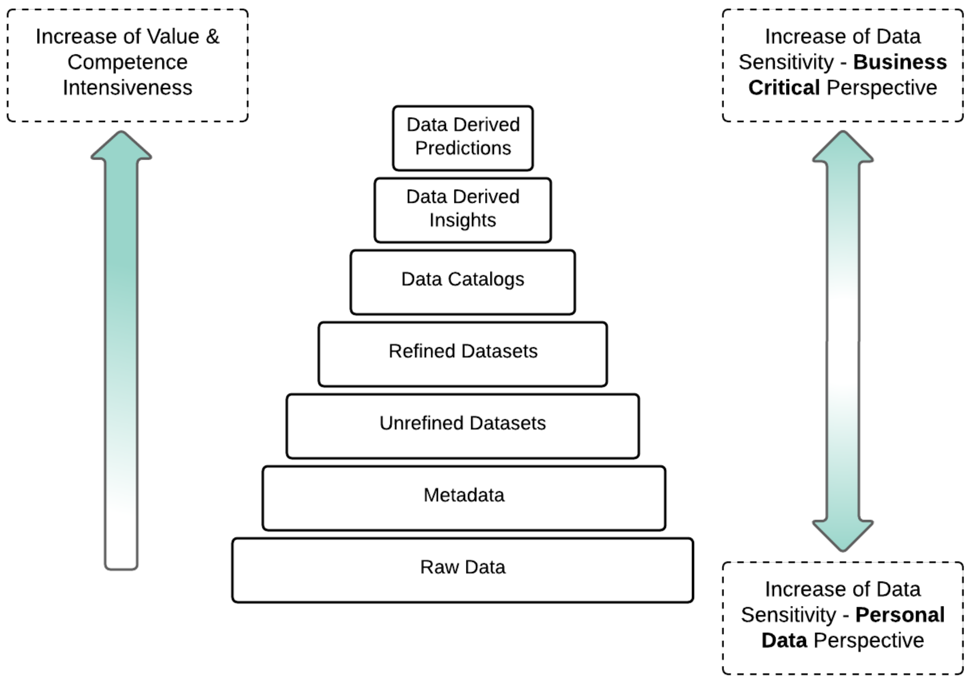


Figure 5.6: The Data Pyramid

The source of sustainable competitive advantage is the possession of assets that are considered rare, valuable, inimitable and non-substitutable. With the increase of applied competence to data, the data assets become more valuable to the company because they are put into context, but they also become more rare and difficult to re-create or substitute. If a company has applied specific competence or algorithmic calculations to their data, the company will essentially be in possession of unique data assets, i.e. insights or predictions, that provide a strong sustainable competitive advantages. This reasoning can be further developed into a statement that sustainable competitive advantage can either come from Raw Data that in itself fulfills the four criteria, or by creating other data assets that fulfill the criteria through competence appliance. Company C, for instance, uses freely available Raw Data, i.e. not rare, inimitable or non-substitutable. But, through the collection, processing and structuring of the data, the company transforms the data into Unrefined and Refined Datasets that fulfill the criteria. This is the source of Company C's sustainable competitive advantage which has enabled them to establish a strong, unique and valuable position on the market.

Two additional concepts that emerged during this study are *static data* and *dynamic data*. Static data is defined as data which has value in itself and is not prone to much change over time. Examples of static data include provided personal data, e.g. name and birth date. Dynamic data on the other hand, often lacks value in itself, meaning that is is either valuable as part of a larger volume of data or that it becomes valuable through processing. In this case, the data serves a certain purpose but is then rendered obsolete. Dynamic data also changes quickly and often, meaning that it usually is characterized by a continuous data flow. This type of data can for instance be generated user data, from a product or service. If data is static or dynamic, can have implications on both the value creation and control of data. In general, dynamic data requires a larger competence investment compared to static data, in the form of e.g. processing and analysis in order to generate value. As mentioned, dynamic data is often most valuable in large volumes, meaning that the company must gather, process and synthesize the data before it creates value. In addition, static data usually requires the same level and type of control throughout its utilization, while the control needed for dynamic data can change because the data is continuously transformed into different data assets.

The double-headed arrow on the right side of The Data Pyramid, in Figure 5.6, illustrates that with the increase of competence intensiveness and value, the data assets also become more sensitive from a business-critical perspective. This indicates that the higher up the data assets are in the pyramid, the more valuable they are and important for the company, which in turn indicates that it as well becomes important to ensure protection and control of the assets. Put simply, the data assets that are the strongest source of the company's sustainable competitive advantage are most important to protect. Therefore, these are considered *Business Critical Data Assets*. Another finding from this study is that personal data must, in most cases, be encrypted in order to fully enable utilization. This is also illustrated in the double-headed arrow on the right side of Figure 5.6, which shows that the data sensitivity from a personal data perspective increases as the data assets are less structured and further down in the pyramid. Thus, as the personal data is processed and encrypted, this sensitivity decreases whilst the business-critical sensitivity increases.

5.5 Comparison between ICDT and DBC

There are many similarities and differences that can be found in DBCs and ICDTs in relation to how data assets drive value and control. When comparing the businesses, an overall finding is how the difference in utilizing data in the companies can affect the companies' abilities to create value from data. This has been a key finding in all four research areas, where the companies utilize, collect and control their data to various extents.

In general, DBCs have come further in the progress of creating value from data in relation to ICDTs. However, ICDTs are starting to understand the value that data can create which leads them to start collecting more data and finding more utilization areas. Many DBCs also utilize the capabilities to create value from data by combining human generated data and data generated by incorporation of AI. This was not as implemented within ICDTs, where many processes relied on manually generated data. In general, the ICDTs have started to make more progress in utilizing data to enable value creation. However, they still have a long way ahead until they reach the same capabilities as the DBCs.

Similarly, regarding control of data, a general finding in the comparison between DBCs and ICDTs is that, although all companies might have implemented similar control mechanisms, DBCs have come much further. For instance, all companies have contractual agreements and company guidelines regarding data, but it is clear that these control mechanisms are much more robust and thought-through at DBCs. However, when ICDT start to implement more utilization strategies it will be assumed that the differences will not be as large between the companies utilization of control mechanisms.

A major key finding in relation to control of data, is the difference between how DBCs and ICDTs work with data sharing. The DBCs have purposely lowered the data-sharing barrier throughout the company to enable simple access to data. In doing so, they have promoted a more innovative environment. However, the lowered barriers also mean lowered control of the data. ICDTs on the other hand, have not enabled data sharing to the same extent, meaning that there are more obstacles in the way of access to data throughout the company. This means that innovation is possibly hindered and opportunities for value creation might go undiscovered, but it also lowers the risk of data being wrongfully shared and of being non-compliant with data regulations. This finding is interesting since it indicates that the companies have weighted the benefits of promoting innovation and agility through data sharing against the risks it invites in relation to control of data and regulation compliance. In this dilemma of balance, DBCs have chosen a more risk-inviting approach in order to ensure that innovation is not hindered.

