Norwegian University
of Life Sciences

Master's Thesis 2021 30 ECTS
Faculty of Science and Technology

An Investment Perspective on Data Quality in Data Usage

Espen Hjelmeland and Kristin Otter Rønnevig
Industrial Economics

Preface

The master thesis is the last part of our studies in Industrial Economics at NMBU (Norges Miljø- og Biovitenskapelige Universitet). Humble, happy, and proud, we now deliver our masterpiece, which at times has been demanding. However, we are grateful that we go out of Ås with awesome friends and look back on a good and memorable time. We believe that everyone will agree that the time at NMBU passed by too quickly.

We want to thank NMBU for some fantastic years and how we have developed ourselves both professionally and socially. We look forward to using the knowledge we have learned to meet new challenges in the future.

We would like to take this opportunity to thank the people who have contributed to the thesis and their support during the process. First, we would like to thank our supervisors, Asmamaw and Per, for letting us immerse ourselves in something we were interested in and for your support and patience when needed in a challenging process. Second, we would like to thank all of the respondents in DNV, Orkla, and Fremtind. Without your cooperation, we would not have been able to conduct the study and analysis. Also, thank you DNV, Norwegian Association for quality and risk management, Brønnøysundregistrene, and Laney for the valuable feedback.

Finally, we want to thank our friends and family who have always shown interest and outstanding commitment to what we do. You have given us energy and made sure that this half-year has not only included a master's thesis. A special thanks to our parents for having aroused interest in us and given us opportunities to succeed. Thank you also for your endless support during the process.

Since this topic has international potential, we were recommended to write the thesis in English. We hope you enjoy your reading.

Oslo, August 30, 2021

Espen Hjelmeland

Espen Hjelmeland

Kristin O. Rønnevig

Kristin Otter Rønnevig

Abstract

Today, businesses must be aware of the value in their data to optimize their business operations and maximize profits. Therefore, it would be beneficial to understand how data quality affects the organization's ability to extract value from its data. The purpose of this study is to identify how an organization may maximize the return on investment in data quality. Specifically, the study analyzes how economic principles are applied to data, what factors should be considered when investing in data quality, and the potential costs of the investment options.

To address this topic, a qualitative study has been conducted with a case study about ship, position and movement data, frequently called "AIS data," involving the Norwegian engineering business DNV. Data are collected from interviews, a literature review, and a document analysis. The interviews were examined using a thematic analysis along with coding to extract themes. Results revealed that various factors influence customers´ willingness to pay for data quality and that there may be a lack of awareness of how economic principles work in the context of data quality in general. In addition, this study discovered multiple factors that can be utilized to optimize investments in data quality and determine which approaches could be used. Companies are mainly basing investment decisions on their experience in the industry.

Five factors are influencing customer's willingness to pay for data quality were found. The study elaborates on the economic principles of supply and demand and price-quality relationship to data assets. It highlights the need for further studies on economic principles evaluated to data quality. For businesses to maximize their opportunities, they need to develop clear and measurable quality criteria, involve the appropriate people in investment decisions, and invest in data management over a more extended period. Types of improvement options vary throughout the data value chain. The study identifies that fixing transponders, improving data maturity, data cleaning, and training the user to use the data is ways to improve the data quality. The cost implications of the different options will vary throughout the value chain. Further research on the topic is recommended to strengthen the findings of this study.

Sammendrag

Bedrifter i dag må være klar over verdien som dataene deres har for å optimalisere driften og maksimere fortjenesten. Det vil være fordelaktig å forstå hvordan datakvalitet påvirker organisasjonens evne til å hente verdi fra dataene. Formålet med denne studien er å finne ut hvordan en organisasjon kan optimalisere avkastningen på investering i datakvalitet. Spesielt analyserer studien hvordan økonomiske prinsipper anvendes på data, hvilke faktorer som bør fokuseres på når man investerer i datakvalitet, og hvilke kostnader investeringene har.

For å utforske dette problemet ble det utført en kvalitativ studie som inkluderte et literaturstudie og intervjuer i forbindelse med en casestudie i AIS-data i den norske ingeniørbedriften DNV. Data er samlet fra intervjuer, et litteratursøk og en dokumentanalyse. Intervjuene ble gjennomført ved hjelp av en tematisk teknikk sammen med koding for å trekke ut temaer. Resultatene viste at det er forskjellige faktorer som påvirker kundens betalingsvillighet for datakvalitet, og at det kan være mangel på kunnskap om hvordan økonomiske prinsipper fungerer i sammenheng med datakvalitet generelt. Denne studien oppdaget flere faktorer som kan brukes for å optimalisere investeringer i datakvalitet, samt hvilke tilnærminger som kan brukes. Bedrifter i dag baserer investeringsbeslutninger hovedsakelig ut fra erfaringen deres fra industrien.

Det ble funnet fem faktorer som påvirker kundens betalingsvilje for datakvalitet. Studien utdyper om de økonomiske prinsippene for tilbud og etterspørsel og forholdet mellom pris og kvalitet relatert til data. Den trekker frem behovet for ytterligere studier av økonomiske prinsipper relatert til datakvalitet. For å maksimere mulighetene fra data må bedrifter fokusere på å sette klare og målbare kvalitetskriterier, involvere riktige personer i investeringsbeslutninger og investere i data management over en lengre periode. Forbedringsalternativer vil variere ut fra hvor i dataverdikjeden man fokuserer på. Studien finner at fysisk reparasjon av transpondere, forbedring av data maturity, rensing av data og opplæring av brukeren i å bruke dataene er måter å forbedre datakvaliteten på. Kostnadene av de forskjellige alternativene vil variere i hele verdikjeden. Videre forskning på temaet anbefales for å styrke funnene i studien.

Contents

Preface **i**

Abstract **ii**

Sammendrag **iii**

1 Introduction **1**

1.1 Background . . . 1
1.2 Objectives . . . 2
1.3 Problem Statement . . . 3
1.4 Contributions . . . 3
1.5 Limitations . . . 3
1.6 Outline . . . 4

2 Literature Review **6**

2.1 Data Management . . . 6
2.1.1 Data . . . 6
2.1.2 Data as an Asset . . . 7
2.1.3 Data Management Principles . . . 7
2.1.4 The Data Value Chain . . . 8
2.2 Data Quality . . . 9
2.3 Data Quality Dimensions . . . 11
2.3.1 ISO 8000 . . . 11
2.3.1.1 The ISO/IEC 25012 Data Quality Model . . . 12
2.3.1.2 Quality Dimensions According to DAMA (2017) . . . 13
2.3.2 Data Quality and Customers . . . 14
2.3.3 Data Quality Assessment . . . 15
2.3.4 Data Security . . . 16
2.4 Economic principles . . . 16
2.4.1 Willingness to Pay . . . 16
2.4.2 Supply and Demand . . . 17
2.4.3 Price Elasticity of Demand . . . 18
2.4.4 Return on Investment . . . 20
2.4.5 Price and Quality Relationship . . . 22
2.5 Uncertainty . . . 22
2.5.1 Risks . . . 23
2.5.1.1 Costs of Poor Data Quality . . . 24
2.5.2 Opportunities . . . 25
2.6 Investment in Data and Data Quality . . . 26

2.6.1 Data Value Drivers . . . 26
2.6.2 Maximizing Return for Big Data Projects . . . 28
2.6.3 Putting Value on Data . . . 31
2.6.4 Creating Value From Data . . . 33
2.6.5 Understanding the Data . . . 34

3 Case presentation 36
3.1 Automatic Identification System . . . 36
3.1.1 Prevalence of AIS . . . 36
3.2 Use of AIS Data . . . 37
3.2.1 Emission Estimation . . . 37
3.2.2 Fatigue Damage . . . 38
3.2.3 Pipeline Damage Through Trawls and Anchors . . . 38
3.2.4 Voyage and Port Stay Information . . . 39
3.3 Quality Issues . . . 39

4 Research Methodology 42
4.1 Research Process . . . 43
4.2 Research Design . . . 45
4.2.1 Type of Research . . . 45
4.2.2 Sample Selection . . . 46
4.2.3 Data Collection . . . 47
4.2.3.1 Literature Review . . . 48
4.2.3.2 Document Analysis . . . 48
4.2.3.3 Interviews . . . 49
4.2.4 Analysis . . . 51
4.3 Reflections . . . 52
4.3.1 Method . . . 52
4.3.2 Validity and Reliability . . . 54
4.3.3 Ethics . . . 55
4.3.4 Researchers Role . . . 56

5 Results 57
5.1 What are the Main Drivers for Willingness to Pay for Data Quality? . . . 57
5.2 What Part of the Data Quality Aspects Should One Invest in to Maximize Opportunities and Minimize Risk? . . . 60
5.3 How can the Quality of Data be Improved, and What are the Costs? . . . 63

6 Discussion 65
6.1 Drivers of customers willingness to pay . . . 65
6.1.1 Drivers for Willingness to Pay . . . 65

6.1.2 Supply and Demand 67
6.1.3 Price-Quality Relationship 69
6.2 What Part of the Data Quality Aspects Should One Invest in to Maximize Opportunities and Minimize Risk? 71
6.2.1 Setting Quality Criteria 71
6.2.2 Comparing Improvement Actions 72
6.2.3 Involving the Right People 73
6.2.4 Invest on a Longer Time Frame 73
6.2.5 Invest in Data Management 73
6.3 How can the Quality of Data be Improved, and What are the Costs? 74
6.4 General Discussion 75

7 Conclusion **77**
7.1 What are the Main Drivers for Willingness to Pay for Data Quality? 77
7.2 What Part of the Data Quality Aspects Should One Invest in to Maximize Opportunities and Minimize Risk? 78
7.3 How can the Quality of Data be Improved, and What are the Costs? 78
7.4 Recommendations 78

Appendices **88**

A Interview Guide **88**

List of Figures

1 Thesis structure 5
2 DIKW pyramid 6
3 The data value chain 8
4 Data Quality Assessment 15
5 Supply and Demand 17
6 Price Elasticity of Demand 19
7 Strategic phase and operational phase in uncertainty over time 23
8 The optimal data maintenance effort 25
9 The data value drivers 28
10 Big data return on investment 29
11 Information yield curve 30
12 Data valuation methodologies 31
13 The data landscape 33
14 The steps to understand the data 35
15 AIS data value chain 37
16 Loss of fatigue capacity due to heavy weather 38
17 Abnormal vessel activity plot from three boats close to a pipeline 39
18 Research design based on maxwell's model of research design. 43
19 The research process in the study 44
20 The data triangulation validation method 48
21 The Six-Step Process for Conducting Thematic Analysis 51
22 Overview of how the different quality aspects may be ranked and compared . 61
23 The different quality improvements and the effect on ROI 61
24 The drivers for willingness to pay and the relation to the observations 67
25 Shift in demand 68
26 Supply and demand for informational assets 69
27 The Willingness to sell curve 70
28 Willingness to sell and willingness to pay curves for data quality 71
29 The data value chain and quality improvement actions 75

List of Tables

1 Data transmitted in an AIS-message 36
2 Data Quality Issues in Automatic Identification System Data 40

1 Introduction

1.1 Background

Data are one of the big buzzwords in the modern economy. They are available everywhere and are generated on nearly every digital device. It is common to regard data as an asset that may be leveraged to generate greater value. The data economy has enormous potential for economic growth and is predicted to be a major driver of economic expansion. The term ”data economy” refers to the use of data to produce commodities or services or as a commodity itself. The European Commission assumes in its data strategy (2020) that the value of the computer economy in the EU27 will increase from 301 billion euros in 2018 to 829 billion euros by 2025. In Norway, the computer economy is estimated to represent an annual value creation equivalent to NOK 150 billion in 2020. If conditions are right, these figures are expected to double by 2030 (Det kongelige kommunal- og moderniseringsdepartementet, 2021). The potential economic effect of big data investment is piquing people’s interest (Tambe, 2014). According to Manyika et al. (2013), the use of open data may have a potential yearly worth of $3 trillion in seven areas in the global economy (Education, transportation, consumer products, electricity, Oil and gas, health care, consumer finance).

Data, according to Digital Norway (2021), are an equally essential resource for businesses, and can be seen as a 5th factor of production in addition to land, labor, capital, and entrepreneurship. Therefore, the data must be taken equally seriously. It may be an IT manager who manages this area, but the company’s data are primarily an extensive management responsibility - which is not taken too seriously. Having a better focus on data management is critically important for an increasing number of companies.

Companies that recognize the actual worth of their data and use sophisticated analytic tools to harness them enjoy sustained growth (Saxena, 2019). For example, the social media giant Facebook paid $19.6 billion in 2014 to purchase the text messaging software WhatsApp. Following the acquisition, Facebook’s share price increased from $68 to $77.56. At the time, no greater transaction had ever occurred in Silicon Valley. Despite being a popular app, WhatsApp lost $138 million and had only $10.2 million in revenue in 2013. Why did Facebook pay such a high price? WhatsApp had over 500 million monthly users, with a daily growth rate of over 1 million users. Seventy percent of its users were active daily (Investopedia, 2021). It seems evident that Facebook may have placed a high value on user data. Another example is that United and American Airlines received valuations on their customer loyalty data that were multiples of the entire business itself (Laney, 2020). It is evident that data is extremely valuable to a business and may also exceed the value of the organization itself. However, evaluating the value of data has proven challenging, and no consistent methodology appears to exist. Businesses do not know the value of their data (Short, 2017).

Data quality plays a critical role in today's organizations since poor data quality can lead to low efficiency, lost opportunities, increased risk, and wrong decisions resulting in poor organizational productivity. According to S. Moore (2018), "organizations believe poor data quality to be responsible for an average of $15 million per year in losses. This is likely to worsen as information environments become increasingly complex."

The requirement for high-quality data is essential for every company that has data-based solutions for decision support. On the other hand, the connection between the expenses of high-quality data and the return on a specific upgrade expenditure poses practical issues. According to Trillium Software (2010) and Experian information solutions (2014), investing in quality to obtain a higher return is of concern to businesses since they require better accuracy from their information.

Numerous articles on the valuation of data have been published stating how data create value, such as KPMG (2019), Li et al. (2019), Rea and Sutton (2019), Diane Coyle and Stephanie Diepeveen and Julia Wdowin and Jeni Tennison and Lawrence Kay (2020), and Short (2017), to name a few recent articles. None of these articles address important concepts from the economic theories for data and data quality improvements. There are limitations of studies and publications about these economic principles for data quality. Laney (2017) specifically wrote about the economic principles in his book "Infonomics", but when Laney was asked to elaborate on the topic of economic principles evaluated to data, he stated:

> I think the application of economics concepts to data is an entirely green field. Forget about research, and develop your own ideas! (D. Laney, personal communication, June 08, 2021)

Additionally, a literature review study consisting of 128 papers on the topic of realizing value from big data concluded that "the current literature on big data value realization is characterized by a limited number of empirical studies and some repackaging of old ideas" (Günther et al., 2017, p.205). Therefore, there seems to be a need for more research regarding the value of data, the economic principles related to data and data quality, and how to make suitable investments for data assets.

1.2 Objectives

The main objectives of this thesis are to

1. Find the main drivers and factors contributing to the willingness to pay for data quality.
2. Find which data quality aspects to focus on in the investments in data to optimize the investments.
3. Find ways of improving data quality and determining the cost.

1.3 Problem Statement

This thesis aims to determine how improvements in data quality can provide value to data as an asset and the organization´s investment. The literature is weak on economic concepts for data and seems to need more studies investigating this topic. Since data can be copied and used multiple times at a low cost, it is of key interest to observe whether increasing data quality can increase data usage and better justify investments. Since the cost of producing data is more or less fixed and independent of the number of usages, the supply-demand model has been introduced as part of the methodology approach.

In addition, it is desirable to determine how to influence organizations to invest more in data quality to provide opportunities and establish a data-driven competitive advantage. Businesses need to be aware of how they should invest in data quality. Identifying which data aspects to focus on when investing in data quality, and the cost implications of different options, will therefore need to be studied.

The problem statement of this study is: "How can an organization optimize its investments in data quality?" The following research questions are developed to address the problem:

1. What are the main drivers for willingness to pay for data quality?
2. What part of the data quality aspects should one invest in to maximize opportunities and minimize risk?
3. How can the quality of data be improved, and what are the costs?

1.4 Contributions

Although much research has been conducted to determine the value of data, little is known about the supply and demand for these assets. This thesis will enhance understanding and awareness of how data value evolves and what role quality plays in this value. The research will also help determine how data investments should be made and what the business should prioritize.

1.5 Limitations

There may be some possible limitations in this study. The topic covered in this thesis is extensive, and since there have been limitations with time, the empirical findings are substantiated by individuals. The informants have been drawn from a relatively small sample size from mainly one industry. Therefore, there may be implications for generalizing the findings to a broader perspective. Having a small sample size may have given the study a less accurate result. By conducting a study with larger sample size, it may be possible to generate more accurate results. Other limitations may be that there seems to be a lack of previous research in this area, making a small study like this more uncertain. Not much numerical data are available, making it challenging to quantify the costs in this study.

1.6 Outline

- **Chapter 1 - Introduction:** Establishes the motivation for the thesis, introduces the topic, and explains why the topic will be investigated in more detail with the help of previous research articles.
- **Chapter 2 - Literature Review:** Concepts and previous research are presented for establishing the background theory from relevant areas within the field of this thesis.
- **Chapter 3 - Case Presentation:** The use case of AIS data is presented with important concepts.
- **Chapter 4 - Research Method:** The methodological approach that was used in the thesis and the reason behind the different methodological decisions that were made during the work of the study is presented. In addition, thoughts on the study's method, data collection, quality, ethical concerns, and the role of the researchers are given.
- **Chapter 5 - Results:** Contains the main data material along with the main finding from the qualitative analysis.
- **Chapter 6 - Discussion:** Presents an overall discussion of the results and theory chapters.
- **Chapter 7 - Conclusion:** Outlines the conclusion of the thesis. In addition, proposals for further research are proposed.

Figure 1 shows the structure of the thesis. First, the introduction (Chapter 1) is presented, then both theory chapters are presented - the literature review (Chapter 2) and the case presentation (Chapter 3). Then, the research methodology (Chapter 4), the results (Chapter 5), the discussion (Chapter 6), and the conclusion of the thesis (Chapter 7) are presented.

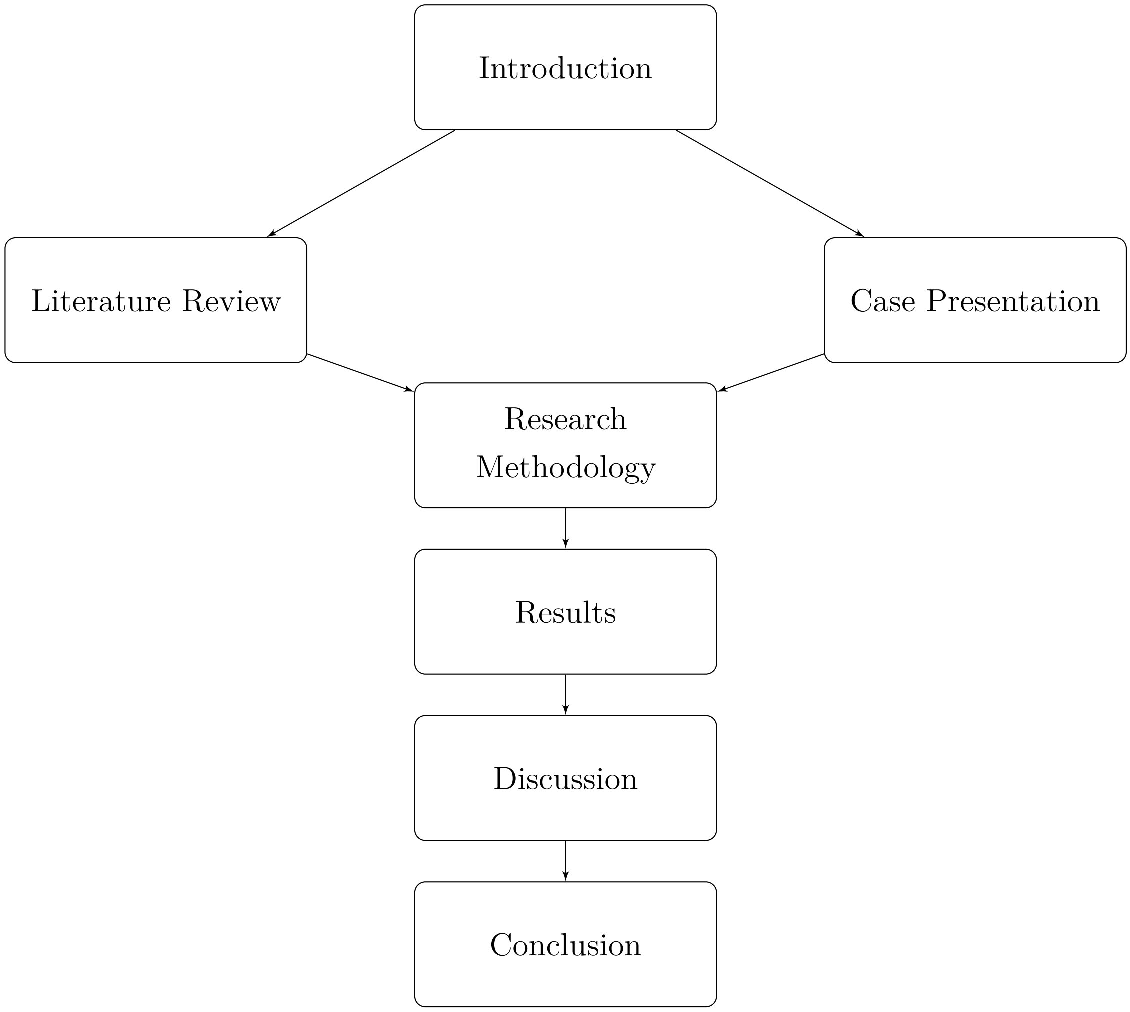


Figure 1: *Overview of the Structure of the Thesis*

2 Literature Review

In this chapter, the relevant literature on the field will be presented. This chapter includes the basic data principles and the current framework for data management, setting a value to data, and economic principles, such as supply and demand, willingness to pay, and ROI (Return On Investment), and how they are interpreted to data.

2.1 Data Management

2.1.1 Data

There is a clear link between data and information. *Data* has been called "the raw material for information," and *information* has been called "data in context." These relationships can be described by a layered pyramid where data is the base layer. Figure 2 shows information, knowledge, and wisdom above data. From this pyramid, it is easy to understand that data needs to be well-managed to use it efficiently (Ontotext, n.d.).