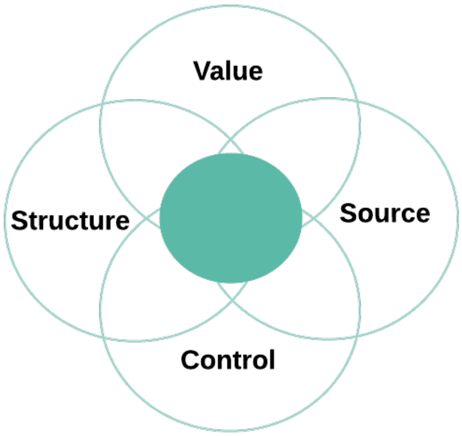


Figure 5.7: The correlation between value and control in relation to Data Structure and Data Source

When comparing DBCs and ICDTs, an intersting finding is the possibilities to control and create value from data is highly dependent on the *Data Structure* and the *Data Source*, as illustrated in Figure 5.7. If the data is not structured as needed, it is more difficult to ensure high quality data. This in turn leads to difficulties in creating value of the data as it does not maintain the quality needed for value creation. For data to be used for analytical purposes, it is of high necessity that the data has high quality since it otherwise might not create the value demanded. If the data does not maintain the quality needed, this in turn also affects the control of the data, both internally and externally. The value creation and control are also highly dependent on the source of the data as it can affect both the data quality and the value creating possibilities. When generating, acquiring or being provided data, it is vastly important to know where the data originates from and what the company is allowed to do with the data. Especially in acquisitions, where a finding is that data providing companies often try to regulate the ownership of the data in their favor. As a data-driven company, it is important to have knowledge about the ownership of the data as it directly affects the value creating possibilities. When retrieving data from various sources, it can inflict the data quality since it is most likely not in the same quality as the data within the company. This is also an important aspect to have in mind, especially in relation to acquisitions and when data is being being provided, as these can have other quality metrics. This in turn also affects the possibility to control and create value from the data. By having knowledge about the data source and data structure, this enables DBCs and ICDTs to create more value and control of their data. To conclude, it is important to have a hand in the middle of the illustration, and not focus on only one of the aspects, as this often affects the abilities to enable control and value from data.

5.6 The Data Asset Framework

The research results from this research study concluded in a Data Asset Framework that illustrates the importance of data as an intellectual asset within a firm, see Figure 5.8. The framework highlights and combines the identified major areas in the research study which are data origin, the company data flow, data utilization, data structure, data assets and lastly the control of data. By thoroughly analyzing and investigating the findings from the theoretical framework as well as the empirical results from the comparative multiple case study, these major areas have been assessed as essential for enabling value creation and control of data assets.

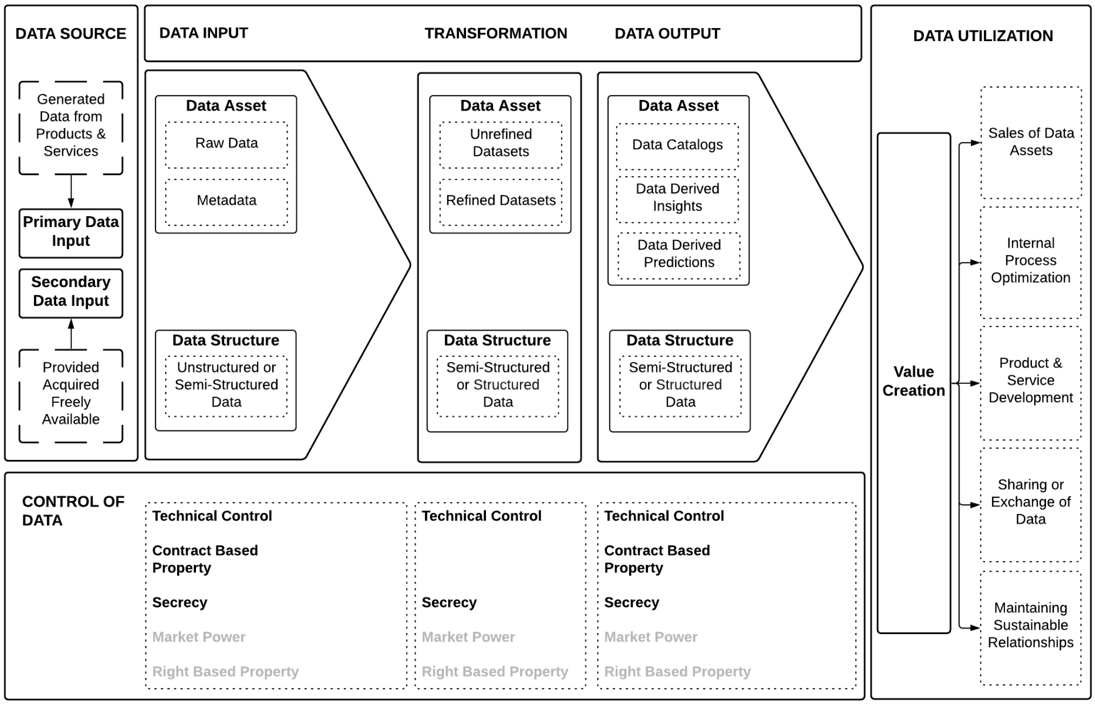


Figure 5.8: The Main Data Asset Framework

The Framework has been developed with the intention of making it as general as possible, in order to enhance its applicability. The purpose of the framework is therefore to increase the understanding of how data is gathered, transformed and utilized through the company, and how this interacts with control aspects. Thus, the framework provides clarity to the meaning of the concept *Data as an Intellectual Asset*.

5.7 Answers to Research Questions

This section will present and provide brief answers to the research questions based on the presented empirical results and analysis.

5.7.1 Research Question 1

What data resources do companies have that are important for value creation?

There are endless volumes of data resources in a firm which can be valuable for a company. The most prominent data resources identified in this research study were user data, user behavior data, raw data, statistical data, user feedback data, event data, location data, third party data, business sensitive data and refined datasets. An important finding from this research is the difficulty of managing the various perspectives in a firm in order to identify the resources that are important for value creation. However, in order to enable value creation from data resources, they need to be captured and utilized as data assets. Through the conducted literature analysis and main interviews, seven types of generic data assets that are important for value creation were identified, see Figure 5.9.

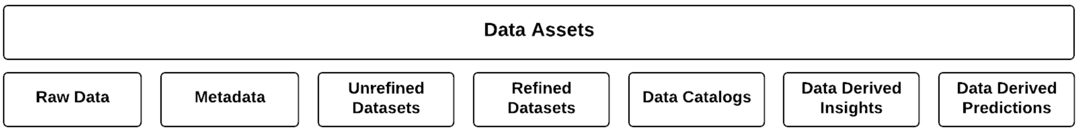


Figure 5.9: The seven generic data assets

5.7.2 Research Question 2

What control mechanisms can be relevant to use in relation to data assets and why?

The most relevant control mechanisms to use in relation to data assets are Technical Control, Contracts Based Property and Secrecy. Mechanisms related to Technical Control are fundamental for the purpose of controlling data assets. More specifically, various data encryption and pseudonymization methods are essential in order to ensure compliance with current data regulations, to enable utilization of sensitive data and protect valuable data assets from unauthorized access. In addition, various tools related to tracking data within the company and managing the access to data have been identified as crucial control mechanisms. Furthermore, in relation to regulation of ownership and rights to utilization, Contract Based Property mechanisms, such as contractual agreements, are the most efficient type of control. This applies both to transactions between the user and provider of a product or service, and between the company and other actors, such as partners, suppliers and collaborators. Control mechanisms related Right Based Property, such as IPRs, have been deemed less relevant for data assets because the control offered is considered relatively weak. Secrecy is therefore considered to be a crucial type of control mechanism that enables the company to keep business critical data within the organization and through that maintain important competitive advantages. Secrecy is also important in relation to collection of personal data to ensure that the company stays compliant with data regulations. Lastly, control mechanisms related to Market Power have been identified as significant from a different control perspective. An important identified source of competitive advantage includes an established data flow and processes needed to manage the data. Mechanisms related to Market Power can enable the company to secure continuous access to data through e.g. their brand and installed base. In addition, exclusive access to certain data types can in turn strengthen the company's market position and enable even greater access to data.

5.7.3 Research Question 3

How can data assets create value in a business setting?

Data assets can create value for companies in many various ways, both internally for the company and externally for other actors in the value chain. The identified utilization areas that can be used are Sales of Data Assets, Internal Process Optimization, Product & Service Development, Maintaining Sustainable Relationships and Sharing & Exchange of Data. The most prominent ways of utilizing the data assets were through Internal Process Optimization and Sharing & Exchange. These utilization areas enable the company to both improve their internal processes and strategically become better at sharing and exchanging data across the company. In addition, these areas can be used to improve the company's ability to work internally with data, which is an important basis to increase the utilization of the other areas. The identified utilization areas can create economic value in terms of new revenue streams or business segments, increase of sales and productivity, cost reductions and detection of fraud or deviation, which can be used to support business decisions. In addition, these utilization areas can increase customer trust and customer loyalty as well as customer satisfaction, which in turn can create a large recurring revenue. By successfully implementing more utilization areas, the data assets will create more value for the firm which can lead to new ways of leveraging upon data, as well as enable more economic benefits. It is also important to emphasize that in order to enable value creation from data assets, it is necessary to complement the utilization strategy with an appropriate control strategy that supports and facilitates the data utilization.

5.7.4 Research Question 4

How does the value creation and control mechanisms related to data assets differ between digital-born companies and industrial companies undergoing digital transformation?

The major difference in relation to value creation and control mechanisms between the companies, is that the DBCs have come further in the process of implementation. Although both types of companies have access to a large amount of data, DBCs have generally identified and implemented more ways to create value from their data compared to ICDTs. However, DBCs have been found to not utilize their data to the full potential, meaning that there are utilization areas which DBCs have not yet implemented. It is however as well important to note, that not all utilization areas are applicable to the company business. However, in this case this indicates that DBCs possess the capabilities needed to become even better in creating value from their data. In relation to control, both DBCs and ICDTs have implemented similar control mechanisms, but in general, DBCs have come much further. This is primarily made clear by the implemented control mechanisms at DBCs being more robust and thought-through compared to ICDTs.

6. Conclusion

Capturing and collecting data in companies is nowadays essential in order to be part of the digital transformation that is occurring as society is becoming more digitalized. Simultaneously, it is as important to understand how to manage and control data to be able to drive value and leverage it. As industrial organizations start to undergo digital transformations, it becomes essential to identify and capture data assets in order for them to drive value from data. However, this is also true for digital-born companies since, as this study has shown, these companies do not utilize or create value from their data to the maximal extent possible, even if they in comparison are much further ahead.

Data resources play a vital role for companies' value creation as they set the foundation for the value creating opportunities. In order for an organization to create value from data, the company must ensure that it has the internal capabilities to collect data. However, as this study has shown, it is not enough to collect a large amount of data; the company must also have strategies for converting the data resources to data assets. Thus, the company needs to understand how to manage data, as well as establish a common vocabulary regarding data. In addition, it is important that the company understands how to utilize and leverage data in order to create economic benefits and monetize on it. Understanding these fundamental reasons on why and how data resources can become valuable for companies sets the foundation for the conversion of converting data resources to data assets. This research study concluded that there are seven main data assets that are important for an organization's value creation, see Figure 6.1

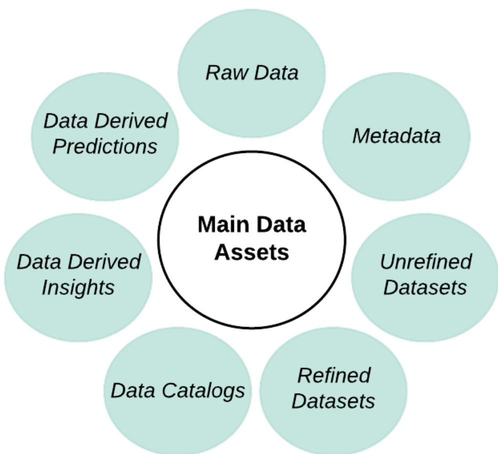


Figure 6.1: The identified main data assets

The study has shown that in order to enable value creation from data assets, a company must understand how data can be protected and controlled. It is important to recognize which control mechanisms are most relevant to use in relation to each data asset. In this research study, all recognized types of control mechanisms are *possible* to use in relation to the identified data assets. However, depending on the value creation strategy, different control mechanisms can be necessary in order to enable, support or facilitate the value creation. In general, the most *useful* types of control mechanisms in relation to data assets are Technical Control, Contract Based Property and Secrecy. These respectively refer to mechanisms such as access management systems, contractual agreements and trade secret management. The only type of control mechanisms not distinctively utilized in the investigated companies was Right Based Property, which relates to IPRs. The most plausible reason for this is that IPR regulations have not yet been adapted to suit the characteristics of data and can therefore not provide adequate protection of data. Lastly, utilizing control mechanisms related to Market Power can help ensure exclusive access to data. This can, in turn, strengthen the company's market position and lead to an increase of competitive advantages on the market. All in all, control of data essentially includes regulation of ownership rights and rights to utilize data, ensuring compliance with relevant data regulations, protecting data assets from unauthorized access and ensuring continuous access to data. Another important aspect to note in relation to control is that there are two types of sensitivity perspectives, *Business-critical* and *Personal Data*, which can require different control strategies. The more unprocessed and unstructured the data is, the more sensitive the data is from a personal data perspective, since it has not yet been anonymized or encrypted. However, the more processed and structured the data is, the more sensitive it becomes from a business-critical perspective, meaning that is even more important to protect from e.g. competitors.

Furthermore, in order to enable value creation from data assets, it is also important to understand how data can be utilized and monetized on. Thus, the company must recognize which utilization areas that can be used in relation to which data assets. The value creation does, however, not always have to result in economic or monetary value. It can for instance also lead to improved customer value or business value. By understanding how data can be used to create customer value, the company can improve the customer experience which, in turn, leads to increased customer satisfaction and willingness to continue utilizing the product or service. If companies can collect large volumes of data, various data assets can be captured and utilized for many different purposes within a firm. The main identified utilization areas that create value are the following five categories Sales of Data, Internal Process Optimization, Product & Service Development, Maintaining Sustainable Relationships and Sharing & Exchange of Data. The most prominently used areas in relation to the identified seven main data assets were Internal Process Optimization and Sharing and Exchange of Data, which included various ways to optimize the business internally or share and exchange data across the company, either internally or externally.

Furthermore, when comparing the two types of investigated businesses, ICDTs and DBCs, it becomes clear that there are many similarities and differences that can be found. An overall finding is the difference in utilization and value creation of data within the businesses, and how this affects the companies' abilities to create value from data. The investigated companies collect, control and utilize their data to different extents, which indicates that all have abilities to create value from data but are implementing this to various degrees.

By summarizing the key findings and results from this research, the main research question *How do data assets drive value for digital-born companies versus industrial companies undergoing digital transformation?* can be answered. Data assets can be utilized in many various ways to drive value for businesses. However, in order to enable value creation from data, it is first necessary to understand which data assets exist and how they can be captured, collected and used within a company. There are also strong correlations between value and control of data. The more structured the data is, the more competence has been invested and, thus, the more valuable it becomes. However, the more valuable the data is, the more important it is to protect it from competitors. The research concluded in the creation of a Data Asset Framework, which was developed to increase the understanding of how data is collected, transformed and utilized within the company, as well as how control of data needs to be taken into consideration. The research also concluded in a Data Pyramid visualization, which can be used to increase the understanding of how value, competence investment and control are highly linked together. By utilizing the logic of the Data Asset Framework and the Data Pyramid, it is possible to gain an increased understanding of the intertwined relationship between data, control and value.