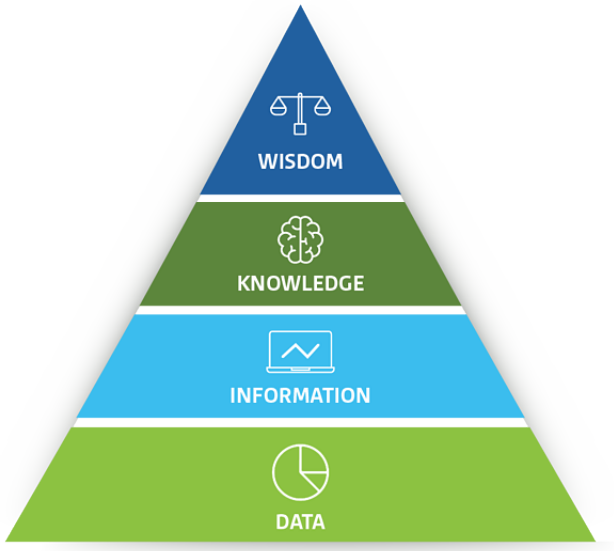


Figure 2: *DIKW Pyramid (J. Moore, 2019).*

- **Data -** A collection of information in a raw or disorganized form, such as numbers or letters.
- **Information -** The DIKW (Data, Information, Knowledge and Wisdom) pyramid's next construction component is information. This is data that has been cleaned for mistakes and further processed so that it is simpler to measure, display, and analyze for a particular purpose.
- **Knowledge -** Information is transformed into knowledge when it is used to accomplish objectives rather than simply seeing it as a collection of facts. This knowledge is often a competitive advantage for businesses. Deeper insights are obtained as connections are

discovered that are not clearly expressed as information, moving further up the DIKW pyramid.

- **Wisdom -** The DIKW hierarchy is topped by wisdom, and to get there, questions like ”why does something?” and ”what is best?” must be answered. Wisdom, in other terms, is knowledge put into practice.

2.1.2 Data as an Asset

IAS (2004) (The International Accounting Standards) define an asset as ”a resource controlled by the entity as a result of past events and from which future economic benefits are expected to flow to the entity.” They further define the critical attributes of an intangible asset as follows:

- **Identifiability -** It originates from contractual or other legal rights and is capable of being split and sold.
- **Control -** Possession of the ability to derive advantages from the asset.
- **Having probable future economic benefit -** Such as revenues or reduced future costs.

Data meets these criteria and similar IFRS (International Financial Reporting Standards) criteria, however, since 2002, IAS 38 (International Accounting Standard 38) prohibits data from being recognized on balance sheets (Laney, 2017). Since data are being treated as an organizational asset, the need for managing data is rising. What it means to manage data as an asset is still a work in progress. Even though data has no universal adoption, their monetization will become increasingly common. The data are used in organizations to make better and more effective decisions to operate more efficiently. These data-driven organizations stay competitive by making decisions based on the data instead of feelings or instinct. ”Being data-driven includes the recognition that data must be managed efficiently and with professi onal discipline, through a partnership with business leadership and technical expertise” (DAMA, 2017).

2.1.3 Data Management Principles

The concepts of data management are similar to those of asset management in certain ways. It entails identifying what data a company has and what can be done and utilizing data assets to achieve corporate objectives effectively. It must strike a balance between strategic and operational requirements. The principles of data management outlined below govern data management practice (DAMA, 2017):

- Data are a valuable resource with its own set of characteristics. Data are an asset, but it differs from other assets in important ways that affect how it is managed. One of these characteristics is that data are not consumed when they are used.

- The value of data can and should be expressed in monetary terms.
- Data management necessitates maintaining data quality.
- Metadata are required for data management. The business will need data on the asset to administer it, just like any other asset. The data asset is referred to as metadata.
- Data management requires planning.
- Data management is a cross-functional task that necessitates a wide range of skills and knowledge.
- Data management requires a business perspective.
- Data management must consider a variety of perspectives.
- Data management is synonymous with data lifecycle management.
- The lifecycle characteristics of different kinds of data vary.
- Managing data also entails managing the risks that is associated with it.
- Information technology decisions must be driven by data management requirements.
- Effective data management necessitates leadership and dedication.

2.1.4 The Data Value Chain

Value chains have been utilized as a decision support tool in business management to represent the sequence of actions that an organization conducts to create a valuable product or service (Curry, 2016). In a value chain, an organization's generic value-adding operations are grouped into a series of categories. Inputs, transformation processes, and outputs are all part of a value chain. It may be used as an analytical technique to comprehend the value created by data technology. A *data value chain* is a sequence of processes necessary to produce value from data. An information system may be modeled using the big data value chain, as shown in Figure 3.

Figure 3: *The Data Value Chain (Curry, 2016).*

Data Acquisition

Data Acquisition is the procedure for acquiring, filtering, and cleaning data before storing them in a data base that can be used for data analysis. In terms of infrastructure needs, data gathering is one of the most significant big data issues. The infrastructure necessary to enable big data acquisition must provide low, predictability of lag in data capture and query execution, and be capable of handling extremely high transaction volumes frequently in a short period (Curry, 2016).

Data Analysis
The goal of data analysis is to make the raw data gathered usable for decision-making and domain-specific applications. A data analyst is someone responsible for examining, manipulating, and modeling data to identify relevant data and synthesize and extract valuable, concealed information with great potential from a business point of view. Data mining, business intelligence, and machine learning are other related fields (Curry, 2016).

Data Curation
Data curation is the active management of data throughout its life cycle to ensure that it satisfies the required data quality standards for successful use. Content generation, selection, classification, transformation, validation, and preservation are some of the tasks involved in data curation. Expert curators oversee enhancing the accessibility and quality of data through data curation. Data curators (also known as scientific curators or data annotators) make sure that data are reliable, discoverable, accessible, reusable, and appropriate for their intended use (Curry, 2016).

Data Storage
Data storage is the scalable persistence and administration of data that meets the demands of applications that need quick access to the data. Relational Database Management Systems have been the primary solution for information storage. The Atomicity, Consistency, Isolation, and Durability characteristics that ensure transactions in the database, on the other hand, lack flexibility in terms of schema modification, performance, and fault tolerance as data quantities and complexity increase, causing them to be inappropriate for big data situations. NoSQL databases were designed for greater scalability, and they provide a diverse set of alternatives based on different data models (Curry, 2016).

Data Usage
Data use refers to data-driven business operations that need data access, analysis, and tools to incorporate data analysis into the business activity. The use of data in corporate decision-making may improve competitiveness by lowering costs, increasing added value, or measuring any other parameter against current performance standards (Curry, 2016).

2.2 Data Quality

The English Oxford Living Dictionary defines *quality* as ”the standard of something as measured against other things of a similar kind; the degree of excellence of something” or, alternatively, ”a distinctive attribute or characteristic possessed by someone or something” (Lexico, 2021).

Quality is essential, but there is little consensus on what it truly is. The following seven definitions show how quality is measured from the perspectives of management, quality assurance, product, marketing, production, and economics (Mar, 2013):

1. **Fit for purpose**
   This definition arose from discussions in the quality management community. It is beneficial since it can be used to describe any process, service, or product. However, it is difficult to quantify.

2. **Conformance to requirements**
   Business users, for example, establish the need for a sales system. The sales system is created, and its quality is assessed against the set of criteria.

3. **Proportional to costs**
   Cost equals quality. Producing a high-quality product is more expensive than producing a product with low quality. Product quality has traditionally been measured in terms of material costs. For certain basic goods, this kind of quality definition works well. However, it does not apply to technology, art, or culture.

4. **Quality is price**
   The amount customers are willing to pay for a product or service is referred to as quality. Economists believe that if something is costly, it must be of good quality.

5. **Quality is a standard**
   The manufacturing sector was the first to take a serious, scientific look at quality. It is critical to ensure that both the quality of the goods and the production process are high. Manufactures cannot afford to create sub-par goods that their consumers will return. They cannot afford product liability problems as a consequence of poor quality. Inefficient procedures are also expensive. To enhance both procedures and product quality, manufacturers utilize standards and continuous process improvement methods. They think of quality in terms of measurements and statistics.

6. **Quality is value for performance**
   A $3 disposable toothbrush, according to this concept, may be of better quality than a $3,000 golden toothbrush since it provides better value.

7. **Quality is an experience**
   Customers' experiences are assessed via the development of connections with them to generate dialogue and feedback. Bottom-line indicators, such as revenue, repeat visits, and lifetime customer value may also be used to assess experience quality.

There are several definitions of what *data quality* is, but usually, it is considered to have high quality if it is "fit for (its) intended use in operations, decision making, and planning" (Redman, 2008). As a result, data quality is evaluated in terms of how well it fulfills the expectations and requirements of data consumers. Therefore, the quality of data is said to be constrained by context and the requirements of the data consumers. One of the most challenging elements of data quality management is that quality requirements are not always apparent. Data management experts must first understand their clients' quality expectations and how to assess them to manage data quality successfully. These specifications may change over time, requiring ongoing discussion (DAMA, 2017; Mestl & Ruud, 1999).

2.3 Data Quality Dimensions

2.3.1 ISO 8000

A common framework to use is the ISO 8000 standards for data quality measurements. In this framework, the following three categories are mentioned : syntactic, semantic, and pragmatic (DNV, 2017; ISO Central Secretary, 2015; Lekanger, 2019):

1. **Syntactic Quality**
   The syntactic quality category describes the degree to which stored data conforms to stored metadata. Legal values, data kinds, referential integrity, such as connections between data components, business language, and any specified business rules are examples of metadata in this context. This category deals with the problem of data quality concerning design (as represented by metadata), for example, as determined by integrity checking (Price et al., 2008).

   The structure of data is referred to as syntactic data quality. The objective of syntactic data quality is consistency, which is achieved by using a consistent symbolic representation for data values for specific data items in the data warehouse (Balou & Tayi, 1996; Wang et al., 1995).

2. **Semantic Quality**
   Semantic data quality aims to ensure that the data one has access to has a consistent, meaningful representation. For example, if a car registers a velocity of 80 $\frac{km}{h}$, then the car's actual velocity should likewise be 80 $\frac{km}{h}$, and if it is not, there is some semantic error. Comprehensiveness and accuracy are the benchmarks for semantic quality (Balou & Tayi, 1996; Wang et al., 1995). The degree to which there is a data value in the data warehouse for each relevant state in the real-world system is referred to as comprehensiveness. The degree to which data values in the data warehouse match the condition of the actual world is referred to as accuracy (Shanks & Darke, 1998).

3. **Pragmatic Quality**
   The degree to which data is appropriate and helpful for a specific purpose is referred to as pragmatic data quality. It supports data users' beliefs about their data's suitability for their intended use. Usability and usefulness are the objectives of pragmatic quality (Kahn et al., 1997). The degree to which each stakeholder can successfully access and utilize the data warehouse is known as usability. The degree to which the data assist

the stakeholder in completing their job within the social environment of the company is defined as usefulness (Shanks & Darke, 1998).

2.3.1.1 The ISO/IEC 25012 Data Quality Model

The data quality model is the foundation upon which the system for evaluating data product quality is constructed. The major data quality characteristics that must be considered when evaluating the qualities of the desired data product are defined in a data quality model. There are two major types of data quality characteristics: inherent data quality and system-dependent data quality. The term ”inherent data quality” refers to the extent to which data quality features have the inherent ability to fulfill expressed and inferred requirements when utilized under defined circumstances. The term ”system dependent data quality” refers to the extent to which data quality is achieved and maintained inside a computer system when specific criteria are met. The ISO/IEC 25012 Data Quality Model consist of 15 different qualities and is as follows (ISO Central Secretary, 2008):

Inherent Data Quality:

1. **Accuracy**: The degree to which data has attributes that correctly represent the true value of the intended attribute of a concept or event in a specific context of use.

   It has two main aspects:

   - **Syntactic Accuracy**: Syntactic accuracy is defined as the closeness of the data values to a set of values defined in a domain considered syntactically correct.
   - **Semantic Accuracy**: Syntactic accuracy is defined as the closeness of the data values to a set of values defined in a domain considered syntactically correct.

2. **Completeness**: The degree to which subject data associated with an entity has values for all expected attributes and related entity instances in a specific context of use
3. **Consistency**: The degree to which data contain characteristics that are not contradictory and are consistent with other data in a given context. It may be among data on a single entity or across similar data for comparable entities, or both.
4. **Credibility**: The degree to which data has attributes that are free from contradiction and are coherent with other data in a specific context of use. It can be either or both among data regarding one entity and across similar data for comparable entities
5. **Currentness**: The degree to which data has attributes that are of the right age in a specific context of use.

Inherent Data Quality:

According to ISO Central Secretary (2008) inherent data quality is defined as the following dimensions:

1. **Accessibility**: The degree to which data can be accessed in a specific context of use, particularly by people who need supporting technology or special configuration because of some disability.

2. **Compliance**: The degree to which data has attributes that adhere to standards, conventions or regulations in force and similar rules relating to data quality in a specific context of use .

3. **Confidentiality**: The degree to which data has attributes that ensure that it is only accessible and interpretable by authorized users in a specific context of use. Confidentiality is an aspect of information security (together with availability, integrity).

4. **Efficiency**: The degree to which data has attributes that can be processed and provide the expected levels of performance by using the appropriate amounts and types of resources in a specific context of use

5. **Precision**: The degree to which data has attributes that are exact or that provide discrimination in a specific context of use

6. **Traceability**: The degree to which data has attributes that provide an audit trail of access to the data and of any changes made to the data in a specific context of use

7. **Understandability**: The degree to which data has attributes that enable it to be read and interpreted by users, and are expressed in appropriate languages, symbols and units in a specific context of use. Some information about data understandability are provided by metadata

System-Dependent Data Quality: According to ISO Central Secretary (2008) system-dependent data quality is defined as the following dimensions:

1. **Availability**: The degree to which data contain characteristics that allow authorized users or programs to obtain it in a particular context of usage.

2. **Portability**: The degree to which data contain characteristics that allow them to be installed, changed, or transferred from one system to another while maintaining its quality in a given context.

3. **Recoverability**: The degree to which data contain characteristics that allow them to retain and keep a defined level of operations and quality in a particular context of usage, even if it fails.

2.3.1.2 Quality Dimensions According to DAMA (2017)

Data quality criteria may be defined using the data quality dimensions. Businesses must define criteria essential for business operations and are quantifiable to assess the quality of data. It is feasible to create quantifiable rules using these dimensions. There are many paradigms for the meaning of the term *data quality.* However, according to (DAMA, 2017), the following dimensions can be defined as:

1. **Completeness**: The proportion of data stored against the potential of 100%.
2. **Uniqueness**: No entity instance (thing) will be recorded more than once based upon how that thing is identified.
3. **Timeliness**: The degree to which data represent reality from the required point of time.
4. **Validity**: Data is valid if it conforms to the syntax (format, type, range) of its definition.
5. **Accuracy**: The degree to which data accurately describes the 'real world' object or event being described.
6. **Consistency**: The absence of difference, when comparing two or more representations of a thing against a definition.

Other factors that affect quality were also highlighted by DAMA (2017). They are not given as dimensions, but they are as essential to comprehend and defined by DAMA (2017) as:

1. **Usability**: Is the data understandable, simple, relevant, accessible, maintainable and at the right level of precision?
2. **Timing issues** Beyond timeliness itsels): Is it stable yet responsive to legitimate change requests?
3. **Flexibility**: Is the data comparable with other data? Does it have useful groupings and classifications? Can it be repurposed? Is it easy to manipulate?
4. **Confidence**: Are Data Governance, Data Protection, and Data Security in place? What is the reputation of the data, and is it verified or verifiable?
5. **Value**: Is there a good cost / benefit case for the data? Is it being optimally used? Does it endanger people's safety or privacy, or legal responsibilities of the enterprise? Does it support or contradict the corporate image or the corporate message?

2.3.2 Data Quality and Customers

Gupta (2021) listed the same data quality dimensions as DAMA (2017), except timeliness. On the other hand, integrity has taken the place of timeliness. The dimension controls whether identical data are saved and utilized in many instances. This is represented as a proportion of values that match across several objects. Data consistency guarantees that the value of data is captured and used properly in analysis.

2.3.3 Data Quality Assessment

DNV (2017) contributes to how data quality assessment should be done in the workplace. The paper emphasizes the importance of defining the scope of the data, which entails specifying the data's intended uses before investigating and profiling them for those purposes. They explain how to profile data and how to assess data quality metrics, organizational maturity, and data quality risk. Figure 4 depicts a high-level summary of the procedure.

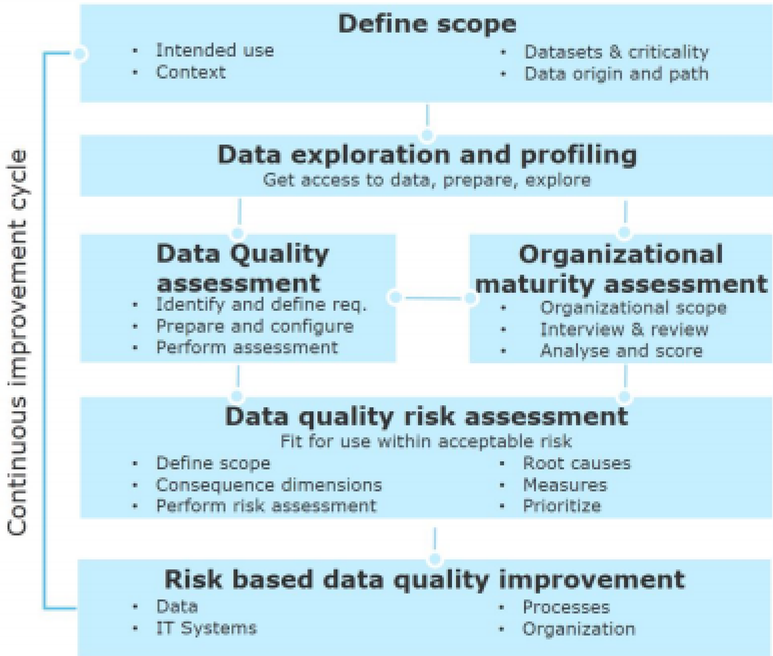


Figure 4: *Data Quality Assessment (DNV, 2017).*

DQI (The data quality index) is a normalized number that expresses data quality as a decimal number between 0 and 1, with 0 representing the most inadequate quality and 1 representing the best. *DQI* is defined as follows: *Rt* is the total number of records examined, and *Rp* is the number of records that passed the test (DNV, 2017).

$$DQI = \frac{R_p}{R_t}. \tag{1}$$

The quality dimensions ISO Central Secretary (2008) and DAMA (2017) outlined in Sections 2.3.1 and 2.3.1.1 are used to calculate this quality index. This way of measuring data quality is also supported by the theory of IBM (2016), which also describes the confidence of each data quality dimension as a value between 0 and 1.

Maturity Assessment

Organizational maturity assessment is done to respond to data quality issues in a predictable and repeatable way. A paradigm for evaluating organizational elements of data quality is provided, with five degrees of maturity. Data analysis, document inspection, and critical people interviews all help to assess the data's maturity level (DNV, 2017).

Risk Assessment

Data quality risk assessment is carried out to determine the potential business impact and should be thoroughly examined. Risk analysis is used to evaluate and prioritize mitigation activities based on the risk score and risk tolerance. The risk analysis should be used as a starting point for risk-based data quality improvement, with the goals of improving digital process performance, gaining business benefits, lowering costs, and extracting opportunities. (DNV, 2017).

2.3.4 Data Security

Data security is an essential aspect of data quality and data maturity. Data ownership and governance are also required to offer the correct degrees of access control to the data. Ad hoc actions can risk integrity, and analysis can reveal personal data (DAMA, 2017; DNV, 2017). Confidentiality, integrity, and availability are three words that are commonly used to define information security. Other attributes that may be engaged in data security are authenticity, responsibility, non-repudiation, and reliability, according to ISO/IEC 27001 (Talha et al., 2019).

However, there may be conflicts between data quality and data security. Flexible read and write access to all data can be required when developing a data quality management system. Because the data quality system can interchange data with other systems or be changed by different users with different profiles who do not necessarily have the same access permissions, this requirement can lead to many security issues. As a result, data security might be a barrier to data quality needing new solutions to be considered (Talha et al., 2019).

2.4 Economic principles

2.4.1 Willingness to Pay

Willingness to pay is the maximum price at or below which a consumer will buy a unit of a product (Varian, 1992). Companies must be aware of the factors that influence the price they may accept for a product. The business will be unable to follow a price plan that is appropriate for its market situation without this information about the customer's willingness to pay (Breidert et al., 2006). The literature examines a variety of methods for determining willingness to pay, including market data analysis, performing experiments (laboratory or field trials), and conducting direct surveys.

2.4.2 Supply and Demand

One of the most basic principles in economics is the principle of *supply and demand*. Adam Smith popularized them in 1776, and they have been proven to anticipate market behavior accurately. The principle comprises a supply curve and a demand curve that show how demand and supply vary when the price of a product or service rises or falls. Because a producer wants to sell their goods for the highest possible price and a consumer wants to purchase them for the lowest possible price, an equilibrium point must be reached for both the consumer and the producer to profit from continuing economic transactions. The place where the two curves meet is called the equilibrium point. When one of the two curves changes, the equilibrium point may shift as well, as seen in Figure 5 (Investopedia, 2020). The fundamental supply curve operates a little differently for digital goods such as information, which have the unique capacity to be reproducible and recyclable. The supply seems limitless if the same unit of information can be sold and supplied to numerous consumers. Laney (2017) gives two formulas for describing the behavior between suppliers and consumers of information assets. The cost value of information (CVI) establishes a minimal price for information based on the costs of producing/collecting, maintaining, and distributing it, whereas the market value of information (MVI) formula explains how market saturation devalues the same information, lowering demand.

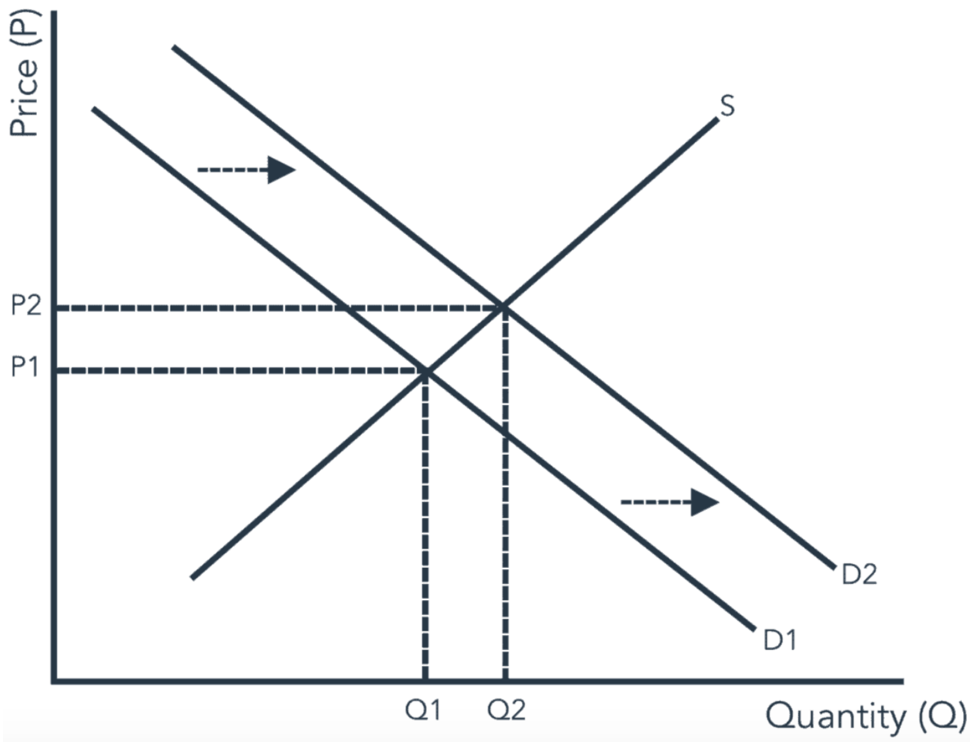


Figure 5: *Supply and Demand (Agarwal, 2018)*

A ccording to Kenton (n.d.) ”the demand curve is a graphical representation of the relationship between the price of a good or service and the quantity demanded for a given period of time.”