7. Discussion

This chapter will include a discussion of the contribution of research, research limitations and lastly some suggestions for further research.

7.1 Contribution of Research

The existing literature in relation to data as an intellectual asset is heavily focused on the business models of tech-savvy companies born in the digital era. More specifically, the literature is focused on how digital-born companies create value from data including various comparisons between digital companies. This research study, on the other hand has contributed to an overarching understanding of the combination of value creation, control and technology. In addition, this study has provided a more IP focused approach which is a not well-covered research area in relation to data. The study has also contributed to highlighting the major differences in various companies utilizing data assets by comparing digital-born companies to industrial companies undergoing digital transformations. By doing this, it has been possible to realize how these different type of companies can learn from each other.

7.2 Research Limitations

As discussed previously, data and its related concepts have proved to be remarkably ambiguous and thus, have definitions which vary greatly depending on the chosen perspective and context. Although the concepts were recognized as social phenomena, difficulties have arisen from the intense level of ontological subjectivity, which were not predicted in the beginning of the study. Due to the short time-frame of the project, the research could thus not fully explore all various definitions to the desirable extent. In addition, to reach the desired level of robustness of the results, the study would have required a considerably larger research sample, i.e. more and longer interviews, both in the primary phase of data collection and in the pressure-testing phase, and also more companies to investigate. This is especially true for the two companies which were not as deeply investigated. The restricted time-frame of the project was therefore again a main limitation affecting these aspect of the research. Lastly, the implications of the COVID-19 pandemic have not gone unnoticed in the development of the project. The most apparent factor has been the impact on the quality of conversation during the conducted interviews, meetings and discussions. Although the fully digital environment probably made it possible to conduct a larger number of interviews and meetings, compared to a non-digital environment, it also affected ones abilities to perceive underlying intuitions and meanings during conversations. Additionally, being limited to a digital environment entailed some difficulties in finding key people for the study, due to the decrease in networking opportunities.

7.3 Suggestions for Further Research

As this research study is one of few that investigate the relationship between data and intellectual property, there are many suggestions for further research within the field. A highly relevant research field to investigate further is the applicability of the proposed Data Asset Framework. As this was only tested on a few people within the investigated companies, it would be beneficial to continuously work with the framework in order for it to be applicable, reliable and useful. In relation to this, it would also be beneficial to broaden the perspective of the research to include more types of data assets, i.e. not only R&D data. Another area of the Data Asset Framework which could be further investigated is the possibilities to identify more valuable sub-categories in relation to the proposed data assets. It would also be interesting to further investigate the applicability of the framework in relation to various business contexts and business models.

Since a large focus of this research study was to investigate the types of data that exist within companies, the control and value perspective could be more thoroughly investigated. In this research study, the focus was to investigate the companies in relation to existing theory, and thereafter evaluate the identified assets in relation to these findings. A suggestion could be to further investigate and evaluate the value and control findings to enable even more reliable and valid results.

In relation to control mechanisms, a further research area could be to assess and investigate to which degree the laws and regulations are applicable to data and how these can be adapted to increase the applicability. This has not been thoroughly investigated in this study due to time constraint. Another interesting area for further research is to analyze the global perspective in relation to control mechanisms and legal requirements, since this study has only focused on the EU legislation. Furthermore, this research study has not analyzed and assessed personal data to a large extent. Thus, this is another area which could be further investigated, especially in relation to the global perspective as the definitions and regulations related to personal data vary greatly between jurisdictions.

References

Ackoff, R. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, *16*, 3-9.

Alchian, A. A. (2021). *Property rights.* Retrieved from `https://www.econlib.org/library/Enc/PropertyRights.html`

Allen, M., & Cervo, D. (2015). Metadata management. In S. Watkins (Ed.), *Multi-domain master data management - advanced mdm and data governance in practice* (pp. 161 –178). Morgan Kaufmann.

Baecker, J., Engert, M., Pfaff, M., & Krcmar, H. (2020). Business strategies for data monetization: Deriving insights from practice. *International Conference on Wirtschaftsinformatik.*

Barney, J. (1991). Firms resources and sustained competitive advantage. *Journal of Management*, *17*, 99-120.

Bekkers, R., & Updegrove, A. (2012). A study of ipr policies and practices of a representative group of standards setting organizations worldwide. *National Academies of Science.*

Bell, D. (1976). *The coming of post-industrial society.*

Bock, M., & Wiener, M. (2017, 12). Towards a taxonomy of digital business models – conceptual dimensions and empirical illustrations.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*, 27-40.

Bryman, A., & Bell, E. (2011). *Business research methods.*

Cavanillas, J., Curry, E., & Wahlster, W. (2015). *New horizons for a data-driven economy: A roadmap for usage and exploitation of big data in europe.* doi: 10.1007/978-3-319-21569-3

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *Management Information Systems Research Center*, *36*, 1165-1188.

Cook, D. P., Goh, C.-H., & Chung, C. H. (1999). Service typologies: A state of the art survey.s. *Production and Operations Management*, *8(3)*, 318-338.

Cusumano, M. (2010). Technology strategy and management the evolution of platform thinking. *Commun. ACM*, *53*(1), 32–34. doi: 10.1145/1629175.1629189

Determann, L. (2018). No one owns data. *Hastings Law Journal*, 1-44.

Earley Information Science. (2019). The business value of taxonomy. *Earley Information Science INC*.

Erevelles, S., Fukawa, N., & Swayne, L. (2015). Big data consumer analytics and the transformation of marketing. *Journal of Business Research*, *69*, 897-904.

European Commission. (2021). *Trade secrets.* Retrieved from `https://ec.europa.eu/growth/industry/policy/intellectual-property/trade-secrets_en`

European Patent Office. (2021a). *Guidelines for examination.* Retrieved from https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g_i_1.htmn

European Patent Office. (2021b). *Hardware and software.* Retrieved from https://www.epo.org/news-events/in-focus/ict/hardware-and-software.html

European Union. (2021a). *Copyright.* Retrieved from https://europa.eu/youreurope/business/running-business/intellectual-property/copyright/index_en.htm

European Union. (2021b). *Database protection.* Retrieved from https://europa.eu/youreurope/business/running-business/intellectual-property/database-protection/index_en.htm

European Union. (2021c). *Patent protection in the eu.* Retrieved from https://ec.europa.eu/growth/industry/policy/intellectual-property/patents_en

European Union. (2021d). *Trade secrets.* Retrieved from https://europa.eu/youreurope/business/running-business/intellectual-property/trade-secrets/index_en.htm

General Data Protection Regulation. (2016). *Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016.* (https://eur-lex.europa.eu/eli/reg/2016/679/oj)

Gimpel, H., Rau1, D., & Röglinger, M. (2018). Understanding fintech start-ups – a taxonomy of consumer-oriented service offerings. *Electron Markets*, *28*, 245-264.

Grant, R. M. (1996). Towards a knowledge-based theory of a firm. *Strategic Management Journal*, *17*, 109-122.

Gregg, M. (2006). Using snort and ethereal to master the 8 layers of an insecure network. In A. Ivernizzi & S. Elliot (Eds.), *Hack the stack* (pp. 285 –352). Syngress Publishing, Inc.

Hannila, H., Silvola, R., Harkonen, J., & Haapasalo, H. (2019). Data-driven begins with data; potential of data assets. *Journal of Computer Information Systems*. doi: 10.1080/08874417.2019.1683782

Hartmann, P., Zaki, M., Feldmann, N., & Neely, A. (2016). Capturing value from big data – a taxonomy of data-driven business models used by start-up firms. *International Journal of Operations & Production Management*, *36*, 1382 - 1406. doi: 10.1108/IJOPM-02-2014-0098

Heiden, B., & Petrusson, U. (2009). Assets, property and capital in a globalized intellectual value chain. In B. Berman (Ed.), *From assets to profits: Competing for ip value & return* (Vol. 1, pp. 275 –292). John Wiley & Sons, Inc.

Hermansson, C. (2020). Intellectual asset management in ventures. *Lecture from Chalmers School of Entrepreneurship.*

Huang, K., Dyerson, R., Wu, L. Y., & Harindranath, G. (2015). From temporary competitive advantage to sustainable competitive advantage. *British Journal of Management*, *26*, 617–636. doi: 10.1111/1467-8551.12104

Hunke, F., Engel, C., Schüritz, R., & Ebel, P. (2019). Understanding the anatomy of analytics-based services – a taxonomy to conceptualize the use of data and analytics in services. *In Proceedings of the 27th European Conference on*

Information Systems (ECIS).

Kaynak, O., & Yin, S. (2015). Big data for modern industry: Challenges and trends. *Proceedings of the IEEE*, *103(2)*, 143-146.

Liang, F., Yu, W., An, D., Yang, Q., Fu, X., & Zhao, W. (2018). A survey on big data market: Pricing, trading and protection. *IEEE Access*, *6*, 15132-15154. doi: 10.1109/ACCESS.2018.2806881

Liew, A. (2007). Understanding data, information, knowledge and their inter-relationships. *Journal of Knowledge Management Practice*, *7*.

Mayer, A., & Ritter, J. (2018, 03). Regulating data as property: A new construct for moving forward. *Duke Law & Technology Review*.

Möller, F., Azkan, C., Iggena, L., Gür, I., , & Otto, B. (2020). A taxonomy for data-driven services in manufacturing industries.

Möller, F., Bauhaus, H., Hoffmann, C., Niess, C., & Otto, B. (2019). Archetypes of digital business models in logistics start-ups. *Proceedings of the 27th European Conference on Information Systems.*

Nickerson, R. C., Varshney, U., & Muntermann, J. (2012). A method for taxonomy development and its application in information systems. *European Journal of Information Systems*, *22*, 336-359. doi: doi:10.1057/ejis.2012.26

Organization for Economic Co-operation and Development. (2014). Data-driven innovation for growth and well-being. *Interim Synthesis Report*.

Otto, B., & Aier, S. (2013). Business models in the data economy: A case study from the business partner data domain. *11th International Conference on Wirtschaftsinformatik*.

Petrusson, U. (2004). Intellectual property & entrepreneurship : creating wealth in an intellectual value chain.

Petrusson, U. (2016). Research and utilization. *Tre Böcker Förlag AB*.

Powell, T. C. (2001). Competitive advantage: logical and philosophical considerations. *Strategic Management Journal*, *22*, 875-888.

PRV - Swedish Intellectual Property Office. (2021). *Copyright.* Retrieved from `https://www.prv.se/en/knowledge-and-support/glossary/copyright/`

Püschel, L., Roeglinger, M., & Schlott, H. (2016, 12). What's in a smart thing? development of a multi-layer taxonomy.

Ritter, T., & Lund Pedersen, C. (2020). Digitization capability and the digitalization of business models in business-to-business firms: Past, present, and future. *Industrial Marketing Management*, *86*, 180-190. doi: https://doi.org/10.1016/j.indmarman.2019.11.019

Rizk, A., Bergvall-Kåreborn, B., & Elragal, A. (2018). Towards a taxonomy for data-driven digital services. doi: 10.24251/HICSS.2018.135

Saunders, M. N., Lewis, P., & Thornhill, A. (2019). Understanding research philosophy and approaches to theory development. In M. N. Saunders, P. Lewis, & A. Thornhill (Eds.), *Research methods for business students* (pp. 128 –170). Pearson Professional Limited.

Schilling, M. (2012). *Strategic management of technological innovation: Fourth edition.* McGraw-Hill Higher Education.

Siddiqa, A., Hashem, I. A. T., Yaqoob, I., Marjani, M., Shamshirband, S., Gani, A.,

& Nasaruddin, F. (2016). A survey of big data management: Taxonomy and state-of-the-art. *Journal of Network and Computer Applications*, *71*, 151-166. doi: https://doi.org/10.1016/j.jnca.2016.04.008

Soto, H. D. (2003). The mystery of capital why capitalism triumphs in the west and fails everywhere else. *Basic Books*.

Spender, J. C. (1996). Making knowledge the basis of dynamic theory of firm. *Strategic Management Journal*, *17*, 45–62.

Sullivan, P. H. (1999). Profiting from intellectual capital. *Journal of Knowledge Management*, *3*, 132-142.

Wang, H., He, J., & Mahoney, J. (2009, 12). Firm-specific knowledge resources and competitive advantage: The roles of economic- and relationship-based employee governance mechanisms. *University of Illinois at Urbana-Champaign, College of Business, Working Papers*, *30*. doi: 10.1002/smj.787

Weibel, S. L., Lagoze, C., & Wolf, M. (1998). Dublin core metadata for resource discovery. *OCLC Online Computer Library Center, Inc.*.

Xie, K., Wu, Y., Xiao, J., & Hu, Q. (2016). Value co-creation between firms and customers: The role of big data-based cooperative assets. *Information & Management*, *53*(8), 1034-1048. doi: https://doi.org/10.1016/j.im.2016.06.003

Zech, H. (2016). A legal framework for a data economy in the european digital single market: rights to use data. *Journal of Intellectual Property Law & Practice*.

Zhang, Y., Ren, S., Liu, Y., Sakao, T., & Huisingh, D. (2017). A framework for big data driven product lifecycle management. *Journal of Cleaner Production*, *159*, 229-240. doi: https://doi.org/10.1016/j.jclepro.2017.04.172