The law of demand states that the demand curve will go down from left to right because quantity of a good purchased is inversely proportional to the price of that good (Kenton, n.d.). For example, as the price increases, the quantity demanded decreases. (Hayes, n.d.). As the price of a commodity rises, the quantity demanded falls, all other factors being constant (Kenton, n.d.).

The willingness to pay and the demand curve are inextricably linked. The price provided by the demand curve determines the willingness to pay since the demand curve tells how much quantity of the good can be sold at a particular price point. The condition of the economy, trends, the customers' personal price points, the rarity of a product, and the quality of a product, to mention a few, may all have an impact on willingness to pay. The factors that influence a person's willingness to pay for a product are crucial to comprehend when determining a product's price (Campbell, 2021).

Relating the economic principles on data is an essential task for a data-driven business. Laney (2017) states that the traditional supply and demand is not suited for information assets and that the price equilibrium is based on a more sophisticated function of information cost, possible use, and market saturation. He states that information suppliers should consider how attenuating market saturation will affect the number of buyers and that consumers should not expect to buy information at a nominal cost premium. Therefore achieving an information price equilibrium would involve some game theory beyond basic demand and supply. Laney (2017) also discusses the pricing of information assets as the problem with the demand and supply graph arises. He says that since information, in theory, is infinitely replicable and the same information can be sold several times, suppliers do not need to limit the supply for any price point. This effect may push the price for the information to zero resulting in no or negative profits. Limiting the supply would therefore be crucial for the company profiting from the information. Therefore, it may make more sense for a supplier of information to increase supply up to the point in which market saturation creates downward price pressure (Laney, 2017).

2.4.3 Price Elasticity of Demand

Price elasticity is a concept of a mathematical calculation that is used to find the relationship between the price and the demand for a specific product (Visma, n.d.). For example, Figure 6 shows the price elasticity of demand which is defined as the percentage change in demanded quantity (ΔQ measured as a percentage) divided by the percentage change in price (ΔP measured as a percentage.

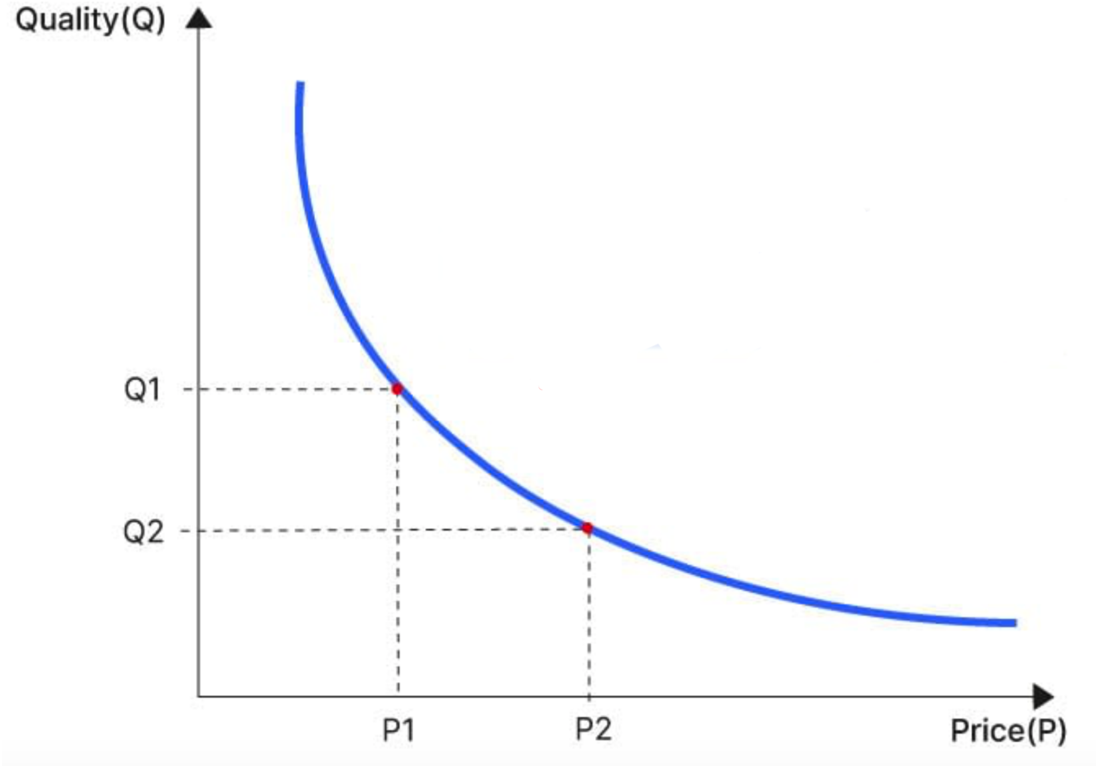


Figure 6: *Price Elasticity of Demand (Hoffmann et al., 2020)*

As Figure 6 shows , price elasticities are almost always negative. This means that the demand for a good decreases when the price increases and vice versa.

Regarding the determination of the reasonable price for a product or service, price elasticity is often utilized. That is, one can determine how much a change in price affects demand. Therefore, price elasticity is part of what is known as "demand-based pricing" (Visma, n.d.).

According to Ross (2021), the responsiveness of a product or service's supply to a change in its market price is measured by the price elasticity of supply (PES). When the price of a product rises, the supply of that good increases, according to the fundamental economic theory. When the price of a product falls, the supply of that good falls as well. The formula is explained as below:

$$PES = \frac{\%\text{ Change in Supply}}{\%\text{ Change in Price}} \tag{2}$$

There are five kinds of PES: perfectly and relatively inelastic, unit elastic and perfectly and relatively elastic. Below is an example of each of the five supply curve pricing elasticities (Ross, 2021):

- Perfectly Inelastic Supply - When the PES formula equals 0, there is a perfectly inelastic supply. When the price changes, the amount provided does not change. Products in limited numbers, such as land or paintings by deceased artists, are examples.

- Relatively Inelastic Supply - For relatively inelastic supply, the PES ranges from 0 to 1. This implies that the percentage change in quantity provided is smaller than the percentage change in price. Nuclear power, for example, has a longer lead time due to plant construction, technical know-how, and the protracted ramp-up procedure.
- Unit Elastic Supply - The PES of unit elastic supply is 1, which means that the amount provided changes by the same proportion as the price.
- Relatively Elastic Supply - When the price elasticity of supply is higher than 1, it implies that the amount provided varies by a more considerable proportion than the price change. A fidget spinner is an example of a product that is simple to manufacture and market. The resources to manufacture more spinners are easily accessible, and the overall cost of ramping up or down manufacturing would be low.
- Perfectly Elastic Supply - The PES for perfectly elastic supply is infinite, which means that the amount can be supplied at a given price is endless, but no quantity can be available at a different price. There are few real-world examples of this where even a minor price differential would deter or prevent product makers from offering even a single product.

2.4.4 Return on Investment

ROI is a metric that compares an investment's profit or loss to its cost. Even though ROI is a ratio, it is most often represented as a percentage. It is often used to assess the likelihood of earning a profit from an investment (Fernando & Mansa, 2020).

The ROI is as follows:

$$\textbf{ROI} = \frac{\text{Current Value of Investment} - \text{Cost of Investment}}{\text{Cost of Investment}} \tag{3}$$

The profits from the sale of an interest-bearing investment are referred to as the ”current value of investment.” Because ROI is expressed as a percentage, it can readily be compared to the returns on other assets, enabling one to evaluate a range of investment kinds.

$$\textbf{ROI} = \text{Gain From Investment} - \text{Cost of Investment} \tag{4}$$

Gain From Investment
The gain from the investment will contain revenue optimization, cost optimization, and existing solution gain (Zola, n.d.).

1. **Revenue Optimization**
   To calculate revenue optimization, which refers to how much revenue one can obtain from spending a certain amount of money on big data, one must first calculate:

- **Existing Revenue Stream**
  One can generate bundles to enhance the existing revenue streams or generate innovative price plans, and this percentage can be added to the annual revenue as an additional gain from the investment (Zola, n.d.).

- **New Revenue Stream**
  If the data can be offered as a new service, the revenue information from the selling budget may be used to calculate the return on investment (Zola, n.d.).

2. **Cost Optimization**

   - **Human Capital Gain**
     Taking advantage of using big data analytics is significant for productivity. To measure how much productivity has changed, the business must know the average cost per employee per hour and then determine the hours of increased productivity (Zola, n.d.).

     Human Capital Gain = Employee Average Cost Per Hour - Total Number of Increased Productivity Hour

3. **Existing Solution Gain**

   - **Phasing out**
     If the business is going to phase out the data's improvement, it should first estimate the cost of having it at the same level. This enables them to measure the cost savings when they get the new data solution (Zola, n.d.).

     Existing Solution Gain = Running Costs + Value of Minimized Revenue Risk

Cost of Investment
The cost of investment that is related to investing in big data analytics is CAPEX and OPEX (Zola, n.d.).

1. **CAPEX**
   CAPEX stands for capital expenditures that are not recurrent. Finding the ideal solution to meet the company's requirements will take time, and it will also need a lengthy roll-out period before it is ready to use. Even though the cost only applies once, it is still necessary to include it in the ROI (Zola, n.d.).

2. **OPEX**
   OPEX means ongoing cost, «which stands for operating expenses or expenditure and refers to the costs incurred by the business via the production of goods and services» (GoCardless, n.d.).

The cost of these investments for data according to Zola (n.d.) can be related to client support, subscriptions, network and security, licenses, software upgrades, customization, and data storage.

2.4.5 Price and Quality Relationship

Instead of the relationship between price and quantity given by the supply and demand curves, the *price and quality relationship* describes the price and thereby measures the customers' willingness to pay. Several studies have investigated this relationship. However, there have been no studies (to our knowledge) on the price and quality relationship for data assets. According to Steenkamp (1988), a general positive correlation was found between the two variables, but most correlations were weak, indicating that there may not be a perfectly linear relationship.

2.5 Uncertainty

The gap between the knowledge required to make a safe choice and the information available at the moment the decision is made is known as *uncertainty* (Prosjekt Norge, n.d.). There are several ways to categorize uncertainty. According to Prosjekt Norge (n.d.), the most common way is to divide it into contextual uncertainty and operational uncertainty.

- **Contextual Uncertainty -** Uncertainty related to the project's surroundings, nature, and basic conditions. These are completely or to a large extent outside the project's control.
- **Operational Uncertainty -** Uncertainty related to the actual implementation of the project and the factors over which the project has much control.

Austeng et al. (2005) also introduces the following categories:

- **Conceptual Uncertainty -** Uncertainty related to the analysis itself and the interpretation of the results from the analysis.
- **Scenario Uncertainty -** Uncertainty related to the stability of the goals or decision criteria in the project. All of these conditions can be radically changed in projects with a long-term horizon.

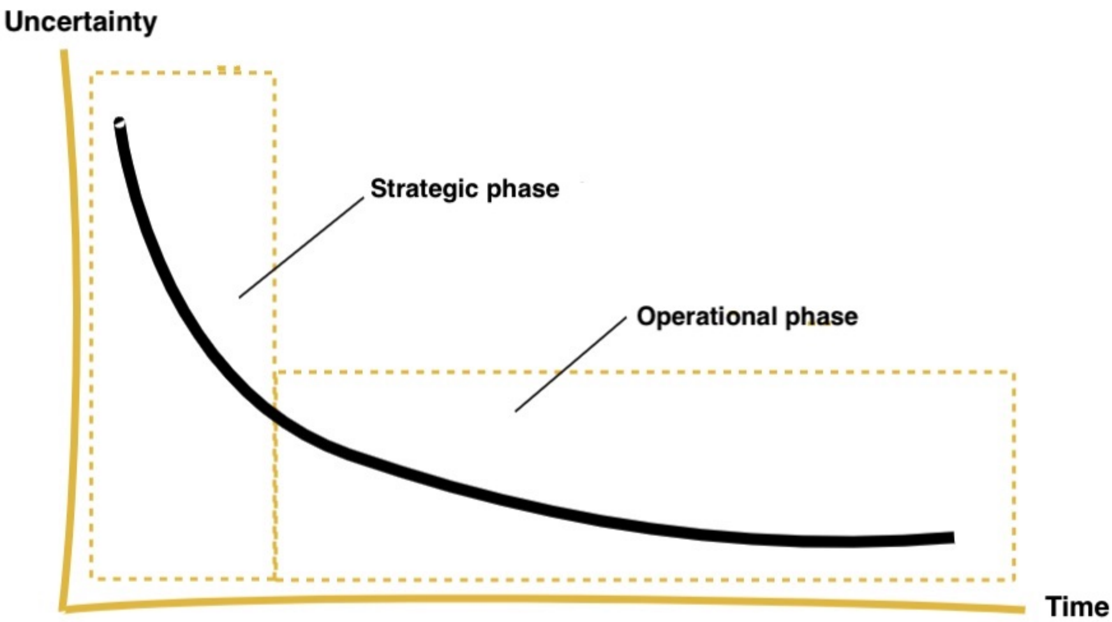


Figure 7: *Strategic Phase and Operational Phase in Uncertainty Over Time (Husby, 1999).*

According to Husby (1999), there is a common agreement that the level of uncertainty in the strategic phase would be higher than in the operational phase of the project, as illustrated by Figure 7. The most frequent reason is that a limited quantity of useful information is initially accessible, but it progressively grows.

Uncertainty has both positive and negative aspects. Negative uncertainty is linked to negative outcomes, while positive ambiguity creates possibilities that may be taken advantage of. If the consequences of uncertainty occurring are considered, risks or opportunities occur. Therefore, negative uncertainty is risks, and positive uncertainty is opportunities (Rolstadås, 2020).

2.5.1 Risks

Risk is the probability that an unwanted event will occur multiplied by the consequences it creates (Rolstadås, 2020).

$$\textbf{Risk} = \text{Likelihood} \cdot \text{Consequence} \tag{5}$$

According to (Shen, n.d.), in financial terms, risk refers to the possibility that a result or investment's real profits may vary from what was anticipated. The potential of losing part or all of one's initial investment is a risk.

Possible positive effects are incorporated in the risk concept in the newest ISO standards, as well as, a variety of different situations. The term "risk" refers to the uncertainty of achieving an objective in general. Threats and opportunities may influence uncertainty, resulting in both bad and good outcomes depending on the goal (Digitaliseringsdirektoratet, 2021).

According to Melanie (2019), the five consequences of poor-quality data are:

1. **Poor decision-making** Bad choices will be made because of low quality. This implies that no choice can be better than the one on which it is based. Furthermore, making important choices based on low-quality data might have fatal consequences.

2. **Business inefficiencies** Poor data quality will lead to inefficient corporate operations that rely on data, from reports to product orders and everything in between, where facts are needed. Instead of concentrating on the essential activities, this may lead to costly rework labor, which must verify and rectify data problems.

3. **Mistrust** Much mistrust can arise because of the low data quality. This is particularly true in sectors where laws govern customer interactions or commerce, such as the financial services industry. Money, as well as effort and reputation, maybe wasted if the data are incorrect. This has a detrimental impact on the company and decreases the costumers' trust.

4. **Missed opportunities** With poor data quality, the risk of losing out on profitable opportunities is high. For example, a competitor with a higher understanding of data can profit from customer needs or new product development, which the business cannot compete with due to low data quality.

5. **Lost revenue** Poor data quality may result in increased expenses and income loss. For example, inaccurate customer data, for example, may result in poor sales. This may lead to erroneous targeting and communication, which is particularly bad for multi-channel sales.

On the other hand, the quality risk is the potential for losses due to quality that fails to meet the business' quality goals (Spacey, 2015).

2.5.1.1 Costs of Poor Data Quality

Cost of poor quality (COPQ) is an unknown term. It is a monetary indicator that measures how poor quality across all business operations has a negative effect on a company's profitability (Kudva, 2020).

Organizations have acquired and retained massive volumes of data because of the advancement of information technology in recent decades. However, as the volume of data grows, so does the difficulty of handling it. Because companies are collecting and managing more significant and more complex information resources today, the danger of poor data quality is increasing. Poor data quality costs many companies much money, although it is difficult to estimate the exact amount. It is difficult to estimate the total cost of bad data (Haug et al., 2011).

Haug et al. (2011) proposes a definition of the optimal data maintenance effort shown in Figure 8. The costs of coping with low-quality data are represented on the vertical axis. The second and horizontal axis is concerned with data quality. The two curves in the diagram reflect the expenses of low data quality and the costs of maintaining good data quality. The combined cost of the two stated curves is the entire cost of data quality. The figure has two assumptions: 1. The costs incurred due to of poor data will drop exponentially, and 2. The expenses of ensuring data quality are proportional to the quality of the data. Figure 8 shows that the relationship between the costs of low data quality and the costs of maintaining good data quality can be classified as a trade-off. The primary premise is that thorough data cleansing, which ensures excellent data quality, eventually becomes less profitable.

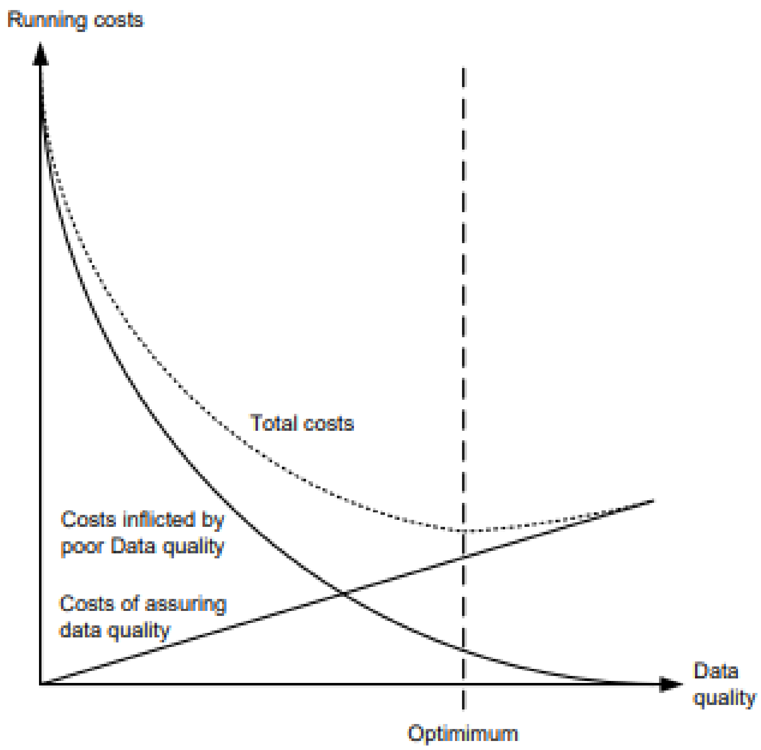


Figure 8: *The Optimal Data Maintenance Effort (Haug et al., 2011)*

2.5.2 Opportunities

Opportunity is the probability that a desired event occurs multiplied by the consequences it gives (Rolstadås, 2020).

$$\textbf{Opportunity} = \text{Desired event} \cdot \text{Consequence} \tag{6}$$

According to Grier (2019), there are three ways one can turn uncertainty into opportunities:

1. **Be agile and build resilience:**
   The first step is to build a flexible organization. Many businesses, for example, review

their portfolios to ensure they have a strategy for navigating a volatile and changing environment. Companies should focus on their core competencies while reinvesting in their transformation agenda. Organizations that do so will have an easier time swinging. They will be better able to fund new technology investments and safeguard their future in the transformative era.

2. **Prioritize your people:**
   When trying to improve a company, the most important resource is its employees. Unfortunately, changes in leadership, elections, nationalist closures, geopolitical shocks and personal liberties might threaten an organization. As a result of this change, organizations must ensure that their employees feel safe, supported, and valued.

3. **Act decisively:**
   When things change due to a tweet affecting markets or a leader making a minor mistake with deadly repercussions, it is critical to have access to the correct information from a variety of reliable sources. Although many of the sources will be found inside the organization, this will not be sufficient. Third-party experts are equally essential. This is because specialists can help discover blind spots, question assumptions, and better understand what the organization does not know. This knowledge may aid in understanding and evaluating important events, allowing organizations to make the best choice based on what is known.

2.6 Investment in Data and Data Quality

Regarding data investment, understanding how to value the asset is critical. Valuation comes easily to certain assets and is simple to determine. For example, the value of a stock is the difference between what it costs to buy and what it is sold for. However, it becomes complex when it comes to data since ”neither the costs nor the benefits of data are standardized” (DAMA, 2017). In addition, data has a contextual value, implying that information may be valuable to one company but not to another. As a result, determining the value of data is difficult. It is essential to develop methods for associating financial worth with data because companies must comprehend assets in financial terms to make consistent decisions (DAMA, 2017).

2.6.1 Data Value Drivers

The degree to which data may generate future economic advantages for an organization is determined by essential data features - data value drivers, such as those shown in Figure 9. Some of these drivers determine the data's quality (completeness, consistency, accuracy, timeliness), while others may either make the data useless or provide the owner or rights holder with unique and essential competitive advantages (exclusivity, restriction, liability, interoperability) (Rea & Sutton, 2019). The data value drivers are defined by Rea and Sutton

(2019) as:

1. **Exclusivity -** having exclusive rights to this data set would be far more valuable than merely being one of multiple licence holders.
2. **Timeliness -** data which describes an individual's location at a precise moment in time is significantly more valuable to a retailer than data which describes their location an hour ago.
3. **Accuracy -** data may pin-point a person's location to the nearest 10m or only to the nearest 200m. Less accurate location data may result in adverts being served when an individual has already left the shopping centre.
4. **Completeness -** data which encompasses location and direction of travel will be far more valuable than data covering only one aspect – location or direction of travel.
5. **Consistency -** the personal identification data which links purchasing habits to an individual needs to be consistent with the personal identification data which provides that same individual's location, otherwise they will be being served with the wrong data.
6. **Use Restrictions -** local regulation may require an individual to 'opt-in' when it comes to pushing location based adverts to their mobile phone. Without this opt-in by the individual, the data will be worthless to a retailer.
7. **Interoperability -** the ability to combine real time location data with purchasing habit data creates significant value through the interoperability of the data sets.
8. **Liabilities and Risk -** the reputational consequences and financial penalties for breaching new data regulations, such as GDPR, can be severe. The greater the risk associated with the data use, the lower its value.