A. Literature Analysis

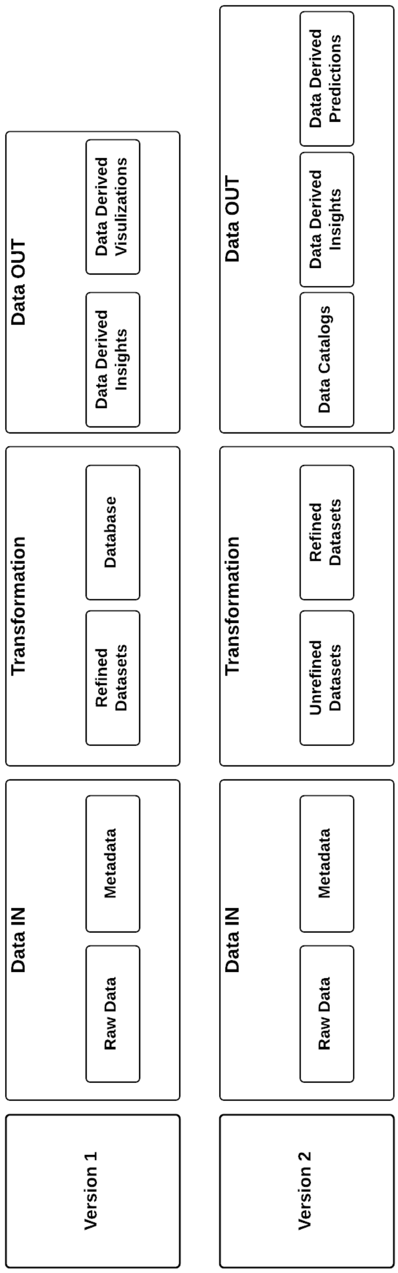
Literature Analysis Table

No.	Author/s	Publication Year	Title	Short Description	Industry Examined	Perspective	Data Categories
1	Kang Xie & Yao Wua & Jinghua Xiaoa & Qing Hub	2016	Value co-creation between firms and customers: The role of big data-based cooperative assets	The article discusses how data assets can be categorized from a consumer and a firm based perspective. The article also identifies and dicusses four types of data resources and how these resources can drive value for digital businesses.	Clothing and Furniture Industry	**Data Source**	Two general categorizations of big data, from a customer generated perspective and a firm-based perspectivs. **Customer generated involved:** Transactional big data, Communication big data, Participative big data, Transboundary big data. **Firm generated involved:** Transactional platform, Communication platform, Participative platform, Transboundary platform.
2	Philipp Max Hartmann & Mohamed Zaki & Niels Feldmann & Andy Neely	2016	Capturing value from big data – a taxonomy of data-driven business models used by start-up firms	The article investigates data as a key resource within companies. Provides a categorization of data as well as a mapping of data sources, data connected business activities and data driven business models, which is valuble in order to understand how data is used to provide value for companies	Web-based startups	**Data Source & Data Structure**	Data categorization with subcategorizations: **Internal data sources:** Existing data, self-generated data, crowd sourced. **External data sources:** Acquired data, provided data, freely available data. Freely available data can in turn be divded into two categories open data, social media data Additional category: **Value of data for customers** (raw data, information/knowledge from processed data, additional non-data products/services) *Note: This is only a segment of the framework.*
3	Sunil Erevelles & Nobuyuki Fukawa & Linda Swayne	2015	Big Data consumer analytics and the transformation of marketing	The article discusses different factors that drive value for big data within a firm, and also discusses how big data can be classified into different categories. The article also discusses the usage of big data in a firm and the competitive advantage big data can bring	General industries utilizing data	**Data Structure**	**Unstructured, Semi-structured and Structured**
4	Hannu Hannila & Risto Silvola & Janne Harkonen & Harri Haapasalo	2019	Data-driven Begins with DATA; Potential of Data Assets	The article discusses how data assets can be categorized and the importance of data assets in a firm, especially in relation to decision-making and busniess strategy	General industries utilizing data	**Data Source & Data Structure**	Data categorization with subcategorizations: **Master data** (products, customers, suppliers), **transactional data**, and **interaction data**. Another categorization mentioned in the article: **Unstructured, Semi-structured and Structured**
5	Aya Rizk & Birgitta Bergvall-Kåreborn & Ahmed Elragal	2018	Towards a Taxonomy of Data-driven Digital Services	The article discusses how data driven digital services can categorize data and to define specific characteristics in relation to data driven digital services	Digital services	**Data Source, Data Activities & Data Utilization**	Data categorization with subcategorizations: **Service Interaction** (application based, product based and embedded services), **Data aquisition mechanism** (tracking, /sensors, crowdsurfing, open data portals, secondary device), **Insights utilization** (visulization, secondary services, features/ recommendations, autonomous decision making), **Data exploitation** (information processing, advanced analytics)
6	Fabian Hunke & Christian Engel & Ronny Schüritz & Philipp Ebel	2019	Understanding the anatomy of analytics-based services - A taxonomy to conceptualize the use of data and analytics in services	The article discusses how to develop taxonomies in relation to data and analytically based services. The article also conceptualizes the utilization pportunities with data and defines characterists in relation to the concepts.	Digital analytical services	**Data Source & Data Utilization**	Data categorization with subcategorizations: **Data Generator** (Customer, non-customer, process, objects), **Data target** (internal, external), **Data origin** (Customer, Non-customer, Processes, Objects, Environment), **Analytics** (Type, Descriptive, Diagnostic, Predictive, Prescriptive), **Portfolio Integration** (Stand-alone solution, Wrapped around product, Wrapped around service), **Service** (User, Role, Recipient, Provider, Interactor). *Note: This is only a segment of the framework.*
7	Henner Gimpel & Daniel Rau & Maximilian Röglinger	2017	Understanding FinTech start-ups – a taxonomy of consumer-oriented service offerings	The article discusses how to create a taxonomy for consumer-oriented fintech startups and what key characteristics are utilized in fintech startups	Digital fintech startups	**Data Source, Data Structure & Data Utilization**	**Data Source** (User, Peer, Public), **Time Horizon** (Historic, Current, Predictive), **Data Usage** (Transactional, Basic Analytical, Advanced Analytical), **Data type** (Structured, Unstructured) *Note: This is only a segment of the framework.*
8	Michael Gregg	2006	Hack the Stack - Using Snort and Ethereal to Master the 8 Layers of an Insecure Network	A book that explains network security through a layer approach and offers an understanding of common vulnerabilities and the ways in which attacker's exploit, manipulate, misuse, and abuse protocols and applications. Chapter 9 focuses on the possible occurrence of human error as it relates to network security, meaning how to protect your company from threats enabled by the actions of employees.	General industries utilizing data	**Data Security**	**Public**: Public available in where anyone can obtain the information. **Internal**: Information not made available outside the company **Limited Distribution:** Information/Data that is ony given to a certain amount of individuals namned on a specific distribtuion list. Each copy is uniquely identified and additional copies are never made. **Personal**: This information regards an employee's individual status (e.g., employment terms, appraisals, benefit claim, and so forth).
9	Mark Allen & Dalton Cervo	2015	Multi-Domain Master Data Management - Advanced MDM and Data Governance in Practice (Chapter 10: Metadata Management)	The article discusses the importance of metadata, how to manage metadata and how metadata can be classified	General industries utilizing data	**Data Structure & Data Source**	Categorization of metadata; **Business metadata** (e.g. glossaries of business definitions and reference libraries of business rules, data qualities, and algorithms), **Technical metadata** (e.g. physical data structures, interfaces, data models, data lineage, and transformations), **Process or operational metadata** (e.g. statistical observations or documentation) Categorization of data structure: **Internal and external data sources**: structured data, semistructured data, unstructured data
10	Yingfeng Zhang & Shan Ren & Yang Liu & Tomohiko Sakao & Donald Huisingh	2017	A framework for Big Data driven product lifecycle management	The book discusses how big data can be classified based on the lifecycle of big data. The book aso discusses practical guidance and specific instructions to help companies manage the challenges of a multi-domain master data management (MDM) implementation: Chapter 10 discusses the importance of managing metadata and how it can be classified in order to facilitate the management.	Manufacturing Industry	**Data Structure**	**Structured Data** (e.g. from sensors, bill of materials), **Semi-structured Data** (e.g. extensible user manuals), **Unstructured Data** (e.g. audio, video, text)

Literature Analysis Table

No.	Author/s	Publication Year	Title	Short Description	Industry Examined	Perspective	Data Categories
11	Frederik Möller & Boris Otto & Lennart Iggena & Inan Gü	2020	A Taxonomy for Data-Driven Services in Manufacturing Industries	The article discusses and generates a taxonomy for a data driven service in the manufacturing industry. It incorporates value creation activities throughout a data driven service and adds on goal oriented data utilization and service design strategies.	Data driven services and data driven manufacturing industries	**Data Source, Data Structure & Data Utilization**	Data specific categorizations including subcategories: **Data Sources** (Self-generated Data, Acquired Data, Customer provided, Free Available), **Data Types** (Product, Process Environmental, Other) **Analytics Type** (Descriptive Diagnostic Predictive Prescriptive), **Aggregation Level** (Single, Multiple) *Note: This is only a segment of the framework.*
12	Louis Püschel & Helen Schlott & Maximilian Röglinger	2016	Taxonomy of smart things	The articles aims to study and complement the work on IoT solutions as a building block for business models, with a large focus on smart things. The study proposes a taxonomy with multi-layers for smart things with 10 different dimensions in relation to architectural layers of current IoT solutions.	Data driven service (smart things)	**Data Source & Data Utilization**	**Data Usage** (Transactional Analytical (basic), Analytical (extended)) **Data Source** (Thing state, Thing context, Thing usage, Cloud). *Note: This is only a segment of the framework.*
13	Frederik Möller & Henrik Bauhaus &Christina Hoffmann & Constanze Niess & Boris Otto	2019	Archetypes of digital business models in logistics start-ups	The paper analyzes and develops an archetypical representation of digital businesses in a startup phase from the logistics sector. It discusses and develops a taxonomy for digital driven businesses and also includes a discussion on the key sources to enable value creation in an e-business	Digital businesses	**Data Source & Data Utilization**	The framework is focused on a taxonomy from a value creating perspective, where the main categories are: Value Proposition, Value Architecture, Value Network, Value Finance. The value architecture part in the framework includes the categorization **Key Data Source** including the subcategories (Tracked & Generated, Customer, External, Multiple). *Note: This is only a segment of the framework.*
14	Aisha Siddiqa & Ibrahim Abaker Targio Hashem & Ibrar Yaqoob & Mohsen Marjani & Shahabuddin Shamshirband & Abdullah Gani & Fariza Nasaruddin	2016	A survey of big data management: Taxonomy and state-of-the-art	The article discusses a taxonomy of big data techniques, which can be used to understand which techniques that can be utilized in a business and also how to categorize the data in relation to these	Digital industries	**Data Activities & Data Security**	**Processing** (Classification, Prediction), **Pre-processing** (Transmission, Cleaning), **Data Storage** (Replication, Clustering, Indexing), **Data Security** (Privacy, Integrity, Confidentiality and Availability)
15	Maximilian Bock & Martin Wiener	2017	Towards a Taxonomy of Digital Business Models – Conceptual Dimensions and Empirical Illustrations	The article discusses a taxonomy of digital business models and describes how to categorize data and also the different value adding perspectives in relation to digital businesses	Digital Business	**Data Source & Data Utilization**	**Data analytics** (Process and product data, Customer data, (Free) external data) *Note: This is only a segment of the framework.*
16	Hsinchun Chen & Roger H. L. Chiang & Veda C. Storey	2012	Business Intelligence and Analytics: From Big Data to Big Impact	The article provides and develops a framework that identifies different industry application and emerging research areas of Business Intelligence and Analytics. Three different research areas within BI&A are defined and describes in relation to key characterits and capabilities.	Mostly digital businesses, however a large sample of different industries	**Data Source & Data Structure**	Data categorization based on different industry settings: **E-commerce and Market Intelligence Data Categorization** (Structured web-based, User Generated Content, network information, Unstructured Informal Customer Opinions, Search & User logs, Customer transaction records), **E-governance & Politics 2.0** (information source and legacy systems, Rich textual content, Unstructured citizen informal conversations, Feedback & Comments), **Science & Technology** (Sensor and network content, Instrument-based data collection,Fine-grained multiple-modality and large scale records, Specific Data formats),**Smart Health & Well-being** (Linked Content, Person Specific-Content, HIPAA, IRB, Health and Patient Social Media, Electronic Health Records, Genomics & Sequence Records), **Security & Public Saftey** (Personal Identity Information, Incomplete and deceptive content, Group and Network Information, Multilingual Content) *Note: This is only a segment of the framework.*
17	Stuart L. Weibel & Carl Lagoze & Martin Wolf.	1998	Dublin Core Metadata for Resource Discovery	Discusses how metadata can be categorized in relation to IP, content and Instantiation. This has also become an ISO standardisation, whilst this is the article which is foundational for the standard.	N/A	**Data Source**	Main category: **Metadata**, which is thereafter defined in three categories with sub-categories: **Content** (Title, Subject, Description, Type, Source, Relation, Coverage) **Intellectual Property** (creator, Published, contributor, rights,) **Instantiation** (date, format, idenfier, language) *Note: This is only a segment of the framework.*
18	José María Cavanillas, Edward Curry, Wolfgang Wahlster	2015	New Horizons for a Data-Driven Economy	The book discusses different ways of monetizing and creating value of data in relation to various industries and also how various parts of the big data value chain is utilized.	Health Care, Public Sector, Finance & Insurance, Energy & Transportation, Media & Entertainment	**Data Structure, Data Activities & Data Utilization**	Many data categorizations from the big data value chain; **Data Aquisition** (unstructured, structured, event processing etc.), **Data Analytics** (Machine learning, Information extraction, Linked Data, Data discovery etc.), **Data Curation** (Data Quality, Annotation, Automation, Interoperability) **Data Utilization** (Decision support, In-use analytics, Simulation, Prediction) *Note: This is only a segment of the framework,*

B. Data Asset Development



Version 1
Data IN
Raw Data
Metadata
Transformation
Refined Datasets
Database
Data OUT
Data Derived Insights
Data Derived Visulizations
Version 2
Data IN
Raw Data
Metadata
Transformation
Unrefined Datasets
Refined Datasets
Data OUT
Data Catalogs
Data Derived Insights
Data Derived Predictions

C. List of interviewee

List of Interviewees				Phase 1	Phase 2	Phase 3
	Title	**Role Description**	**Company Description**	**First Interview**	**Second Interview**	**Third Interview**
1	Senior Program Manager	Senior Manager for connected services in the B2B segment.	Company A	2/18/2021	10/3/2021	
2	Concept Innovation Manager	Manager working as an innovator and project team leader to create new solutions within data analytics and AI.	Company A	2/17/2021	10/3/2021	29/04/2021
3	Product Manager	Manager working with strategic initiatives and accelerating innovations by identifying new business opportunities to enable high customer value and value for the company.	Company A	11/2/2021	16/3/2021	29/04/2021
4	Vice President, Solutions Manager	Senior Manager working with linking R&D to identified customer needs and demands.	Company A	19/2/2021	15/03/2021	
5	Solution Manager	Manager working with linking R&D to customer needs and demands. More focus on holistic solutions rather than specific products.	Company A	19/2/2021	15/03/2021	
6	Product Manager	Manager for connected services in the B2B segment.	Company A	2/18/2021		
7	Concept Innovation Manager	Manager working as an innovator and project team leader to create new solutions within sustainable energy harvesting.	Company A	2/22/2021		
8	Concept Innovation Manager	Manager working as an innovator and project team leader to create new solutions within software, connectivity and access control.	Company A	2/18/2021		
9	Concept Innovation Manager	Manager working as an innovator and project team leader to create new solutions within wireless sensors.	Company A	11/2/2021		
10	Global Product Manager Software, Network, Integrations	Manager working with products and services related to software, network and integration.	Company A	19/2/2021		
11	Operational Information Security Officer	Information security officer working with key security compliance issues.	Company A	23/02/2021		
12	Director of Innovation	Director responsible for the innovation program for products and services and analyzes global trends in order to follow key innovation trends within the industry and technology offerings. Former responsibilities include data analysis and AI for products and services.	Company A		16/03/2021	
13	Product Manager	Manager working with developing and managing a cloud service for the B2B and B2C segments.	Company A		18/03/2021	
14	Software Developer	Software Developer working with system architecture for connected services in the B2B segment.	Company A		11/3/2021	
15	Chief Technology Officer	CTO working with the development of products and services at the B2B segment of the company.	Company A		22/3/2021	
16	Director, Solutions Architecture	Director working with analyzing the solution level of the products and service in relation to security, software and IT systems and helps create a holistic ecosystem of all products and services.	Company A		17/3/2021	