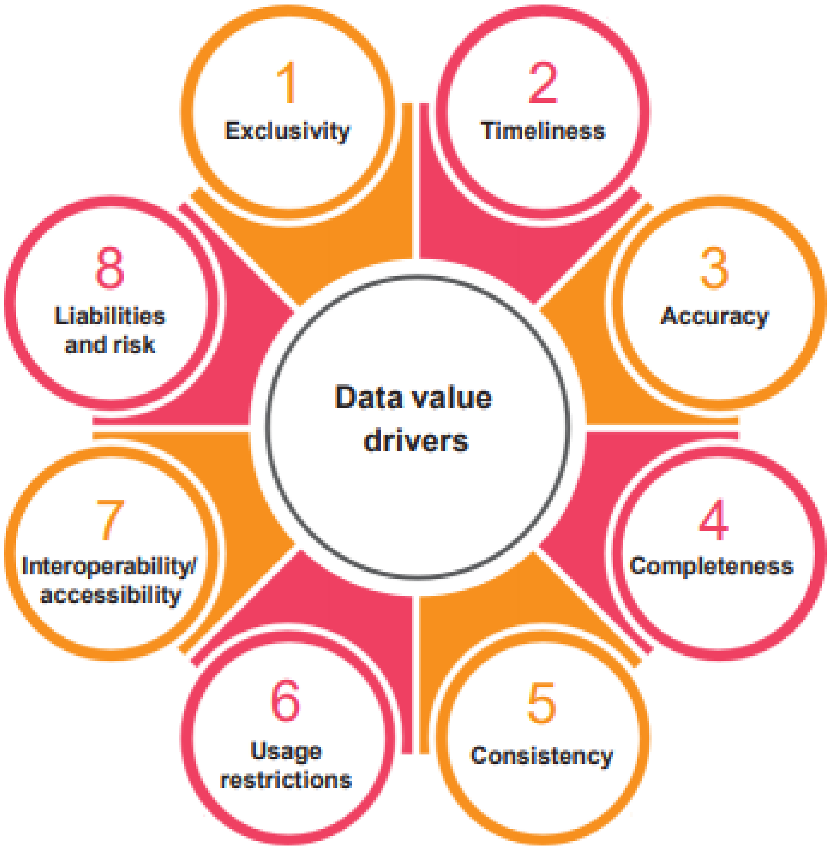


Figure 9: *The Data Value Drivers (Rea & Sutton, 2019).*

However, according to (White, 2020), organizations should consider the following question: "What makes data understandable or usable?" The other drivers aim to increase the usability of the data. If there is a problem, semantics and context must be provided for the model to be interoperable, which is challenging to do in such a simple model without complicating it.

Regarding interoperability, (White, 2020) believes it is no longer about technology. People have been able to connect systems using electronic interoperability standards for years, with electronic data interchange giving them this choice. The real challenge is semantic interoperability, which is made up of data and process (or use/context) semantics.

2.6.2 Maximizing Return for Big Data Projects

A study by Shim et al. (2015), discusses how to maximize the return on big data projects. They mention that organizations should consider four key areas of planning to increase ROI:

- **Skill -** To reduce the learning curve and achieve full knowledge of new technology, the company should include resources in every project to provide appropriate training for employees.
- **Measurement -** All big data projects should have well-defined project metrics and

performance indicators. The initiative will align with and enhance those metrics. Because it is difficult to quantify the advantages of a big data project, this measurement area is essential.

- **Technology -** All essential needs should be considered in big data projects, and staggered deployments of new technologies should be planned to gain expertise.
- **User profile -** Big data projects have an impact on a wide range of employees. During a big data project, each employee has a unique skillset and job description that should be prepared for. The project should keep track of the different kinds of users who engage with the platform and make sure it caters to their requirements.

According to Shim et al. (2015), big data projects must both match the needs of the organization and have metrics unique to the rapidly changing technology and changing customer landscape. According to Tata Consultancy Services (2013), some organizations have seen returns for big data projects as high as 10 times the investment. Figure 10 illustrates the costs and ROI.

Figure 10: *Big Data Return on Investment (Shim et al., 2015)*

Pitney Bowes Software (n.d.) highlighted attention to data quality as part of the data investment and gives the following recommendations for optimizing ROI:

- **Minimal investment:** Many of today's data-quality solutions may be implemented without costly investments in people, process, or technological expenditures. The most effective of them tap into data where it is stored. Rather than requiring system replacement, they interface with current legacy systems. Consequently, costs are kept low, but implementation time is increased.
- **Measurable results:** To make a case for program expansion, it is essential to choose initial initiatives that can yield precise, quantifiable results. This will assist in making the argument for the significance of improving data quality.

- **Powerful business impact:** The most compelling outcomes may be achieved by focusing on changes that have the most significant potential for a high ROI. Demonstrating significant ROIst attracts management attention and creates cross-organizational demand for more high-quality data programming.
- **Clear opportunity for program expansion:** Initial programming that sets the groundwork for enterprise-wide data quality improvement helps to pave the way for future development. It is critical to avoid "one-offs" and instead choose initial programming that can be readily adopted throughout the organization.

Laney (2017) presents a curve that depicts the expected return on data investments depending on how mature the organization is with data. He introduces the *information yield curve* for expressing the rate of improved value per unit of information-related investment. The information yield curve, as shown in Figure 11, shows how investments influence information asset maturity (IAM) in information. As they approach an optimization ceiling, low-maturity companies will experience accelerating gains in the rate of return on their information assets from information-related investments, whereas high-maturity organizations will see decelerating rates of return. A company may utilize a mix of IAM-scoring and information asset valuation and some comparable industry performance indicators to determine where it is on this curve. According to a study by Zhang et al. (2017) on big data investments in knowledge and non-knowledge intensive businesses, companies with varying degrees of knowledge intensiveness should proceed with care when making big data investments. According to the research, non-knowledge-intensive businesses have a higher chance of benefiting from these investments.

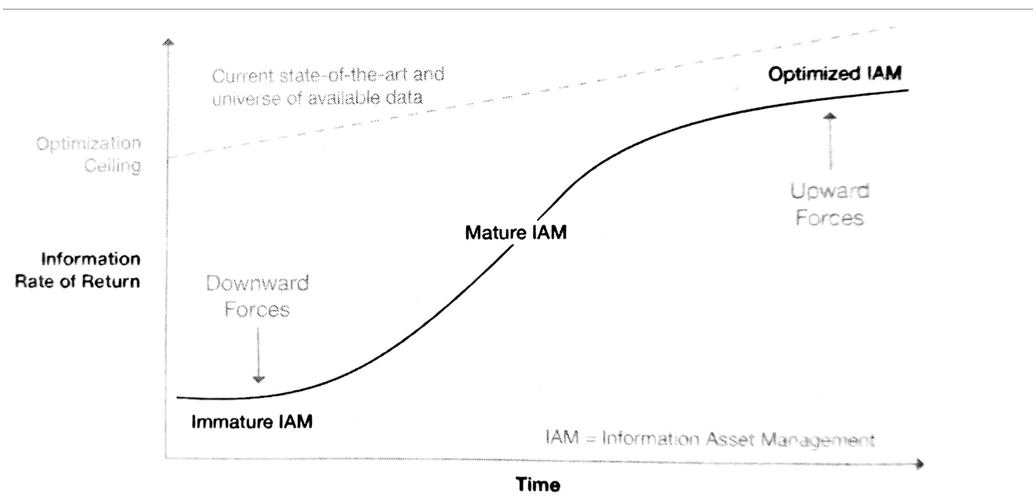


Figure 11: *Information Yield Curve (Laney, 2017).*

2.6.3 Putting Value on Data

The literature provides several models for evaluating data in investing strategies. This section will explain some of the most relevant models. Figure 12 presents an example of data valuation methodologies. It shows three approaches typically used to value any asset;income, market, and cost according to Rea and Sutton (2019)

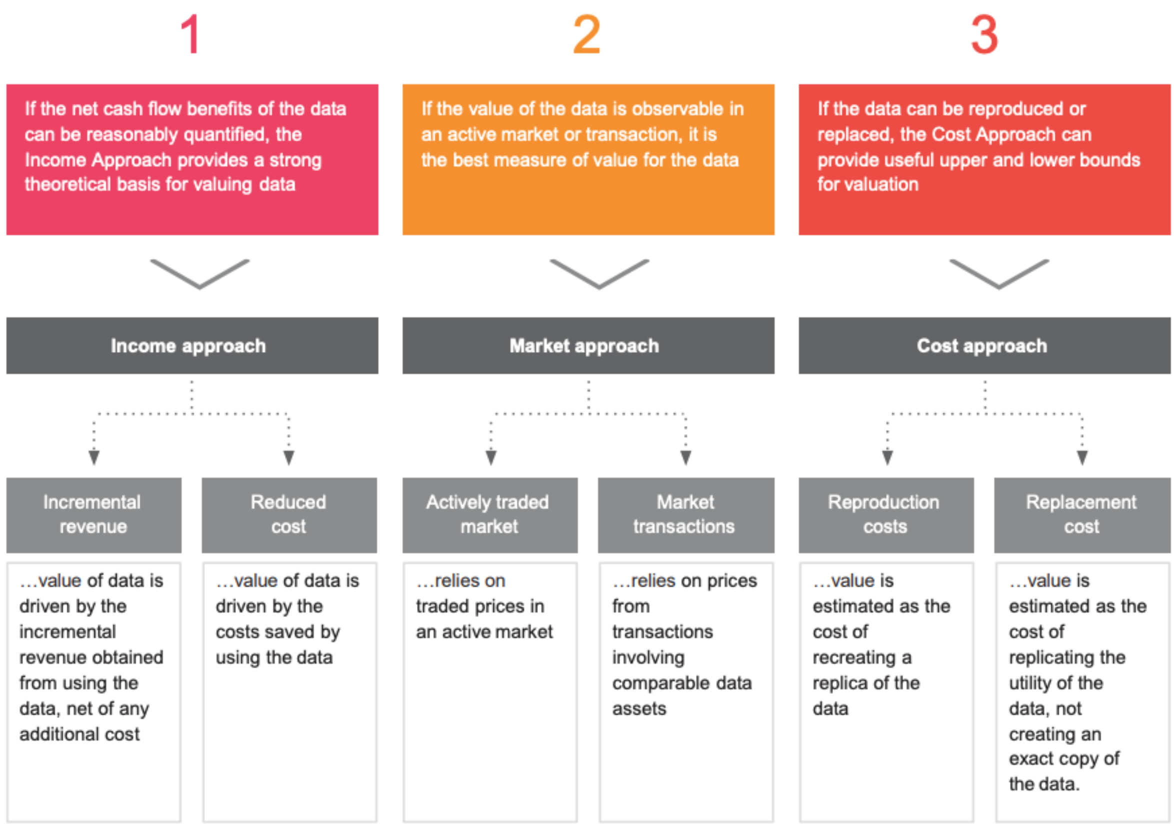


Figure 12: *Data Valuation Methodologies (Rea & Sutton, 2019).*

The income approach

The income approach evaluates the additional cash flows that the use cases are anticipated to produce in the form of increased revenue or reduced expenses by comparing the organization's cash flows with and without the data. Data can also help businesses save money by allowing them to better plan and optimize operations, as well as minimize and manage risk (Rea & Sutton, 2019).

The market approach

If relevant information is available, a market approach adds the most value. Today, active markets are uncommon. Those who do exist are either dealing with unlawful or illegitimate data or involved in cybercrime. It is still in its infancy and is not particularly liquid. As a result, obtaining such information is presently difficult and restricted for businesses (Rea & Sutton, 2019).

The cost approach

The cost method is more straightforward, but it often fails to account for future economic benefits. It may provide some useful benchmarks, but it is not suggested as a primary method (Rea & Sutton, 2019).

Laney (2017) describe different valuation models, both fundamental and financial, for information in an organization that has a connection to Rea and Sutton (2019) model. Regarding financial valuation models, he presents the CVI, which assesses an information asset as the financial expense required to generate, capture, or collect. It is given by

$$CVI = \frac{ProcExp \cdot Attrib \cdot T}{t} \{+ \sum_{p=0}^{n} \text{Lost Revenue}_p\} \tag{7}$$

Where:

- *ProcExp* = The annualized cost of process(es) involved in capturing the data.
- *Attrib* = The portion (percent) of *process expense* attributable to capturing the data.
- T = Average lifespan of any given instance of data.
- t = Time period over which the *process expense* is measured.
- n = The number of periods of time until the information is reacquired, or until business continuity is no longer affected by the lost or damaged information.

The next model is the MVI (Market Value of Information). In an open market, the MVI assesses the prospective or current financial worth of an information asset. Data monetization is usually done between commercial partners in exchange for money, products, or services, or other considerations like preferential contract terms and conditions. Unless the information is licensed or bartered, this approach is not relevant to most forms of data. This strategy should be considered when businesses grow more skilled and active in using their data outside. Because information is not sold but rather licensed, a consideration for reduced marketability of information will be required as it becomes more widely available in the marketplace. This factor is called *premium* in the following formula (Laney, 2017):

$$MVI = \frac{\text{Exclusive Price} \cdot \text{Number of Licensees}}{\text{Premium}}. \tag{8}$$

Finally, the EVI (Economic Value of Information) analyzes the realized revenue change when a particular information asset is integrated into one or more revenue-generating processes. It is given by (Laney, 2017):

$$EVI = [Revenue_i - Rvenue_c - (AcqExp + Admxp + AppExp)] \cdot \frac{T}{t}, \tag{9}$$

where:

- $Revenue_i$ = The revenue generated using the information asset (informed group).
- $Revenue_c$ = The revenue generated without the information asset (control group).
- T = The average expected life span of any given information instance or record.
- t = The period during which the EVI experiment or trial was executed.

2.6.4 Creating Value From Data

A growing number of companies are considering how they might enhance the value of their assets. As a result, it has been shown that companies should not begin with data, but rather with where and how they want to generate value in their company before moving on to data (PWC, 2019).

The value of information assets, according to Rea and Sutton (2019), has never been higher. Organizations that do their homework and take the time to understand the value of their data will unlock substantial value. If they do not, they risk wasting money on ineffective initiatives with little ROI. Figure 13 illustrates the cycle and six success criteria for development in the data landscape:

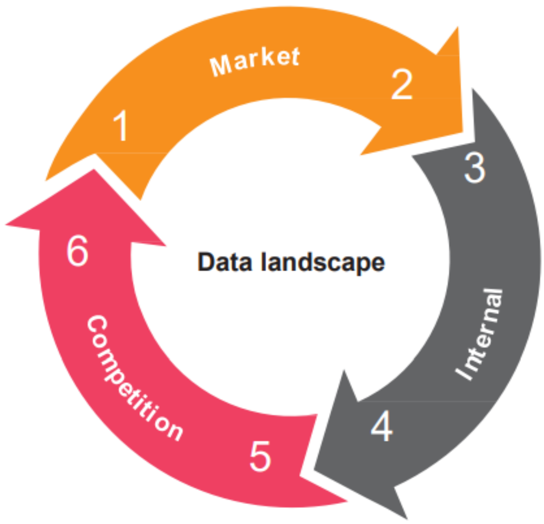


Figure 13: *The Data Landscape (Rea & Sutton, 2019)*

Market

1. Organizations are starting to understand the value of being viewed as data-centric (Rea & Sutton, 2019): Data-driven companies have substantially greater valuation multiples than other sectors. In addition, within the same sector, data-driven companies have greater valuations than their counterparts.

2. Organizations are getting requests from others to use/buy their data (Rea & Sutton, 2019): The use of data analytics to influence decision-making and enhance capabilities is becoming more common. As a result, dependence on third-party data sources is also increasing. In addition, the rapid pace of change necessitates fresh data sets and insights to keep up with changing circumstances (e.g., more sophisticated targeted marketing).

Internal

- Organisations are continually on the lookout for new revenue opportunities (Rea & Sutton, 2019): It has turned out that diversifying income sources and increasing margins are under pressure. Moreover, customer acquisition and retention techniques used in the past are becoming less successful, and many new business models rely on the monetization of data from IoT sensors.
- Organizations are finding that technology is improving, while costs are dropping (Rea & Sutton, 2019): Open source software for data storage and processing (such as Hadoop) is becoming more popular. Furthermore, the cost of cloud storage is just a few cents per giga bytes per month.

Competition

- Organizations are witnessing the success of others with strong analytics capabilities (Rea & Sutton, 2019): Companies with strong data analytics skills are twice as likely to be in the top quartile of their sectors' performance.
- Organizations have rightly started to fear big data players (Rea & Sutton, 2019): It has turned out to be a high concentration of data aggregators; fragmentation of suppliers and consumers. In addition, disruption from big data players is becoming more likely (e.g., Facebook, Uber, Amazon, Google).

2.6.5 Understanding the Data

According to Deloitte (2020), organizations must understand the value of their data assets. The value of their existing data assets and the underlying levels that may increase data value are often overlooked by companies looking to gain a competitive edge via their data assets. Figure 14, shows the steps that are recommended for an organization to avoid this:

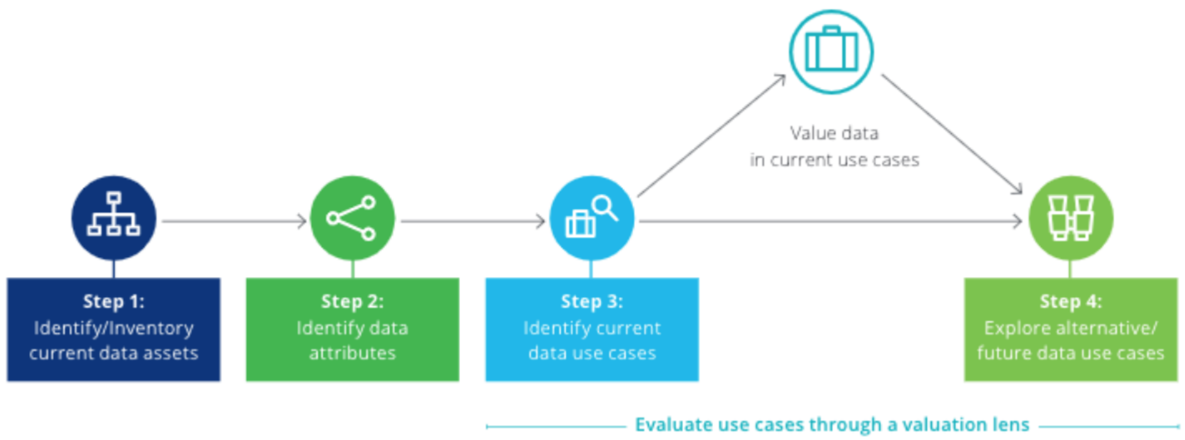


Figure 14: *The Steps to Understand the Data. Source: (Deloitte, 2020)*

Step 1: Identify current data assets and their attributes

The first step is to determine where the data is kept. It may need the usage of tools to update and maintain the company's data registries as well as, more importantly, determining how the business is currently utilizing its data—if it is being used at all (Deloitte, 2020).

Step 2 - Identify the data's key attributes

The data's key characteristics should then be investigated. Understanding the main characteristics helps in the creation of any use cases to optimize the effect of the data on the company. Data quality, targetability, source, universe, use case/ROI, market demand, uniqueness, and exclusivity are examples of these characteristics (Deloitte, 2020).

Step 3 and 4 - Identify current and alternative/future use cases

When data evaluation methods are used, they may uncover new use cases that can lead to the development of new commercial applications for alternative and defensive uses of data. The techniques of valuing data are defined by the data's current and future use (Deloitte, 2020).

3 Case presentation

3.1 Automatic Identification System

The *Automatic Identification System (AIS)* is a communication system that transmits ship movement and technical data at predetermined intervals over the maritime very high frequency band (Smestad et al., 2015). At the same time, AIS provides public authorities with an overview of ship traffic in various waters, which is shared with other government agencies and ports (Kystverket, 2021). The signal is sent by a ship's transponder and received by AIS receivers, which may be located on ships, buoys, land, and satellites (IALA, 2016). In addition, the receiver is often connected to radar and electronic mapping systems (ECDIS), where the ships' names and positions can be displayed in real time (Kjertsad, 2019). Thus, the data transmitted are both static data such as the ship's name, draught, destination, and estimated time of arrival and dynamic data from the ship's sensors such as speed and direction (Smestad et al., 2015).

Having access to AIS data expands research and development opportunities. Learning about shipping trends and using them in optimization models or simulations, new services, and apps, for example, may reduce uncertainty (Costa et al., 2020). Depending on the kind of boat providing the data, the data sent through AIS may vary significantly. Table 1 summarizes the kinds of data that may be transmitted.

Table 1: *Data Transmitted in an Automatic Identification System Message (SVB, n.d.).*

Static Data	Dynamic data	Travel-related data
Vessel's name	Position of the vessel	Current maximum static draught in dm
call sign	Speed over ground (SOG)	Port of destination
MMSI number and international call sign	Course over ground (COG)	Planned arrival time (ETA)
IMO number	Coordinated universal time (UTC)	Specification of cargo category if applicable
Type of vessel	Navigation status	
ship size	Heading	
	Rate of turn	

3.1.1 Prevalence of AIS

Regulation 19 of SOLAS Chapter V - The carriage requirements for Ship-borne navigational systems and equipment establishes the navigational equipment that must be carried on board ships based on the kind of ship. The law mandates that all ships with a gross tonnage of 300 gross tons or more involved in international journeys, cargo ships with a gross tonnage of 500 gross tons or more not engaged in international voyages, and all passenger ships, regardless of size, be equipped with AIS (International Maritime Organization, 2019). In addition, national standards will generally require ships not covered by these regulations to carry AIS transponders. The requirement implies over 85 000 ships sending AIS data across the globe (Mantell et al., 2014).

3.2 Use of AIS Data

AIS was created to enhance ship and environmental safety and, increase marine traffic services and monitoring, by enabling ships in close proximity to automatically communicate information about their course, speed, type, cargo, and other factors. The officers could prevent collisions by exchanging this information, which made it simpler for one ship to communicate with the other through radiotelephone. However, today's usage of AIS has well outgrown its initial purpose, and it is challenging to envision marine operations without it.

(Kystverket, n.d.)

Figure 15 shows how the signals are transmitted from the ship's transponder and received by the AIS receivers, subsequently sending them to the data provider. DNV uses AIS data in its services and has a contract with the data provider, which gives them access to the data. Before being utilized in a service or product, the data are aggregated and cleaned for their intended purpose. DNV obtains other data, such as weather data from external sources, which is bundled with AIS data in different use cases. DNV offers a variety of services for these data's many applications, as described in the subsequent sections.

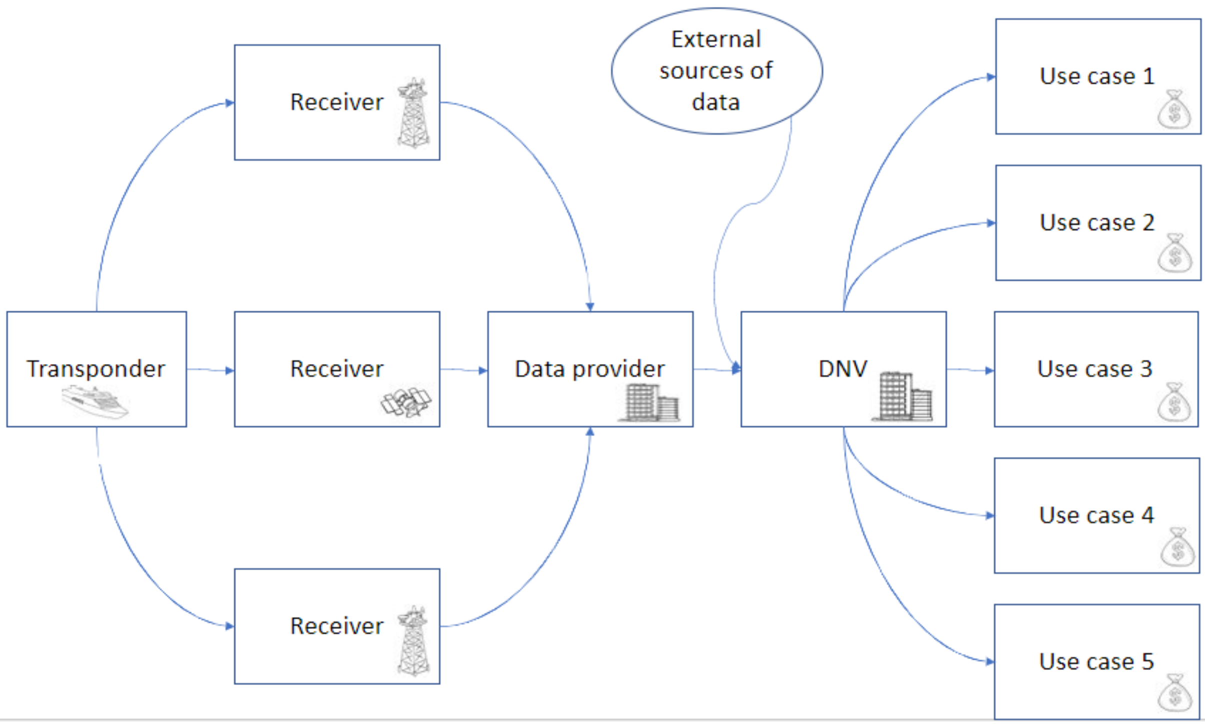


Figure 15: *The AIS Data Value Chain.*

3.2.1 Emission Estimation

All offshore businesses that travel in Norwegian waters are required to pay tax on the emissions they generate. In shipping, the tax covers emissions from vessels operating inside Norwegian territorial seas and domestic shipping, regardless of whether a portion of the trip occurs outside Norwegian territorial waters (Skattedirektoratet, n.d.). The Norwegian Coastal Administration is responsible for tax collection from ship owners and calculates emissions using DNV's emission model, which is based on a combination of AIS and ship data (Mestl, 2020).

3.2.2 Fatigue Damage

One use of AIS data is to monitor fatigue damage caused by voyages in rough seas, which cause the ship to twist and turn somewhat, causing tiredness in various locations throughout the ship. A boat is required to survive a certain amount of these twists and turns throughout its life. The AIS data is combined with meteorological data, and by applying a ship fatigue model to the data, they can determine how fatigue damage affects the ship's total lifespan. This may be used to sell secondhand boats, for example. Figure 16 shows how a fatigue model may be calculated (Mestl, 2020).

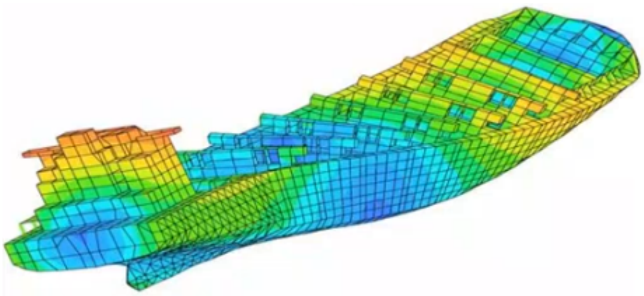


Figure 16: *Loss of Fatigue Capacity due to Heavy Weather (Mestl, 2020)*

3.2.3 Pipeline Damage Through Trawls and Anchors

This use case involves quantifying pipeline damage caused by boats colliding with a pipeline with their anchors, trawls, or anything similar. A critical element of quality is having sufficient data with a high temporal resolution. They detect anomalous ship motions near the pipeline, as shown in Figure 17, which may signal that the boat has been stuck in the pipeline. Data sets used along with AIS data are Fairplay data, which contains generic ship information such as length, height, and pipeline position data. The model is used to forecast when a pipeline will need repair (Mestl, 2020).

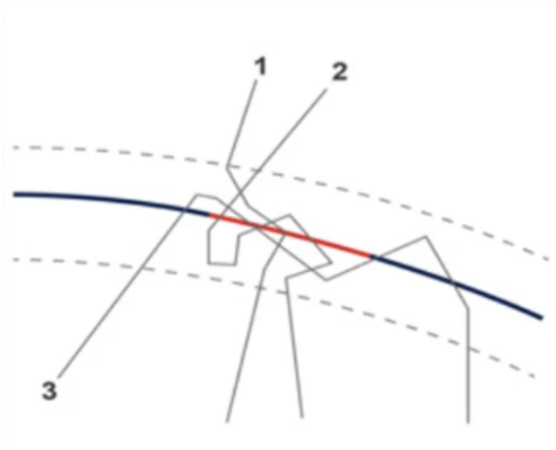


Figure 17: *Abnormal Vessel Activity Plot From 3 Boats Close to a Pipeline (Mestl, 2020).*

3.2.4 Voyage and Port Stay Information

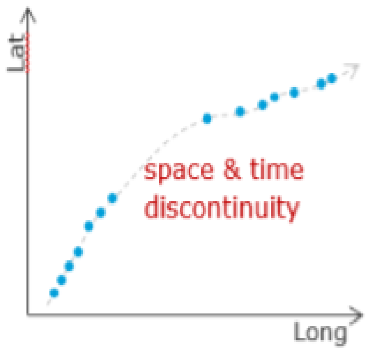
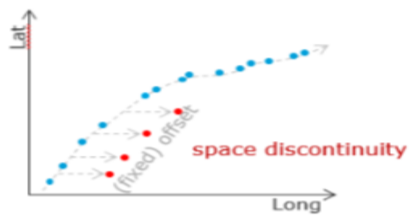
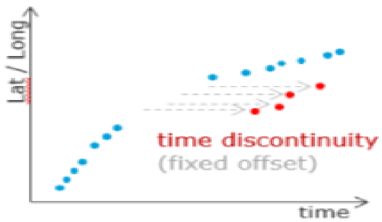
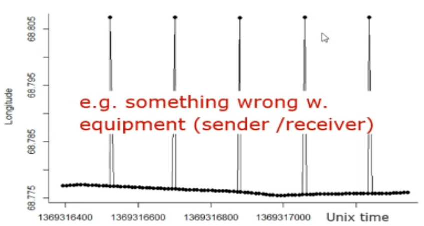
This use case involves the combination of AIS data with port information. Port shapes define the borders of a port. Historically, port shapes were drawn manually, but owing to quality issues associated with, for example, port construction and extension, DNV can compute the appearance of the port using an algorithm (Mestl, 2020).

DNV can monitor where boats are traveling, how they are traveling, and how long they are at a port by combining port shapes or port locations with AIS data. These data may subsequently be used to analyze traffic flow and for competitor analysis (Mestl, 2020).

3.3 Quality Issues

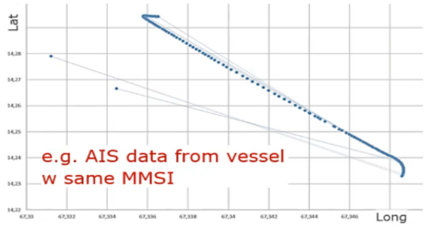
As with every data collection, the quality of AIS data is subject to specific problems. For example, around half of all AIS static data broadcasts include mistakes, and one-third of them have at least one inaccurate vessel identification (Winkler, 2012). In addition, there are a variety of errors that impact the quality of data (Mestl, 2020). A list of the most prevalent quality problems in AIS data is provided in Table 2.

Table 2: *Data Quality Issues in Automatic Identification System Data*

Graphic	Explanation
	This figure shows how a vessel moves across the globe at various time stamps. Discontinuity in space and time may occur due to the transponder being destroyed or switched off for various causes. These reasons may be genuine, such as protecting the location from pirates (Gard, 2019), but can also be due to illegal activities such as fishing in protected areas or entering another country's waters without permission (Oceana, 2018).
	A space discontinuity is a difference between the vessel's actual and reported position. This may be due to inconsistencies in GPS signals (Patroumpas et al., 2017).
	A difference between actual and reported time may be seen, resulting in this pattern of deviation from real time.
	Speed discontinuities are devations from the true speed at a particular time, as shown in this image, where time is plotted on the x-axis and speed is plotted on the y-axis. This may be due to incorrect speed readings from the ship's sensors.
	Significant outliers, such as shown in this figure, may indicate that something was wrong with the transponder at the given time, indicated by the x-axes, and the given position, indicted by the y-axes.

Continued on next page

Table 2 *Continued from previous page*

Graphic	Explanation
	Certain transponders share the same MMSI-number, which may make it seem as if a vessel has traveled a great distance in a short amount of time, but is really just the tracking of two different vessels.

There may be interference issues between the ship's AIS signal and the satellite signal, which has a considerably broader coverage area than AIS was designed for (Smestad et al., 2015). Interference may impede the detection of individual signals from a vessel in high-traffic areas. Multiple base stations on the ground may monitor the same ship, resulting in the base station sampling AIS data from the same vessel and tracking various time stamps for the same location. In Section 6, a discussion about how the above-mentioned data quality problems are related to ROI is given.

4 Research Methodology

The purpose of this chapter is to explain how the study was conducted. The chapter explains the research process, demonstrates how the study is designed to answer the research questions, and ensures that the decisions taken are transparent. The methods used for data gathering and analysis will also be discussed. Finally, thoughts on the study's method, data collection, quality, ethical concerns, and the researchers' roles are given.

The term ”research” refers to the methodical examination and examination of materials and sources in order to establish facts and reach new conclusions (Lexico, n.d.), and it should be conducted systematically to gain knowledge (Belbin., 1981). Research methodology refers to how a researcher plans a study in a methodical way to guarantee valid and reliable findings that meet the research aims and objectives (Jansen & Warren, 2020), and it is designed to achieve the research purpose best and to answer the research questions (Saunders et al., 2016). There is a distinction between methodology and method. A method is an instrument utilized to answer the research questions, and the methodology should influence the methods used to produce the persuasive evidence for a research endeavor (Brookshier, 2018). The credibility and reliability of a study depend on the choice of method (Johannessen et al., 2010). Therefore, specifying the technique and explaining why a specific method was chosen, according to Jankowicz (2002), is excellent research practice. To meet this criterion, the methodology was carefully considered, including an explicit and deliberate selection of approaches, methods, and techniques, as well as justifications for why they were chosen.

Following Maxwell's model of research design (J. Maxwell, 2013), the most significant connections between goals, conceptual framework, research questions, methods, validity, and reliability in this research are shown in Figure 18.

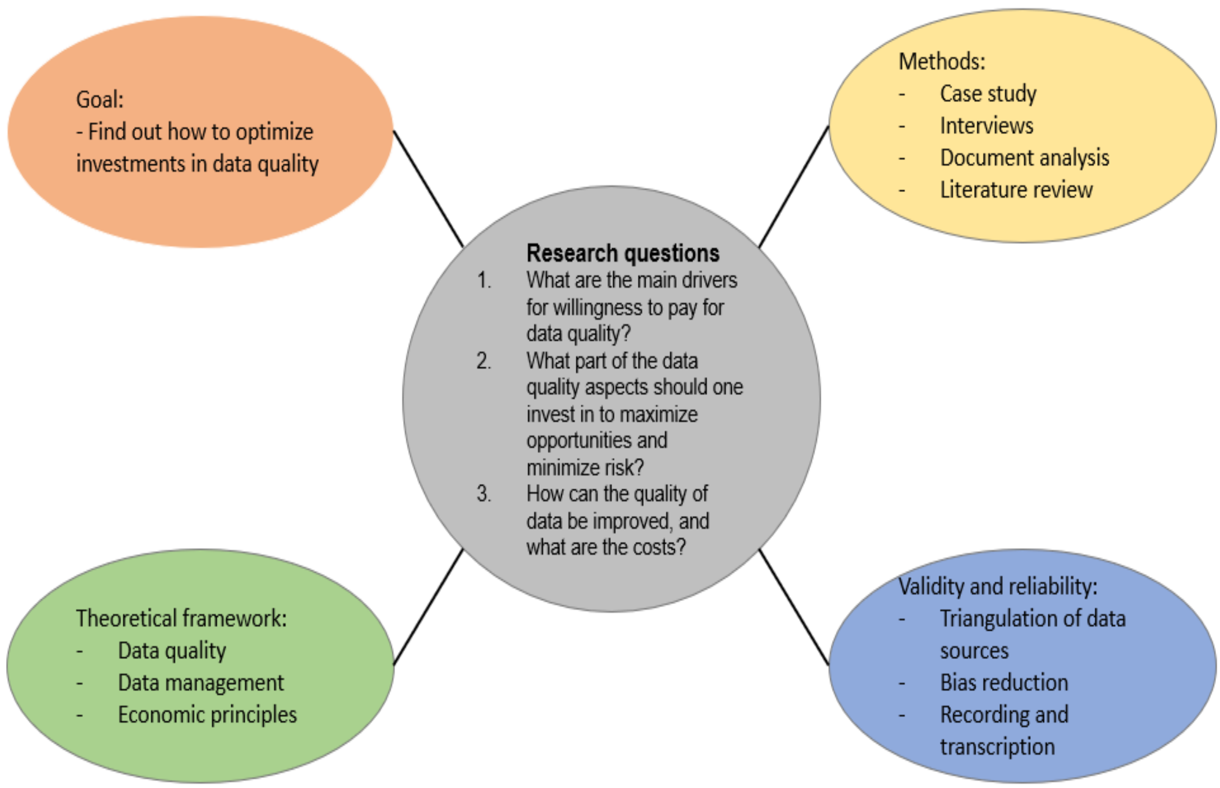


Figure 18: *Research Design Based on Maxwell's Model of Research Design (J. Maxwell, 2013)*

4.1 Research Process

A brief overview of the research process is required before the specifics of the study methodology and methods are given. The purpose of this section is to provide a picture of how the study was carried out. The research process is a series of processes that are used to gather and evaluate data to obtain new knowledge of a problem. Any research program's research process is split into two parts, according to Jankowicz (2002): the project process and the research process. The project process covers all study activities that rely directly on reading theory skills, and the second phase contains new material that is not immediately obvious.

The research process comprises a series of related activities or phases that must be completed to conduct the research successfully. A research process, according to Rummel (1963), consists of six closely linked stages. These include identifying the general area of study, choosing the topic, formulating a research plan and methodology, collecting and then analyzing the data, and writing the study. These processes may be divided into three phases: planning, research, and presentation (Coventry University, n.d.). With this information, Figure 19 depicts the overall research process for this thesis.

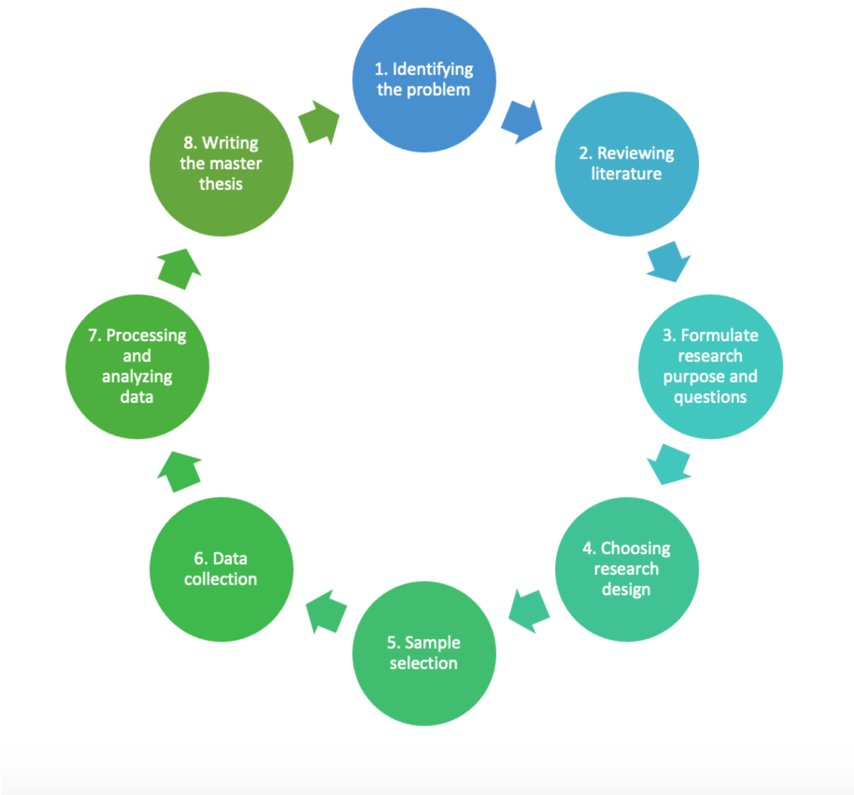


Figure 19: *The Research Process in the Study (iedunote, n.d.).*

We contacted DNV early in the process, and there were several conversations between some of the employees and us. The ability to combine industrial economics expertise with an area that would give value and benefit DNV was essential. Our experience in finance and technology, along with DNV's data quality expertise, prompted us to choose a topic at the crossroads of data quality and finance. We began to get an overview of the situation and map out which results had been produced before. During conversations with experts, investigations of past research, research papers, and theory, it was discovered that the combination of data quality and economic concepts was an area that had been rarely studied and about which there was little information.

First, the specific research area and the research problem statement were defined from the literature once the general study area was identified. The research problem statements were developed via discussions with experts and reading numerous publications on data quality, data management, data value, and data economic concepts. According to Ellis and Levy (2008), the research problem is the central and most essential part of any quality research. This informs the study's objectives, research questions, literature evaluation, methodology, results, and conclusions.

Second, the research purpose and the research questions were established. The research purpose denotes the study's primary aim or intention in addressing the problem (Creswell, 2008), and the research questions narrow the research purpose into specific questions that the researcher want to address (Creswell, 2008). Thus, the research purpose has been broken down into three research questions, as outlined in Section 1.3.

The research instrument was then selected. This indicates that the research strategy's data collection, analysis, and platform have been identified (C. University, n.d.). Afterward, data was collected from three companies (DNV, Orkla, and Fremtind) based on the research questions and research purpose.

4.2 Research Design

The research design and justifications for the research design are presented in the following sections. The research design is the strategy for fulfilling the research objectives and answering questions (Cooper & Schindler, 2008). It constitutes the plan for collection, measurement, and data analysis. The design was prepared with the goal of achieving the purpose and providing answers to the research questions (Choudhury, n.d.). Several factors influence the choice of research design. It is decided, among other things, by the research project's framework conditions, the problem, data needs, research perspective, approach to theory and empiricism, a method for data collection and analysis, and requirements for validity and reliability (Sander, 2020). These factors will, therefore, be discussed in more detail in Sections 4.2 and 4.3.

4.2.1 Type of Research

Pragmatic considerations influenced the research method and analysis that was used in this study. The relevance of quantitative and qualitative research methods, for this research in particular, was compared. At the same time, the choice of the research method was influenced by professional considerations, practical conditions, and limited resources.

Qualitative research is diverse. It can be characterized by being both empirical and theoretical, but usually in the interaction between these. It can be delimited by the environment (so-called case studies), phenomena, type of informants or perspectives, and theories (Tjora, 2015). The purpose of a qualitative research method is to gain in-depth knowledge of incidents, the course of events, opinions, assessments, arguments, decisions, measurements, or development trends (Jacobsen, 2005a). A qualitative research method emphasizes the understanding instead of an explanation of the topic as well as an open interaction between the researchers and the informants. Qualitative studies also focus on data in the form of text, rather than numbers (Tjora, 2015). It was discovered that little research had been done on the economic aspects of data quality before. There was also little data on the topic, and quantification would be difficult. According to (Ayiro, 2012), a qualitative technique was appropriate for this sort of research and was thus chosen.

As a qualitative method was chosen, an inductive approach (exploratory and empirically driven) was utilized rather than a deductive approach (theory - and hypothesis-driven) (Tjora, 2015). This is in conjunction with a desire to move from data to theory (Saunders et al., 2016). Furthermore, an inductive research technique is typically utilized when there is little to no existing literature on a subject according to McCombes. (2021). Therefore, an inductive technique was appropriate for this study.

The research design is determined based on whether the research design is explanatory, descriptive, or exploratory (DeCarlo, 2018). When it is critical to gather new knowledge in an area where little information exists, exploratory is the most suitable option (Jacobsen, 2005b). No previous similar research was found regarding the economic aspects of data quality. In addition, the research aimed to obtain fresh insight into the topic and it was important that changes could be implemented during the research process. Therefore, exploratory investigations were appropriate.

According to Grønmo, 2020, case studies may serve as a springboard for new discoveries, and case studies focuses on obtaining a holistic understanding of the situation (McCombes., 2021). Furthermore, the subject is broad and it was preferable to narrow it down by utilizing a case study. Therefore, the study was limited to AIS data, which DNV is familiar with and utilizes in many situations. It was decided to undertake a qualitative case study using various data sources, including a document analysis, literature review, and interviews to answer the research questions.

4.2.2 Sample Selection

In specific qualitative designs, such as case studies, sampling may not be relevant. The goal is to get a thorough understanding of a particular situation rather than generalize to a large group of population. Rather than sampling, the research might try to gather as much information as possible on the topic being researched (McCombes., 2021). Therefore, the informants were carefully chosen because competent informants were desired. Attempts were made to connect with individuals who had dealt with data quality, data management, risk management, financial elements of data, and contract negotiations with providers of AIS data. Consequently, a strategic selection was carried out. Strategic selection is the process of selecting informants who, for different reasons, will be able to express themselves reflectively on the topic at hand and who are not randomly chosen to represent the population (Jacobsen, 2005a).

During the interviews, the informants suggested other informants who might be interviewed. This method of selection is known as snowball sampling (Crossman, 2019). Snowball sampling is a non-probability sampling method (which includes purposed sampling) in which a researcher starts with a small group of known people and asks them to select others who should be included in the study.

Initially, there were 10 informants, but towards the end, there were 14 informants and

11 interviewees as three interviews were conducted in pairs since some informants worked closely together and had similar backgrounds. This was because it was not known how many informants would be required to answer the research questions. There exists no guideline for how many informants are required to answer the research questions. Steinar Kvale (2019) states that the number of informants is as many as you need to determine what you need to know.

Twelve informants work at DNV, one for Orkla and one for Fremtind, each in a unique role. Some work with AIS data, others sell data, some are data scientists, and others are involved in finance and risk management, while others are involved in research and data management. The common feature is that everyone has something to contribute on the topic of data quality within their field. The purpose of inviting representatives from other businesses is to increase diversity and see whether their perspectives differed from those of DNV's participants.

4.2.3 Data Collection

There are several methods used for the collection of data in qualitative research, such as participant observation, focus groups, qualitative interviewing, language-based approaches, collection and qualitative analysis of texts and documents (Bryman, 2012). Interviews, document analysis (documents from DNV), and a literature review were utilized to gather data for this research. As a result, both primary - and secondary data were analyzed.

Triangulation is a term that refers to the employment of several techniques and is used to obtain a holistic knowledge of phenomena (Patton, 1999). Rugg (2010) distinguished four different kinds of triangulation. Data triangulation is the utilization of numerous data sources in a single study. Investigator triangulation is the use of more than two researchers in an investigation, theoretical triangulation is the use of multiple theories in the interpretation of a study's findings, and methodological triangulation is the use of multiple methods to conduct a study. This research prefers to use data triangulation because three data sources are used in the same study. The utilization of several data collection methods and triangulation is beneficial for data validation - ensuring the research's quality and trustworthiness. Figure 20 illustrates the triangulation validation approach.

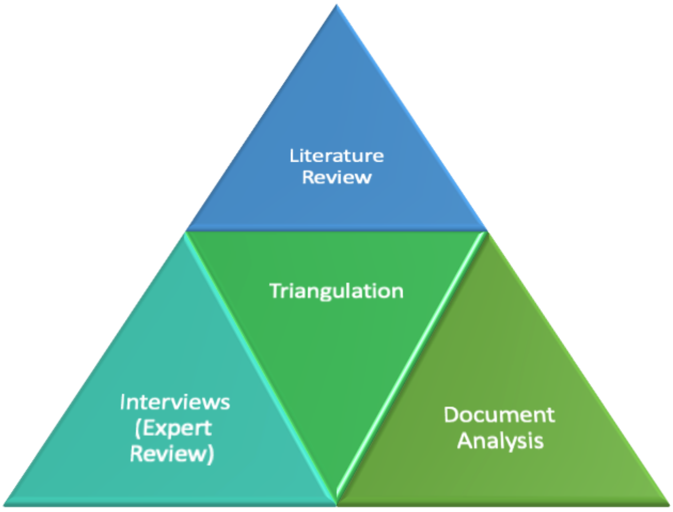


Figure 20: *Data Triangulation Validation Method (Patton, 1999).*

4.2.3.1 Literature Review

A literature search on relevant subjects in the study area was used to perform the literature review. According to (A. University, 2018), a literature review should be conducted for several reasons: to ascertain what has been investigated and what has not, to identify data sources that have been used by other researchers, to contextualize the work, to provide evidence to support the conclusion, and to advance the field through research advancement. All these reasons were desired in this research, and therefore, the literature review was appropriate.

The data-gathering process began with the examination of several relevant papers. A substantial amount of literature was examined in the broad area of the study and other basic theories underlying the research's primary objective. Significant time was spent assessing sources, and every piece of literature was scrutinized. Checkpoints, such as who published the source, its purpose, the year it was created or updated, and its content, were all thoroughly examined. Due to the lack of theory in the area, some papers were cited even though they were published several years ago, and some had no dates. Literature by DNV and material from universities, prominent research organizations, and businesses were studied and used in this study.

4.2.3.2 Document Analysis

According to Bowen (2009), document analysis is a kind of qualitative research in which the researcher interprets documents to give them voice and meaning about a specific evaluation subject (Bowen, 2009). Document analysis involves categorizing information into themes in the same way that focus group or interview transcripts are examined. This was a useful tool for data control when combining the interviews. Several informants in DNV shared links to websites, publications, and other materials supporting literature throughout the research.

This has aided in obtaining a more excellent knowledge of the subject, AIS-data, DNV, and the data value chain.

4.2.3.3 Interviews

Based on the research questions, it was determined that an interview-based approach would be appropriate for this study. Furthermore, the study emphasizes a phenomenological and hermeneutic interpretive framework. This entails an interpretation of the phenomena reported by the informants, which is why the interviews were performed in a semi-structured manner (Krumsvik, 2014). A semi-structured interview is a technique often used in qualitative research. It is defined by its openness, which allows for the introduction of new ideas throughout the interview because of what the informants say (Tjora, 2015). For semi-structured interviews, an interview guide is used, which includes a list of subjects to cover during the interviewer's discussion with the respondent, and the question formulations are tailored to each individual respondent (Malt & Grønmo, 2020). The approach allows for the desired informal conversation (Johannessen et al., 2011). Therefore, this enabled informants to express themselves freely. Additionally, this enabled the informants in the research to more readily follow the questions that were desired to explore (Yin, 2014), and it allowed for the possibility of following up on broad inquiries with more particular questions, a process known as ”tunneling” (Krumsvik, 2014). We structured the questions by creating the interview guide, and at the same time, the informants could speak about whatever they wanted. It was therefore semi-structured.

The question formulations were tailored to each respondent's background and job responsibilities. The interviews were conducted individually and in groups. The group interviews were conducted by two people working on the same subject of interest for a long time. From Tjora (2012) framework, an in-depth interview was used to create the interview guide. The structure of in-depth interviews, according to Tjora (2012), goes through three phases: warm-up, reflection, and closing. The warm-up questions were simple questions to ”warm up the conversation” regarding the background and work tasks. This was to get to know each other and to provide a transition into the topic. Therefore, the interviews began with basic, fact-based questions. In the reflection phase, according to Tjora (2012), the researchers address questions, so-called substance questions, that are more demanding of the informant's thoughts and reflection. These were questions related to the research questions. Therefore, the questions needed to be open-ended to ensure the informants could speak for an extended time. This is because the participants are more knowledgeable in their area than the researchers, and there might be information that the researchers have overlooked that is critical to the study's success. The questions were written in a way that was as easy to understand as possible, in accordance with the theory of Jacobsen (2005a), which states that all informants must understand the questions for them to provide relevant information and confirm or deny something that one was unsure of. The closing questions, according to Tjora (2012), are intended to check whether the informants have anything more to say or if they have any suggestions for additional topics that should be covered. The informants were therefore asked if they had anything to add.

It was not feasible to meet the informants in person during the COVID-19 pandemic because of the restrictions. As a result, the meetings were held on Microsoft Teams, which is a collaboration platform. The meetings were recorded to capture audio, video, and screen sharing activity in consent with the informants. Because of the confidentiality agreement with the informants, only trustworthy services for recording the interview material were considered. The platform was determined to be reliable and had a desirable recording quality based on prior experience with Microsoft Teams. Moreover, DNV, Fremtind, and Orkla utilize this platform.

The videos and audio recordings have been stored in closed folders to maintain the material for security reasons. Our access to this material is likewise restricted to the study time for the company's safety.

The interview participants were invited for interviews via email. A paper describing the study's topic, the problem, and the research questions were sent out to all informants. We called it a "one-pager." They also received information via email about the procedure, including how the material should be utilized, anonymity, re-reading, and other information, including a request for their willingness to participate in the interview. This was done to examine the expectations we had for the study and to double-check whether the subject matter matched some of their expertise. In addition, to emphasize that the participants were treated with dignity and respect (Brinkmann & Kvale, 2005).

If the potential informants agreed to participate, the interview guide was sent well before the interview, at least 24 hours before. The interviews did not have to be conducted in any particular order. Therefore they were conducted at a time convenient for the informants. Each interview was scheduled to last about one hour. At the end of the interview, several of the informants volunteered to answer more questions in the future if needed.

According to Jacobsen (2005a), the further process should take place immediately after the interview has been completed (e.g., same or next day). Consequently, the transcription of the interviews into text began no later than 24 hours after the interview ended. The interviews are structured when transcribed from oral to written form, making them more suitable for analysis (Steinar Kvale, 2019). Furthermore, Kvale (1997) claims that there is no accurate translation from oral to written form, and he recommends considering how to transcribe based on what is helpful in the specific situation. Therefore, the transcription method was considered based on utility value in this situation. To perform a comprehensive analysis, it was found necessary to have a high degree of detail in the data material. As a result, the interviews were meticulously transcribed, including filler words. Filler words, according to Worthy (n.d.), may occasionally provide a reader with better knowledge and insight into what was going on in a respondent's thoughts while answering a question. The transcript was written in the same language as the informants. This was to ensure that no errors or misunderstandings occurred during the transcription process.

4.2.4 Analysis

Thematic analysis and discourse analysis are the two most used methods for analyzing qualitative research (McCombes., 2021). This study uses a thematic analysis. According to Braun and Clarke (2006), this study is characterized by a heavy reliance on interviews and textual data, the creation of themes based on recorded interviews, coding and text analysis, and a range of 4 - 30 informants. The method therefore suited this study. According to (J. Caulfield, 2019), the most frequent method for performing thematic analysis is a six-step procedure, as shown in Figure 21.

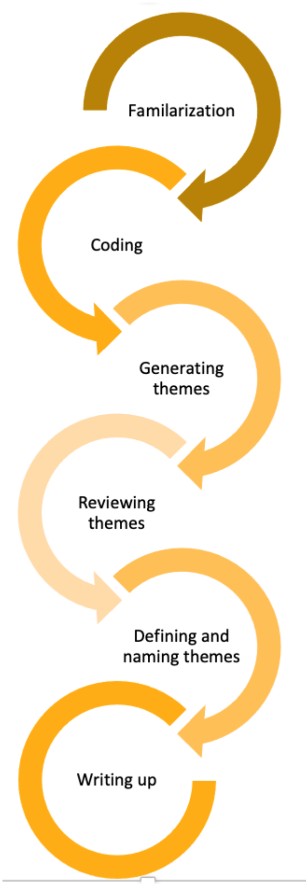


Figure 21: *The Six-Step Process for Conducting Thematic Analysis (J. R. Caulfield, 2019).*

Familiarization is the first step in the analysis. First, we must understand our data. Then, before analyzing individual items, we need to obtain a good picture of all the data we gathered. This may include transcribing audio, reading text and taking notes, and generally familiarizing ourselves with the data.

The next step is coding. According to Tjora (2015), the aim of the coding is threefold: to extract the essence of the empirical material, reduce the volume of the material, and facilitate idea generation based on details in the empirical data.

Coding involves highlighting text and giving it brief labels or ”codes” to explain its content. In this excerpt, we have colored-coded several sentences. Each code represents the text’s concept or emotion. At this point, we want to be thorough: we read through every interview’s transcript, highlighting everything necessary or possibly intriguing. We may keep adding new codes as we move through the text and highlight existing ones. After reading the text, we organize the data into groups by code. These codes provide us a quick summary of the data’s key features and common meanings.

Then we go through our codes, search for patterns, and start creating themes. Their scope is more significant than codes. Most of the time, a theme is made up of multiple codes, which they were in this situation.

Now we must ensure that our themes accurately reflect the facts. Returning to the data set, we compare our themes to it. Is something missing? Are these themes present in the data? What can we do to improve our themes? We may divide, merge, delete, or develop new themes if they are not helpful or accurate.

Now that we have a list of themes, we name and define them. Defining themes requires defining each topic and how it helps us comprehend the data. Theme naming includes giving each theme a short and simple name.

Finally we will write out our data analysis. Writing a theme analysis needs an introduction to define our research question, aims, and approach. We should also include a methodology section. This section should describe how we gathered the data (semi-structured) and how we performed the thematic analysis. It typically handles each topic individually. We illustrate how often the themes appear and explain their meaning using data samples. Finally, our conclusion summarizes the essential findings and answers our research question.

4.3 Reflections

There will always be advantages and disadvantages to using a specific method to answer the research questions (Hoepfl, 1997). In the following section, the suitability of qualitative methods, case studies, data collection, and their benefits and drawbacks will be discussed along with the study’s quality, ethics, and our responsibilities.

4.3.1 Method

There is a perception that qualitative research is more subjective than quantitative research. On the other hand, according to Tjora (2015), this is a wrong statement on a broad level. To be sure, quantitative research’s mathematically based analysis techniques are objective in the sense that they are unaffected by the analysis. The interpretation of the results, on the other hand, will depend on the kind of theories and perspectives the researcher uses.

What is true, however, is that qualitative analysis reflects a more significant degree of researcher subjectivity than quantitative design's strictly mathematical analysis does. Theory and researcher subjectivity become important in quantitative studies only after the interpretation of the findings and the fact that they influenced data collection in the first place. Qualitative analysis often begins with (theoretically inspired) interpretation (Tjora, 2015).

One of the most challenging aspects of research, in general, is to delimit the empirical work, according to Tjora (2015). The approach used in this research is utilizing a case study in the form of AIS data. This has enabled an exploratory approach, at the intersection of economic aspects of data quality. Although this is a phenomenon of limited understanding, we have chosen specific theoretical methods to examine it based on gaps in the literature. This has allowed the information gathered during the research to be organized. However, it may also mean that the discovery has been interpreted in light of the conceptual perspectives that were chosen, and a limitation thus lies in not finding out what happened (Widding, 2005). The case was selected pragmatically based on its accessibility and the informants' familiarity with it. However, it must be noted that the example chosen for convenience is not always optimum for maximum generalizability. Furthermore, companies were selected for strategic and convenient reasons. Because this is a tiny sample, this can weaken a generalization to a larger population (Yin, 2014).

All the interviews were transcribed before being analyzed. Therefore, it is conceivable that some information got lost. Moreover, since the study used a semi-structured method, much unnecessary material was included resulting in lengthy transcripts. On the other hand, measures such as incorporating breaks, ambiguity and filler words, have been implemented to avoid losing information and to remember better.

A method using open interview questions linked to the study topics was selected to capture the aspects that the informants feel are most relevant surrounding this subject. This method has allowed the informants to record aspects that are significant to them both implicitly and explicitly.

On the other side, the informants may misinterpret the jargon and speak about different topics (Krumsvik, 2014). Therefore, the risk of losing the essential information is present. Therefore, after the findings were finalized, the informants were given a presentation from us. The presentations mainly served as polo interviews, with comments on whether the informants grasped the findings and ideas. As a result, the risk of misunderstandings during the interviews was reduced, consequently maintaining the reliability of our results.

4.3.2 Validity and Reliability

The terms reliability and validity are used to assess the quality of research. They indicate how well a method, methodology, or test measures something. Validity is concerned with its accuracy, whereas reliability is concerned with its consistency (Middleton, 2019). According to Krumsvik (2014), validity can be split into intern and external validity. Internal validity is the degree to which the study has minimized bias in the research, and external validity refers to the degree to which research findings can be generalized.

Internal validity was maintained by concentrating on bias reduction throughout the interview process. According to Steinar Kvale (2019), the continuous process validation is divided into seven stages; thematization, planning, interview, transcription, analysis, validation and reporting. Thematization is concerned with theoretical robustness and the degree to which theory and research questions are linked. We provided theory in a framework directly related to the research questions, including data quality and economic considerations. There seems to be coherence between the different components in this manner. Theoretical robustness has been improved thanks to a thorough literature review. Internal validity has been considered throughout the design process of this study, as it relates to the selection of a research method (case study) that can answer the purpose of the study. In addition, the technique (semi-structured interview) used to address the research questions is described. Internal validity concerning interviews is also discussed in Section 4.3.1. Validity in transcribing entails treating the transcripts with care and replicating the informants' responses precisely as they were given (Krumsvik, 2014). This has been ensured by being detailed with the transcripts.

One aspect of the research that is more difficult to separate from preexisting notions is the interpretation. According to (Svartdal, 2019), this influence is unavoidable in any qualitative investigation and is referred to as the "confirmation bias." Even though the impact is present, we were aware of it. First, to reduce the bias, we meticulously coded the data. Second, gave a presentation to the informants who supplied the data to see whether our interpretations are consistent with their views. Third, triangulation occurred, by validating the data through cross-verification from three different data collection instruments, which implies that other sources of data were utilized to corroborate the interpretations. However, the literature review discovered that studies provide widely disparate results, indicating that there is "no general agreement" on economic principles for data quality. As a result of efforts to debunk data and assumptions, the study has escaped the confirmation trap to the greatest extent feasible. Another discovery from Peredaryenko and Krauss (2013) is that by interpreting the findings in light of current theory, the prior knowledge has permitted new perspectives. Therefore, the literature review might help to avoid preexisting notions.

When the study and discussion were finished, the thesis was submitted to experts in the field to contribute to the validity. These were some of the individuals who had been interviewed and those from other businesses who had knowledge of the subject. In this approach, the findings were guaranteed to be of high quality.

External validity is often associated with the word "generalization." According to Widding (2005), generalizability has been cited as one of the most challenging aspects of case studies. A problem with generalizing from case studies is that it's possible to discover variables that lead to specific results that are absent in other comparable instances. Because this thesis looked at one use case, generalizing results will be difficult. On the other hand, the findings are the result of a series of trials, each of which gets the researcher closer to the solution (Yin, 2014). There are additional parallels to be seen with the replication logic, and AIS data may be examined as a stand-alone example (Eisenhardt & Graebner, 2007).

According to Ferreira et al. (2020), the replication of research is linked to reliability, and they stated: "As it is generally not possible to replicate a case study under the conditions in which it occurred, its reliability is fundamentally demonstrated by the triangulation of data." Specific approaches may be employed to improve the study's reliability, such as recording the interviews, coding the answers, or using analytical data processing methodologies. We recorded the interviews, coded, and used a thematic analysis to improve reliability. Furthermore, the transcription was done in great detail to be as reliable as possible. When informants searched for words, this could highlight an uncertainty that could impact the analysis. As a result, it was included first and then deleted from the extracts. Some of the loss from interview to transcription were visual clues and mood during the interviews. However, since the interviews were recorded and videotaped, important information was not lost throughout these "translations."

Thematic analysis was chosen as an analysis method and is a flexible method. This means it is not bound by any certain research traditions or theoretical framework. This, according to Wold (n.d.), is one of the method's advantages. In addition, it provides for a variety of theoretical approaches and additional analytical options. However, the great degree of flexibility may make developing precise recommendations challenging. Non-fixed criteria for what a thematic analysis should include may be problematic, and as a result, maintaining reliability may be complex. According to Braun and Clarke (2006), a rigorous thematic analysis can produce trustworthy and insightful findings. As a result, considerable time was set aside to accomplish this adequately.

4.3.3 Ethics

Research ethics is concerned with principles, norms, and guidelines for determining whether acts are excellent or unethical, particularly in interpersonal interactions (Johannessen et al., 2010). Ethical decisions do not belong in any single part of the interview but must be considered throughout the whole research process (Brinkmann & Kvale, 2005). Because the chosen technique requires contact with a range of people, privacy was essential throughout the study. According to Tjora (2015), the ethics of research in the actual conduct of the interview is primarily linked to the requirement that the information should not be harmed.

The informants were reminded that they could end the interview at any time without explanation. Participants could also withdraw at any moment, also during and after the interview, or request that parts of the interview should not be included. In addition, the informants had the option of not answering any questions if they desired.

Because DNV, Fremtind, and Orkla are large corporations, no name or job title was provided; thus, this will be perceived as an anonymous interview. After transcription, personal data was anonymized and collected securely, and audio recordings were securely deleted. In addition, the citations have been rewritten, which reinforces anonymization. Anonymization in this study would not remove relevant information, and the analysis, therefore, does not suffer from this. Another reason that the quotes were rewritten was that the quotes were not so oral that the participants could appear to be poorly worded. According to Tjora (2015), there is an ethical requirement to always present informants in a respectable manner, and experience shows that some adjustment and clean-up must be expected in the transition from oral to written language.

4.3.4 Researchers Role

Great demands are placed on the researchers' bias and ability to be neutral (J. A. Maxwell, 2012). Both subjective experiences, expectations, attitudes, and what the researchers anticipate discovering may affect the information collected. Academics must understand their function as researchers in this situation.

Since the informants worked for companies with a high presence among NMBU students, we will inevitably have preconceived attitudes toward the companies. However, it is vital to separate these opinions from data collection (Peredaryenko & Krauss, 2013) and avoid nepotism. Therefore, the interviews were approached with an open and impartial mind. Furthermore, by setting protections for the design of the interview guide and the makeup of the informants, these attitudes were limited.

The usage of proprietary corporate communication tools such as Microsoft Teams and Zoom may make employees more available in various ways. At the same time, COVID-19 has brought a phenomenon to life: Zoom fatigue (Jiang, 2020). The strange echoing, delays in connection, screen freezes, and other technical problems that degrade the sound and image quality are blamed for this anguish. Furthermore, virtual meetings limit our ability to interpret the subtle body language signals of those in the meeting (Stepanyan, 2020), lowering the research's confirmability. Therefore, measures were taken to place the camera further away from the face so that more of the whole body was visible. In addition, we were careful to have eye contact so that the informants felt seen and listened to. Furthermore, the presentation for the informants may have helped enhance the confirmability.

5 Results

This chapter presents the results from the analysis of the interviews. The first section covers the first research question, the factors influencing people's willingness to pay for data quality. The results of the second research question are given in the second section, which contains a list of data quality factors to concentrate on to maximize opportunities and minimize risk. Finally, the last section summarizes the results of the third research question, which focuses on improving data quality and how much it costs. The observation is given first, followed by a more detailed explanation.

5.1 What are the Main Drivers for Willingness to Pay for Data Quality?

Observation 1: Increasing confidence in data in order to enhance company operations is one of the motivations for willingness to pay for data quality.

According to the interviewees, customers' confidence in the data influences their willingness to pay for data quality. Among several factors, accurate statistics were mentioned several times. When the informants were asked why they desired more accurate statistics, most respondents replied that they needed to trust their data. Several respondents remarked on this driver, stating that the higher the quality of the information, the more usable it is and the more likely people are to become dependent on the information. Additionally, superior quality is considered to offer greater security for the company and the products or services it produces. It was also stated that understanding the data quality makes companies aware of their weaknesses and may compensate for the quality issues. Two respondents said that they tended to use a statistical approach when using the data for decision-making. They highlighted that they wanted to increase their statistical model's likelihood of success to obtain better confidence in their decisions.

Observation 2: Reduced operational cost is one driving factor for willingness to pay for data quality.

The informants identified the second driving factor as the desire to reduce the company's operational costs. Thus, according to respondents, the primary reason for improving data quality was to decrease operational costs, at least for the master data.

Observation 3: Being aware of the uncertainty inherent in data is one factor contributing to a willingness to pay for data quality.

The third driving factor for willingness to pay was the user's awareness of the data quality. There are two different degrees of awareness of data quality. First, the understanding that the data quality is improved allows the user to use their data better. Second, if the users know how good the data quality is, they may be willing to pay for a data set even though the data quality is low. This knowledge makes the user aware of the weaknesses and how uncertain their decisions based on the data are. According to the respondents, individuals in the business are entirely reliant on insufficient data to make decisions, and knowing that there is uncertainty makes them familiar with data quality issues.

Observation 4: A greater level of knowledge in data quality gives a higher willingness to pay for data quality.

There was a clear relationship between data quality, competence, and willingness to pay. Respondents said that when their customers had more expertise in data quality, there would be a higher willingness to pay for the quality. Additionally, the informants stated that the industry lacked competency for those who are not professionals.

Observation 5: Awareness of the current state of data quality and how it may be improved to increase the value is one driving factor for willingness to pay for data quality.

Other respondents also mentioned awareness of the value of data. The respondents pointed out that one driver would be to know that there lies much value in data. If they do not focus on it, they would have garbage in, garbage out.

Observation 6: Demonstrating to the customer how quality improvements may enhance profit at each step of the value chain contributes to the willingness to pay for data quality.

The respondents emphasized the importance of customers understanding their business processes. Many companies are struggling to see the benefits of improved data quality. According to the respondents, the willingness to pay for data quality increases by describing how the improvements will affect profits at each step in the value chain.

Observation 7: An increased number of use cases can increase the overall willingness to pay for data quality.

Better quality improved the willingness to pay in different ways, according to the respondents. An increasing number of use cases for the data may increase the willingness to pay. By improving the data quality, new uses for the data may become apparent, and each of these new use cases may have an increased willingness to pay. The more the willingness of consumers to pay for the data, the higher turnover and earnings for the business.

Observation 8: Possessing the ability to analyze in the desired manner is one driving factor contributing to the willingness to pay.

Another driver for the willingness to pay for data quality cited was the customers' preferences for which analysis they wanted to do. For example, the customer may be interested in a combination of different data. The customer will then go to the data provider with the best bundling of data, the best visualization options, and the most flexible ability to analyze data. The informants also pointed out the importance of flexibility regarding clients' contracts with the data provider. They said that the more restrictions in the contract on how the user can use the data, the less likely it is that customers will pay for the data.

Observation 9: One driving factor that influences willingness to pay for data quality is the data provider's reputation.

Another point that was raised was the brand's reputation as a data provider. According to the respondents, a consumer will typically choose to do business with a company that has shown stability in the market over a more extended period. These are factors such as service, financial solidity, and whether the supplier has an exciting ownership structure. Although another rival may have better data quality, the customer will still choose a reliable competitor.

Observation 10: Possessing data in a desired format is one driving factor for willingness to pay for data quality.

The respondents said that the data providers' expertise in data handling, management, and delivery was essential. They said that the service providers might have a limited understanding of the data format for analyzing them. Consequently, the data customer needs to try to clean the data because the data is in the wrong format. This cost could be handled by the data supplier and thus raise the willingness to pay by the data customer.

Observation 11: There is a positive correlation between the importance of the data and the customers' willingness to pay.

Another factor mentioned was that the more critical the data are for a use case, the more the customer's willingness to pay for data quality. Therefore, according to respondents, potential profit will depend on how central the data are in the use case.

Observation 12: Drivers for the supply and demand curves are vaguely stated by the informants.

Answers to questions regarding the impact of data quality on supply and demand proved to be ambiguous. The respondents gave only partial answers. Several respondents said that they had not thought about it from this perspective.

Observation 13: The price and quality relationship seems to be nonlinear.

The correlation between price and quality does not seem to follow a linear curve. In the respondents' opinion, increasing the quality of the AIS data from good to excellent, for example, would not result in a statistically significant rise in the customers' willingness to pay.

5.2 What Part of the Data Quality Aspects Should One Invest in to Maximize Opportunities and Minimize Risk?

Observation 14: It is critical to define and measure the quality criteria for the use case to determine whether or not the service is of sufficient quality.

When the informants were asked about investment strategies for data quality, quality standards were mentioned as a common theme. Most respondents agreed that quality criteria should be explicitly stated for the specific use case for which the data should be used. If the criteria were not correctly specified for the desired service, it would not be possible to determine whether the solution meets the requirements or not, nor can one assess how efficient the service is for the business. It was clear that no quality requirements could be made until a set of quality criteria was established.

Observation 15: Prioritizing improvements are based on the time and cost of the particular improvement. A way to reverse-engineer the impact of various improvements can help to identify how to improve the quality.

The informants proposed an approach that used the quality dimensions after the respondents agreed on how to handle the problem in a more general way. For example, this approach may identify which elements of a data set's quality impact and strength on the data model. According to the respondents, the use of this method in combination with a cost and benefit analysis shows how to "climb up" on the ROI graph, as shown in Figures 22 and 23. DNV is conducting research on this technique and developing expertise in this type of method.

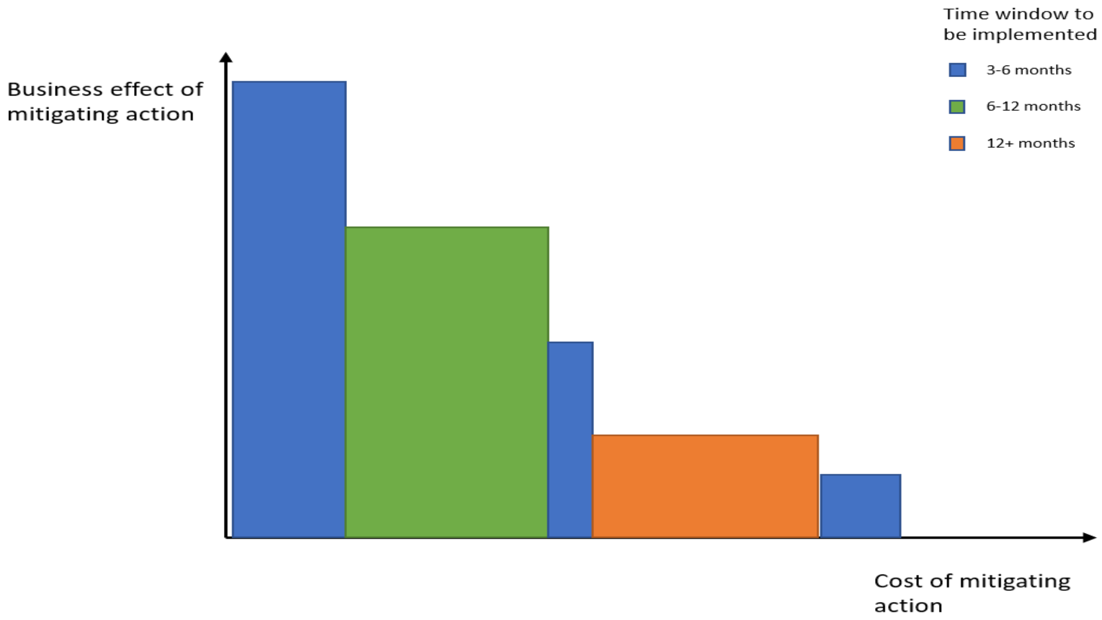


Figure 22: *Overview of How the Different Quality Aspects may be Ranked and Compared*

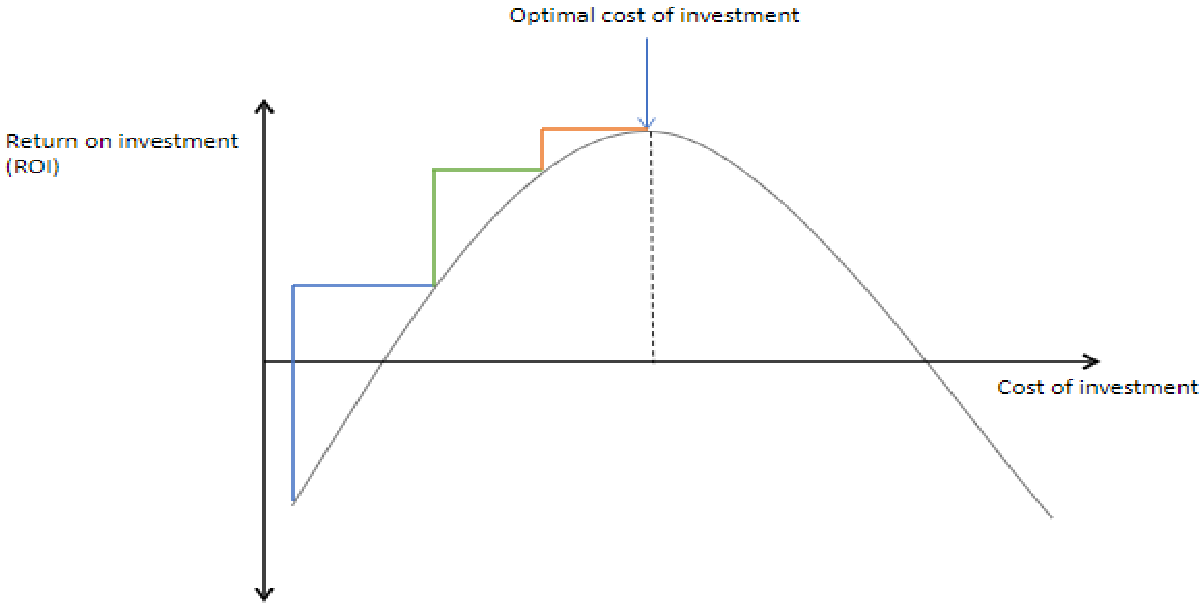


Figure 23: *How the Different Quality Improvements in Figure 22 Affect the Return on Investment.*

Observation 16: Involve the individuals who sell the product to help decide which improvements should be prioritized and which should be ignored.

According to the respondents, it was not always the same people who worked with data who got involved in the investment choices. They believed that the inclusion of people in the process of working with data quality could increase the likelihood of making appropriate investments. The respondents' responses indicated that it is essential to include people who have access to the data or are in the process of selling it in the investment discussion. A common occurrence is that decisions about quality investments are made in departments other than those that are responsible for handling and selling the data.

Observation 17: It is critical to invest in a professional, highly skilled team environment to succeed in data quality investments.

According to the respondents, it is critical to understand that high-quality data has significant value. The informants stated that having a professional environment that understands data quality, its theory, and collaborates effectively with other business services is critical for success with data quality initiatives. In other words, for data quality to be successful, the organization must invest in a professional atmosphere. Most respondents agreed on the significance of considering data as an asset.

Observation 18: Data quality investments should be made over a more extended period than the next quarter, and the ROI should be calculated over a longer time frame.

The informants said it was essential to understand which period the company has for its services. Data quality should be considered to have a long-term aspect because the data are not only used tomorrow but for a more extended period in different known and unknown use cases.

Observation 19: Investing in a flexible contract and managing it reasonably internally in the organization was necessary for the investments.

The contract with the data supplier is one of the most often cited issues by the respondents. They agreed that the contract should contain terms and conditions that permit the commercial use of the data while also guaranteeing that the data are of good quality. DNV has just changed the source of its AIS data. According to the contract conditions, DNV can only use the data in specific ways in its services. DNV changed the contract's conditions to have greater freedom regarding what it might do with the data. Because the data quality on the different contracts was almost similar, the sole reason for changing the data provider's contract was entirely commercial.

The management of various contracts internally was also mentioned. DNV has previously had little influence on the company's contracts. As a result, there was little awareness of the value of data across different goods and services in other departments. DNV then saw the value of an increase in legal and commercial awareness. Collecting all the data that was purchased into the organization in one place and then distributing the cost to the various departments that used and sold the data was profitable.

Observation 20: To cope with data quality, it is necessary to invest in security.

The importance of data security was also addressed in connection with Research Question 2. The respondents said that managing data quality requires data security. In addition, internal processes, such as management, control, and administration, were considered essential by the respondents. Data security is critical because the organization may risk losing control if someone outside manipulates or corrupts the data.

Observation 21: When a company's data quality maturity is low, it generally implies that quality enhancements provide significant returns.

Some respondents emphasized the company's maturity of data quality. They discussed the need to conduct a comprehensive study of the ROI. Furthermore, they stated that this requirement varies depending on how mature organizations are with this type of investment. One respondent said that it was not necessary to make an ROI analysis on improvements, as a result of which it was clear that they would profit from it anyway due to the lack of maturity. According to the respondent, there would be much profit in the beginning when organizations invest in data quality, but as the data quality increases, the willingness to pay decreases.

5.3 How can the Quality of Data be Improved, and What are the Costs?

Observation 22: To know whether the organization is investing in data quality optimally, it will need a deep understanding of the business and experience with it.

As stated in Section 5.2 relating to Observation 15, the cost of improvements is highly dependent on how the company prioritizes which aspects to invest in. Nevertheless, it is essential to recognize when the organization is not investing as efficiently as it might do. That is the point at which any further investment does not provide a profit. According to the respondents, it depends on the use case and mainly on the experience of the people in the company. They tended to lack a clear understanding of the signals to look for when they are not investing optimally.

Observation 23: The improvement cost is usually related to physically fixing transponders and receivers, which is significantly high.

The cost of improving AIS data, according to the respondents, is closely linked to physical objects, such as transponders and receivers. The transponder may be poorly configured and damaged, causing it to send the wrong signals. Some areas of the earth may also be inaccessible to the recipients. The only option to improve the quality is then to meet on the ship and fix the transponder physically or to build new base stations or satellites to receive signals from all over the world. These improvements have a high cost.

Observation 24: It is possible to compare the data with different data sets with a known quality.

Comparing the data with various data sets was one method suggested. DNV, for example, maintained a separate data set for comparing the length and breadth of data signals. In addition, the signals may be analyzed to determine which data source is the most reliable.

Observation 25: Cleaning data may be done using a variety of data science techniques.

The informants also suggested computer science methods as a technique. First, the dataset was sent for explorative analysis. Then they examine the data and evaluate methods depending on the context. This can be, for example, what they will perceive as the advantage of cleaning the data and what they think will affect the model being trained on the data later.

6 Discussion

This chapter discuss the results of the study presented in Chapter 5. It is designed to explain and evaluate the results of the study, showing how it relates to the literature framework in Chapter 2 and case presentation in Chapter 3.

Chapter 5's key findings suggest that there are various elements that drive the willingness to pay for data and how it influences the supply and demand graphs. The findings also highlight how to invest in data quality to improve the ROI from the data. The findings describe how to manage data to extract the greatest amount of value from it.

The quality issues of AIS data presented in Section 3 and Table 2 will have implications for what value those data possess. The quality issues addresses an important topic, as respondents commented on the data quality several times. By bundling these data with different data sets, there are costs for data cleaning and for extracting information from the AIS data. It is clear how bundling different data sets adds value for DNV by combining different data sets for use in different use cases. It is critical for its customers to pay for the services it provides. There will be a cost associated with the preparation and cleaning of the data prior to use. There are also costs associated with obtaining a license to use the data. Several of these topics were also addressed and discussed by respondents.

6.1 Drivers of customers willingness to pay

6.1.1 Drivers for Willingness to Pay

According to the findings of this study, customer willingness to pay for data quality appears to be influenced by several different elements. The respondents said that the factors were increased confidence in the data, lowered operational cost, higher expertise on data quality within the company, awareness of how good the data quality is and its potential, increased number of use cases, having a better bundle of data, having flexibility to conduct the required analysis, the brand and credibility of the firm, and the more important the data is for the use case. These findings illustrate the factors that contribute to increasing willingness to pay for data quality, and each of them is important for understanding what drives the willingness to pay. Some observations, such as Observations 1, 2, and 11 relate to making the business more efficient and effective in business operations. According to Observation 11, the more relevant the data are for a use case, the more willing people are to pay to improve the data's quality. This is because the data become more important to the firm.

The fact that clients are aware of how data quality influences the value of the information they receive is also mentioned as a driver. Observations 3, 4, 5, and 6 indicate that making a data user aware of how data quality can make the organization benefit more from the data its possesses will increase the willingness to pay. Previous research on the topic by Rea and

Sutton (2019), reveals that organizations that take time to understand the value of their data will be able to unlock substantial value, which is thereby confirmed by the findings in this study.

When the data value drivers proposed by Rea and Sutton (2019), are compared to the data quality dimensions, presented by DAMA (2017), Gupta (2021), and ISO Central Secretary (2008), the relationship between data quality and the data value drivers is evident. Gupta (2021) is more related to customer relationships but still contains the same dimensions as DAMA (2017) except that timeliness has been replaced with integrity and is therefore seen as comparable. However, it has been demonstrated that data quality is critical for the value of the data collected. As a result, it appears reasonable to say that one primary driver of increased willingness to pay for data quality is the desire to increase the value of the data itself and thereby make them more efficient and effective for business operation. Some of the findings also relates directly to the data value drivers, such as is Observations 7, 8, 9, and 10, where 8 relates to usage restriction, 7 and 10 relate to Interoperability/accessability, and 9 relates to liabilities and risk.

This section can then be summarized into some drivers that are present for customer willingness to pay for data quality. The drivers are:

- Making the business more efficient and effective in its operations
- Increasing awareness of the uncertainty and opportunity that lies in the data
- Interoperability/accessibility
- Usage restrictions
- Liabilities and risk

A summary is shown in Figure 24. A link between these drivers and the data quality dimensions in the literature review is observed. Some of the dimensions mentioned in DAMA (2017), Gupta (2021) and ISO Central Secretary (2008) may also be important in answering Research Question 1, but more research might be needed to conclude on this.

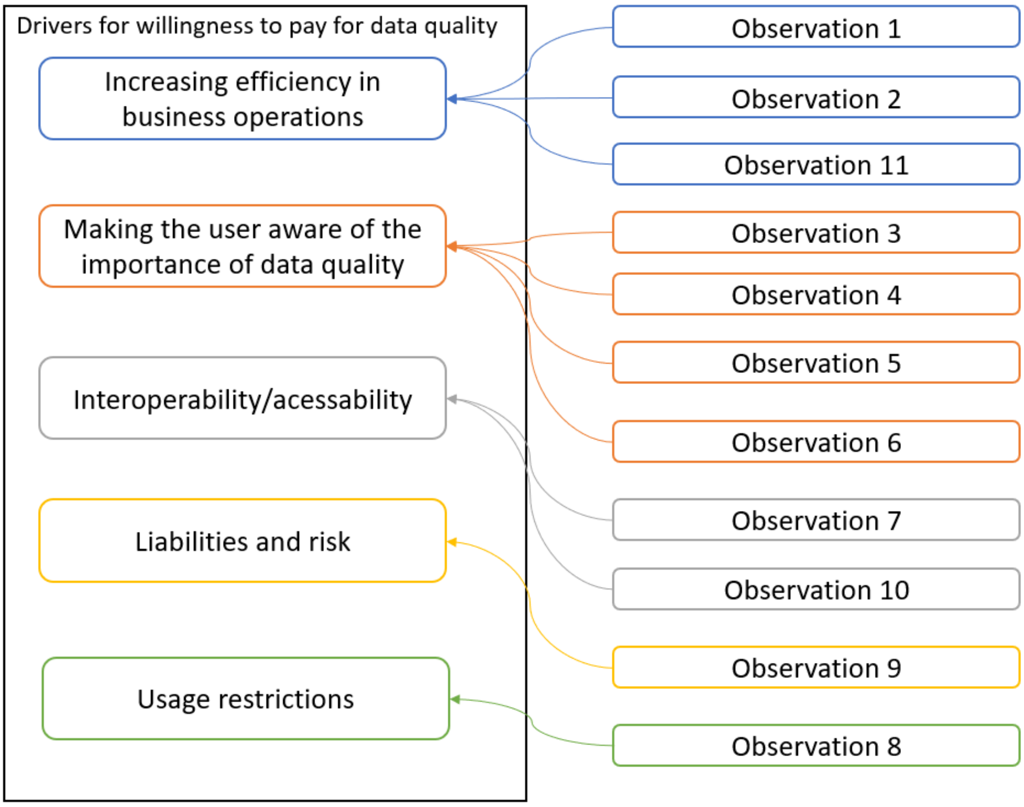


Figure 24: *The Drivers for Willingness to Pay and the Relation to the Observations.*

6.1.2 Supply and Demand

As evidenced by Observation 12, it is not common to think about supply and demand in this manner for data.

From the literature review, it can be seen that the topic on economic principles related to data is little studied. Laney (2017) appears to be at the forefront of the field, since he illustrates how the supply - and demand graph may work for data. However, we will attempt to explain how the supply and demand curves could behave, as well as the price elasticity of these curves, based on a review of the literature and the response from the informants of this study. For data, it is seen from the literature that the data value drivers could be used to increase the willingness to pay for data. It is possible to move the demand curve from D1 to D2 by investing in the value drivers, as shown in Figure 25, thereby increasing the quantity demanded.

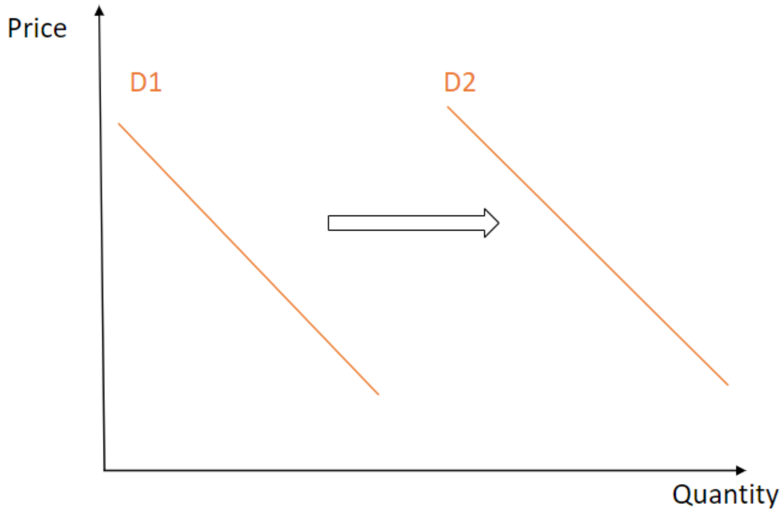


Figure 25: *Shift in Demand.*

Given that data is intrinsically reproducible and that there incurs little or no additional cost for each additional unit, the supply curve must follow a perfectly elastic supply curve capable of meeting any demand at any given price above a certain point. Perfect elasticity may seem artificial, and the price which the business would want to sell at will be determined by whether it can cover its costs of obtaining the data or increasing its quality. The pricing may be computed based on Laney (2017) financial models for valuing the data or the related framework of Rea and Sutton (2019). Graphs depicting the supply and demand are displayed in Figure 26. As shown by Laney (2017), there will be a devaluation of the data as the data will be more ubiquitous in the market, which drives the demand down and hence the price. Therefore, the concept of demand and supply seems to be more complex than demand and supply for traditional goods, and other factors need to be considered to understand it for informational assets. However, demand and supply is a naive understanding of the world, as it assumes that no relevant economic factors, other than the product's price, are changing.

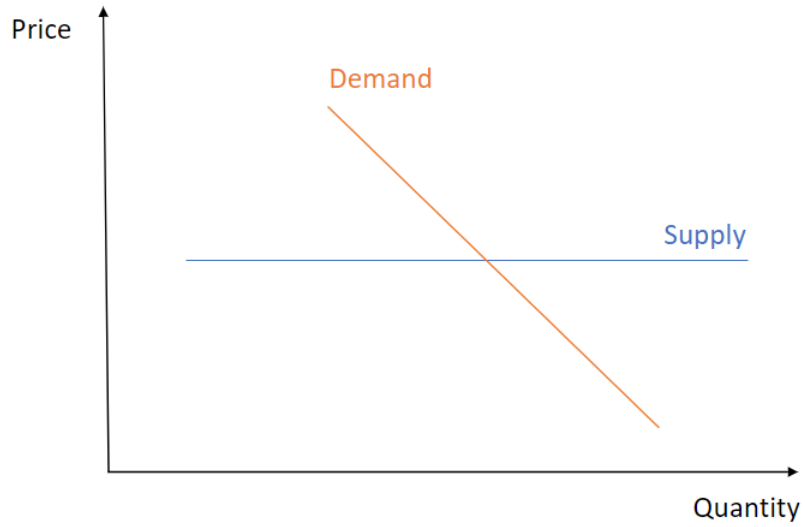


Figure 26: *The Supply and Demand Crossing for Data Assets*

The findings have implications for the problem since they provide a better understanding of what influences the change in the willingness to pay for data quality, as well as how the data requirements curve will change. Although there seems to be some understanding of the reasons for the demand curve, there seems to be little understanding of how data supply and demand behaves. Knowing how the data supply and demand curves behave seems crucial for a data-driven society and companies. This is because the trend suggests that companies are becoming increasingly dependent on data as a basis for their decisions.

6.1.3 Price-Quality Relationship

Another topic the informants mentioned, as seen in Observation 13, was how the price and quality relationship works. The answers from the respondents indicated that price and quality are not necessarily positively correlated, neither is the relationship linear. In the literature review, no studies were found that illustrated this principle for data. From traditional markets, studies have shown that the price and quality relationship does not follow a linear curve (Steenkamp, 1988). It seems reasonable to say based on the discussion with the respondents that the willingness to sell curve of quality may follow an exponential curve. This is because the efforts to increase the data quality will increase as the quality gets better, and having a DQI presented in DNV (2017) at 100% is said to be almost impossible by the informants. Therefore, we would propose that the willingness to sell curve will look like the curve in Figure 27 and describe the lowest possible price the business will charge for the data, thereby represented by the willingness to sell curve.

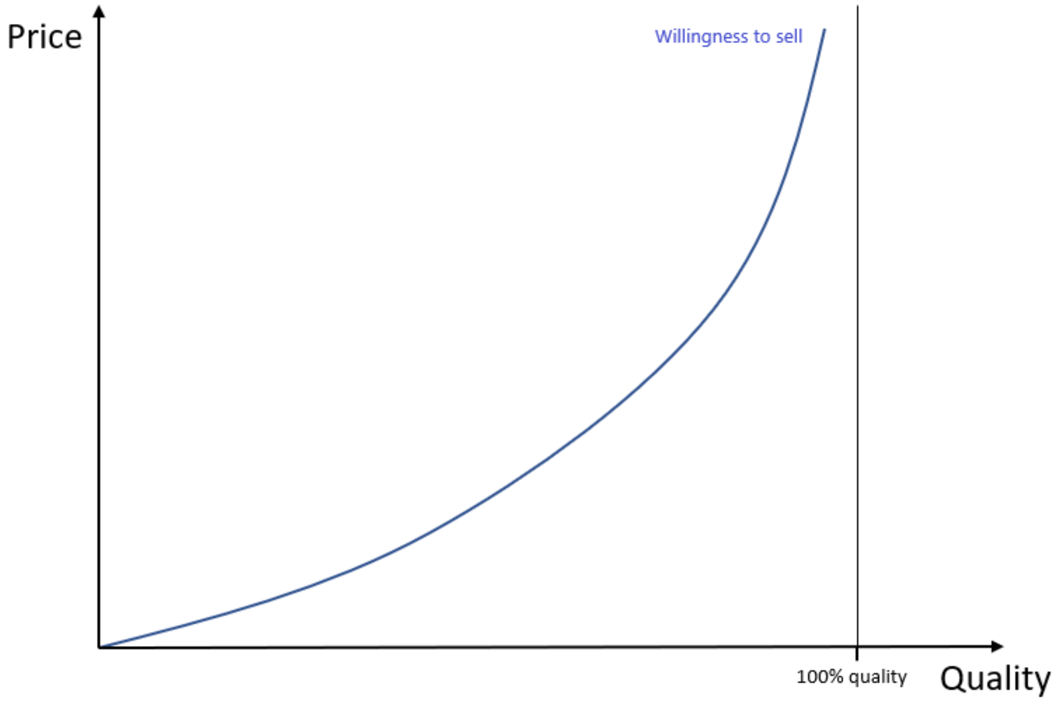


Figure 27: *The Willingness to sell curve for price given by the quality of the product.*

Due to an absence of research, predicting the willingness to pay curve in this scenario is more challenging and needs more research. The results show that the willingness to pay curve may not be linear and can therefore cross the willingness to sell curve at several points. There will be no clear saturation point like in the supply and demand curves. The willingness to pay curve is believed to depend on the data set and the specific use cases that exist for the data set. However, we can try to draw a willingness to pay curve to illustrate how it may behave. The demand is believed to increase when the quality increases but will start to level off as the quality approaches 100%. This is because there will be fewer new use cases for the data, as they are already good enough for the use cases that have arisen at lower quality levels. An overview of how the graphs look like can be seen in Figure 28. For a business selling data, it would be of interest to have the highest possible ROI, and raising the quality to the point where the difference between these graphs is greatest is desirable. Therefore, the integral between the curves will be the earnings model. It will consequently be the optimum ROI where there is greatest distance. There may be a time dependency here as well, as the business would first lose money on the service to test the service and get on board a customer base. After a while, it may want to follow this principle, as shown in Figure 28. Figures 28, 23, and 22 are all related to finding the optimal point of investments and are seen to have strong relations and are important to understand when investing in data quality.

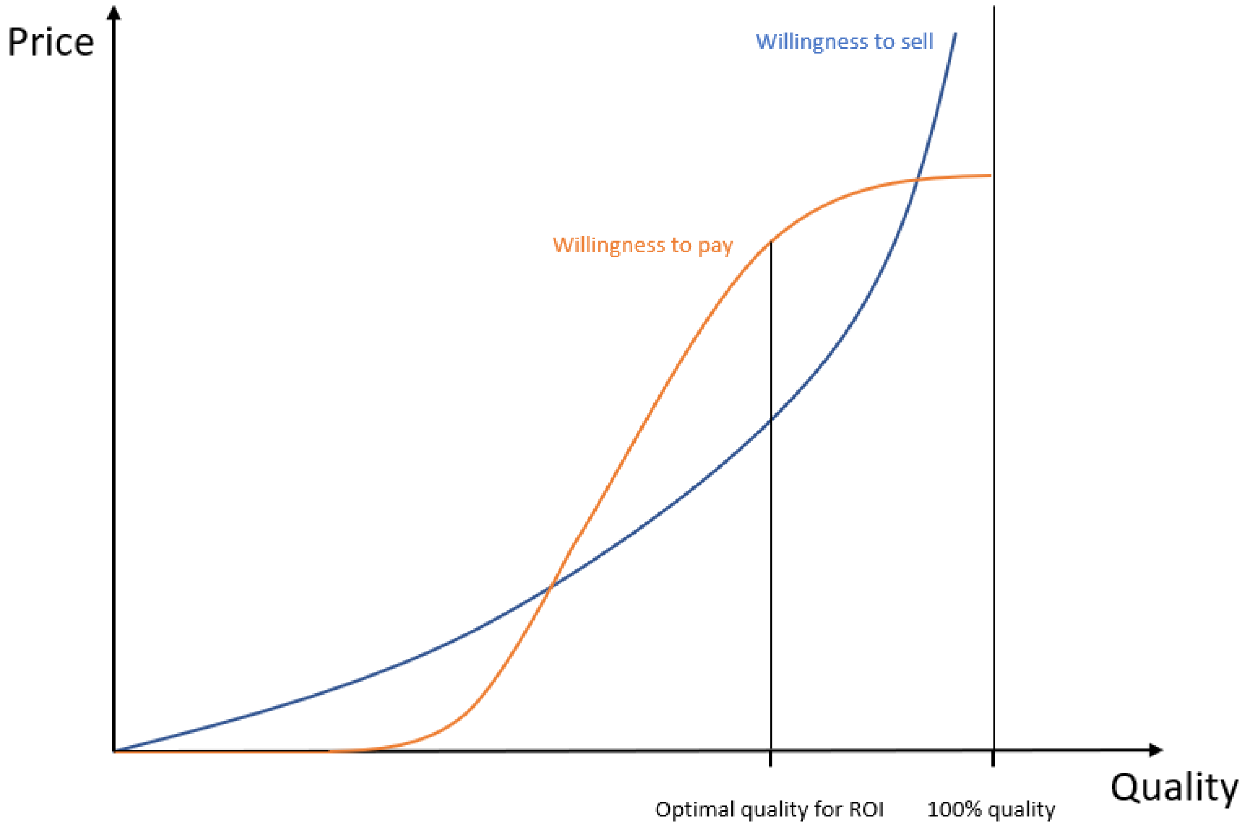


Figure 28: *The Willingness to Sell and Willingness to Pay Curves for Price vs Quality*

6.2 What Part of the Data Quality Aspects Should One Invest in to Maximize Opportunities and Minimize Risk?

6.2.1 Setting Quality Criteria

It seems reasonable that increased knowledge of the drivers for the value of data will increase the confidence of making the right decisions, which was stated in Observation 1. It is clear from both the theory and the interviews that the investment in data quality is not always straightforward. It is highly dependent on the use case of the data. It will be of great interest for a data-driven business to know how to prioritize these investment opportunities. Observation 14 in Section 5 mentioned the need to formulate and measure the quality criteria for the data. This observation is in line with the theory, which states that management professionals need to understand their customer quality requirements and how to measure them (DAMA, 2017). Setting the quality criteria, measuring them based on these criteria, and making a data quality baseline makes it easy to spot improvement opportunities and measure and monitor data quality. There are several different standard frameworks for dealing with the quality dimensions, but none of them are widely used, as seen in Section 2. Different sources have different opinions on which dimensions are important, indicating that this is an immature field. Since the literature does not give one standard framework for the quality dimensions, the organization needs to find which quality dimensions are important for its data sets and use them to measure its data quality. This finding is significant for identifying which data aspects to improve in the data. According to the existing literature on maximizing return for data investments, both Shim et al. (2015) and Pitney Bowes Software (n.d.) mention that data projects should have well-defined project metrics and performance indicators. Therefore, this finding confirms the existing theory. Measuring the data quality of the data the organization possesses is found in the literature as well. It is also seen from the literature review and the respondents that measuring based on quality dimensions is a way to measure the quality. The data quality assessment framework by DNV (2017) shows how to measure data and confirms the need for setting clear quality criteria and using data profiling to obtain the DQI.

6.2.2 Comparing Improvement Actions

The findings indicate that a cost and benefit analysis of each investment opportunity needs to be conducted. Mainly the cost and time of improvements are evaluated against the business benefits for the specific improvements. Observation 15 proposed an approach for reverse-engineering the quality dimensions to see which impact it had on the precision of a specified service. The method is illustrated in Figures 22 and 23 and gives significant contributions for answering Research Question 2. The method gives a clear and better picture of how the different data aspect improvements are compared and evaluated before investment. No such approach is found in the literature and therefore gives contributions to the topic. This approach may be of great interest for further practice and research in the industry. From the literature review, it is clear that there are not only costs of obtaining better data quality that should be included in the analysis; there are also costs of running the business on bad quality data (Haug et al., 2011; Melanie, 2019). The organization will need to identify the risks and costs associated with its data to obtain a clear picture of what value it can generate from quality improvements. Figure 8 shows how the cost and benefit of quality improvements behave.

Another observation that brings value for answering Research Question number 2 is Observation 21. The observation supports the existing theory that there may be more value to obtain from data investments in percentage terms when the company has a low maturity in data quality investments (Laney, 2017). It indicates the need for evaluating and comparing the potential cost and benefits increases as the company becomes more mature in data quality investment. For data maturity there are also several frameworks, such as DAMA (2017), DNV (2017), and ISO Central Secretary (2015), for quantifying the quality maturity of an organization. Pitney Bowes Software (n.d.) points out that the business should first focus on the more straightforward tasks and the improvement options with the most influential business impact. This finding also relates much to understanding the data and their weaknesses and how they influence the business. This concept is also raised by Deloitte (2020) but it does not have the same approach of reducing the quality of one dimension as proposed by Observation 15.

6.2.3 Involving the Right People

Observations 16 and 17 mention the need for involving the people that are hands-on with the data and having a professional environment inside the company, that can handle data management. According to the answers to this problem, this is not always the case. The people who make financial decisions do not store, capture, maintain, use or sell the data. The need to involve the correct individuals appears to be significant for obtaining the desired return from the investments. This finding may imply that existing practices in this area could be improved. Shim et al. (2015) says that it is essential that big data projects cater to all the requirements of the people using the service that is enhanced. Other than this statement, the finding was not expressed in different places in the literature.

6.2.4 Invest on a Longer Time Frame

Observation 18 states that investments in data quality should have a longer time frame. Having a longer time frame means the organization can profit from solid data quality for the present data use and in the future. Therefore, there may be conflicting perspectives on the investment's time horizon. Financial officers may be more concerned with the following quarterly results than with having data of more excellent quality, which implies having data fit for use in other future use cases. The requirement for investing on a more extended period may suggest a different approach for practice than what is now used. The literature by Deloitte (2020), which has created a framework for how to comprehend the data the business possesses, supports a focus on future use cases. A thorough understanding of the data in the organization is also mentioned by DNV (2017).

6.2.5 Invest in Data Management

Regarding Research Question 1, it was also recommended to invest in a contract that allows the company to perform the analysis it prefers. It was also emphasized in Observation 19 that the company's contracts had to be organized in a reasonable manner. It was mentioned that by arranging the data adequately so that the cost can be distributed proportionally to the different departments, the awareness of costs about data throughout the organization was increased. This is a finding described in Section 6.1. This practice was not documented explicitly in the literature, but it is a subtopic of data management and data governance. Data managers need to be aware of how they structure their data contracts within the organization and may therefore require a data maturity assessment to determine what maturity level they may be at (DNV, 2017). According to Observation 21, organizations with low data quality maturity outperform mature organizations in terms of percentage returns on their data quality investments. This means that businesses that begin to invest in data quality will quickly see a return on their investment. This confirm the information yield curve described by Laney (2017). Focusing on the maturity of the business may, thus, be significant when selecting investment methods, as low maturity organizations are more likely to have more ”low hanging fruits.”

The importance of investing in data security was also noted as a critical finding. The data quality is strongly dependent on data security, as indicated in Observation 20. There will be no data quality if the organization is attacked and the data is changed or destroyed. In data quality investments, there is a need to focus on data security. However, conflicts between data quality improvements and data security, as indicated by Talha et al. (2019), must be considered.

6.3 How can the Quality of Data be Improved, and What are the Costs?

Observation 23 revealed that, in the case of AIS data, the improvements are frequently connected to the physical repair of the transponders and receivers themselves. Physically repairing transponders and receivers is costly, both in terms of human costs on board the ships and in terms of the cost of new sensors, transponders, and receivers to replace those that have been lost or destroyed. This endeavor’s cost must be evaluated against the cost of other efforts to improve the quality of existing data. For example, the cost of experimenting with alternative data cleansing procedures as those mentioned in Observations 24 and 25 may be cheaper than the cost of ensuring the input data for AIS data. Furthermore, there are several cost drivers with varying cost magnitudes to consider throughout the data value chain proposed by Curry (2016) and further elaborated for AIS data in Section 3.2. When investing in data quality, it is critical to consider the costs associated with each cost driver. Most often, people will type it, or technology will generate it where the data was created. The expense related to repression will be prohibitively expensive, at least for AIS data.

Many errors can be made across the value chain, encompassing a wide range of diverse companies and stakeholders. The approach will be to consider data as a value chain, with roles, responsibilities, and tools in place to ensure adequate data along the way. As stated in Observationa 21 and 17, this consideration is to guarantee that the organization’s data management is mature. Organizational development, roles, and technology tools are examples of cost drivers in this case. Investing in data management maturity will have an impact on every step in the data value chain, as shown in Figure 29. This is because data management influences the management of how data can be obtained to when it is consumed (DAMA, 2017).

Cleaning the data is another option referring to Observations 24 and 25. The cleaning requirements of AIS data are determined by a dialogue with the customer, who specifies their quality requirements. Hiring personnel to develop and clean data for a model will be costly. Historical data is an example of this; these are usually difficult to recreate and as a result they may need to be cleaned before use.

When none of the previous strategies for improving data quality are viable, the last option for ensuring reliable data use is to train the user to account for the data's inherent uncertainty. As stated in Observation 3, there will be value to the user in knowing the quality of their data so that they can account for uncertainty when making decisions based on the data. The final cost will be the time spent training staff on dealing with the uncertainty produced by the data used to develop the model. When compared to the other options, this technique proved to involve the least amount of effort. Figure 29 shows a summary of the various approaches used by Curry (2016) along the data value chain.

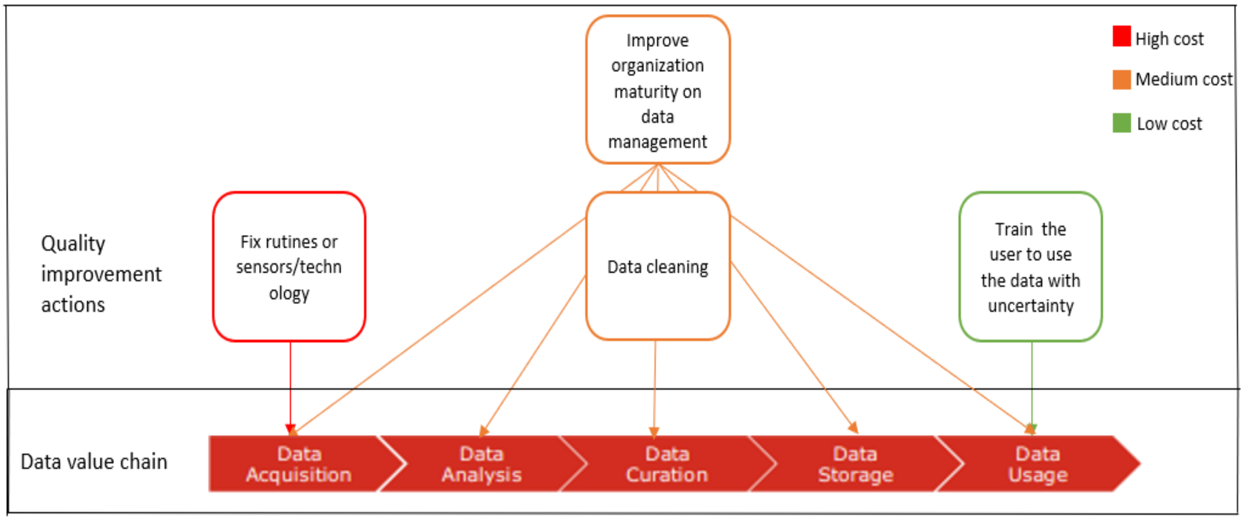


Figure 29: *The Data Value Chain (Curry, 2016) and Quality Improvement Actions*

6.4 General Discussion

The above discussion demonstrates that investing in data quality is strongly reliant on experience. To avoid over-investing, business experts familiar with the industry should review data quality investments. The absence of a response to indicators that a company is investing optimally is worth noting. As stated in Observation 22, it is based on personal experience with the company. However, there are no obvious signals to look for when investing. The findings of the literature research and the findings of this study revealed diverse viewpoints on what matters in the investing process. The concepts offered in this thesis can assist businesses in determining which parts of the value chain may require improvement and the financial implications and risk mitigation that this may entail. A company's understanding of the risk and cost of poor data quality will be critical in deciding whether or not to proceed with data quality initiatives. Market forces and stated quality criteria would be used to determine when its data is suitable for usage. Determining which data features are critical to its business would be an obvious advice from this research.

When it comes to data as a product created by one organization and consumed by another, the consumer organization depends on the maturity and quality of the creators. Some improvements in data quality can only be made where data is born. Since this is from a consumer perspective beyond its own control, the remaining cleansing mitigating actions must be sufficient.

7 Conclusion

In this concluding chapter, we present the most important themes that were discussed in this study. The themes are structured in four sets. The first theme highlights the main drivers for customers' willingness to pay for data quality. The second theme focuses on the aspects of data quality that one should invest in to maximize opportunities and minimize risks. The third theme focuses on how the quality of data is improved. Finally, in the last theme, we present recommendations for further investigation.

The findings of this study have broad implications for investments in data quality. We interviewed several highly competent individuals in the shipping industry and other fields related to data quality. However, the literature on the subject appears to be weak, and more research is needed to fill the gaps.

7.1 What are the Main Drivers for Willingness to Pay for Data Quality?

This study identifies several drivers that influence the willingness to pay for data quality. From the results of this study in Observations 1 to 11, we can conclude that the drivers for willingness to pay for data quality are:

- Increased efficiency in business operation
- Awareness of the importance of data quality
- Interoperability/accessibility
- Liabilities and risk
- Usage restrictions

The supply and demand graphs related to data were found to be less studied before. Observation 12 indicates that supply and demand is not considered in the industry. Therefore, this study has investigated the relationship between supply and demand and proposed a possible explanation to this economic principle evaluated to data as shown in Section 6.1.2. The supply curve is believed to be perfectly elastic, considering data's unique properties.

The price and quality relationship was also found as an essential result regarding Research Question 1. The literature on the relationship between price and quality for data was found to be weak, and Observation 13 indicated that this relationship is non-linear. We explain how the customer willingness to pay and the organization's willingness to pay evolve when quality is increased, as shown in Figure 28. The optimal quality for investments is the point where the distance between the two graphs is the greatest.

7.2 What Part of the Data Quality Aspects Should One Invest in to Maximize Opportunities and Minimize Risk?

Several factors are important when making investments in data quality. The study found that setting quality criteria for the data would increase the organization's confidence in making the right decisions for investments as shown in Observation 14. Observation 15 proposed a method for determining which factors influence the quality as shown in section 5.2 and Figures 22 and 23. This approach can be used to discover which aspects of data quality that should be prioritized. Furthermore, Observations 16 and 17 indicated that the organizations need to involve people who work closely with the data such as the people selling them or the chief data manager. Investments should also have a longer time frame than the next quarter and thus make the organization able to account for future use cases according to Observation 18. Finally, the study concluded that the maturity of data management affects the possible return the organization could obtain from the data referenced to Observations 19, 20 and 21. The organization will generally see quicker and higher returns from its investments if it has low maturity in data management.