17	Senior Engineering	Senior Engineering Manager working with finding new opportunities to create value from the already existing products and services, especially in relation to company data.	Company D	01/03/2021		
18	Product Lead Research	Product Director working with developing new solutions for personalized customer experiences.	Company D		23/03/2021	
19	VP Engineering	Chief Architect working with the overall responsibility for data architechture, tech stack and engineering practices.	Company D	02/03/2021	07/04/2021	04/05/2021
20	Director Data Science/Analytics	Director working with managing a data science team in the financial engineering team that provides data science solutions to problems such as forecasting and predicting outcomes.	Company D		30/03/2021	
21	Data Scientist	Data Scitentist working with hardware products and services, with the purpose of enhancing the experiences through analysis of customer needs and demands.	Company D		24/03/2021	
22	VP Engineering	Senior Engineering Manager working with a focus on the suppliers of the company and has an overall responsibility for the technical aspects, procedures and standards regarding the products and services.	Company D		22/03/2021	
23	Data Program Manager - Data Infrastructure	Manager working with leading a team that develops standardized data infrastructures to ensure high quality data.	Company D		29/03/2021	
24	Senior Product Manager - Data Management	Senior Manager working closely with data and insights within the company, with a special focus on data access and data management.	Company D		06/04/2021	
25	Senior Manager, Global Affairs - Strategy & Operations	Senior Manager within Global Affairs, working with internal audit and data strategy development in relation to global business, compliance, legal affairs and data protection.	Company D		25/03/2021	04/05/2021
26	Head of Product	Manager working with the internal content of the company to facilitate and enable better product and service development.	Company D		26/03/2021	
27	Product Designer	Product designer working with data collection and the input data that is collected by the company.	Company D	11/03/2021		
28	Thesis Intern	Thesis Intern conducting a master thesis regarding categorization of data in relation to sensitivity, as well as identifying strategies for encryption and utilization of data.	Company D	12/03/2021		
29	Product Manager	Product manager working close to data collection and managing the data infrastructure in this segment of the company.	Company D			
30	Senior Legal Counsel	Senior Lawyer working with a focus on legal matters relating to privacy and the users of the company. Helps the company to navigate between various regulations regarding data privacy.	Company D		27/04/2021	
30	Senior Business Technical Leader	Senior Manager working with leading the business team and is responsible for marketing and product strategies at the company.	Company C		20/4/2021	20/4/2021
31	Innovation Leader	Manager working as a business developer with identifying new monetization possibilities in relation to data and data-driven innovations.	Company B		19/04/2021	19/04/2021

D. Interview Guide

Introduction

Short introduction of the authors and their academic backgrounds.

"This project is our final thesis project in the master program Intellectual Capital Management at Chalmers University of Technology in Gothenburg. The program is focused on how to evaluate and create value from Intellectual Assets and create business strategies anchored in intellectual property, in order to leverage their value."

Short description of the purpose of the study

"The purpose of the study is to investigate how data can be managed as an intellectual asset, meaning how it drives value for businesses and how it can be protected and controlled. The project is a collaboration with [OTHER COMPANY] and we will deliver one company specific internal deliverable each and one academic paper. We have both signed contracts and NDAs so we treat all information collected confidentially."

Description and clarification of how the data gathered from the interviews will be used, in order to ensure that consent is achieved and given for publication.

"With that said, we also want to clarify how the information you share during this interview will be used for our master thesis, which will be published at Chalmers University of Technology. Your responses will be presented as qualitative data on the perspectives of how data is used to drive business value and how it is managed. Since it is a sensitive subject, the information that is collected in the interviews will be presented with a high level of abstraction in our academic paper - meaning that there will be no attribution to a particular individual (i.e. you) or company (i.e. [COMPANY]).

With this in mind, please let us know if there is anything you want us to specifically exclude from our thesis."

Request interviewee to record the interview, for easier management of transcription on interview material.

"If it's OK with you, we would like to record the interview for our own personal reflection. We will not share it, create any transcripts or publish any quotes and

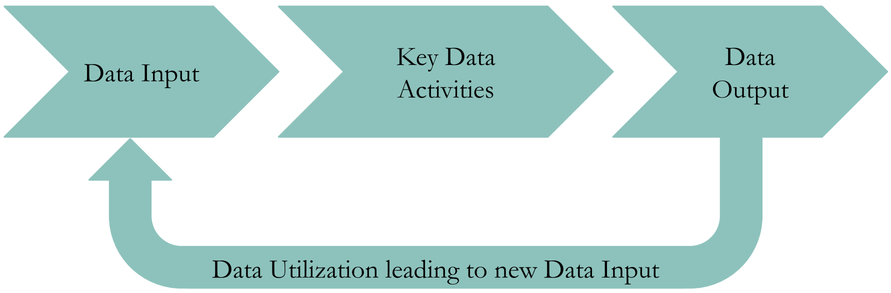
we will delete the material immediately after the project is completed in June. The purpose of recording is for us to be able to fully focus on our discussion. Is this OK for you?"

Interview Questions

About the candidate

Could you in short describe your role at [COMPANY]?

Show Data Flow Chart



"We have made this simplistic illustration of the data flow within a organization, which shows that data comes into the company in one way and undergoes a value adding transformation of some sort and then goes out again in a different form and creates value - sometimes the utilization of data can in turn lead to new input of data - based on this thinking we, would like to hear your thoughts on the data flow at [COMPANY]"

1. Data Input a) If we start at the beginning, what types of data do you know of that comes into [COMPANY DIVISION] from products and/or services?"

b) How and from where is the data collected?

c) An important aspect to consider in relation to data collection is regulating the ownership and rights to utilize the collected data. How does [COMPANY] work with regulating the data that is collected?

2. Key Data Activities

a) How is the collected data transformed, processed and structured, in order to facilitate utilization at [COMPANY DIVISION]?

b) Data can typically be controlled and protected either through intellectual property rights, such as trade secrets or copyright, or structural control mechanisms, such as contracts and access. Is this something that your team or [COMPANY] works with?

If **Yes**, through which mechanisms is the data controlled and protected?

If **No**, do you see a need for guidelines on this?

3. Data Output

a) Can you think of any type of data that is unique for [COMPANY] or that [COMPANY] has exclusive access to? What data is unique and in what way is it unique? In what ways is data utilized at [COMPANY DIVISION] to create value and for whom?

b) In relation to what we discussed earlier regarding ownership and rights to utilize data the collected data, how does [COMPANY] work with managing this in terms of output data?

4. Data categorization / Data vocabulary

a) We have read articles and concluded that we think that a large obstacle when working with data is that there is a lack of a common “data vocabulary” throughout organizations. One way of resolving this is to have a set taxonomy or categories for data - If you would categorize the data at [COMPANY], what do you think would be useful main categories?

b) In your work at [COMPANY], do you currently use an existing way of data categorization? Or do you know if there is a [COMPANY] way to categorize data in general?

DEPARTMENT OF TECHNOLOGY MANAGEMENT AND ECONOMICS
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY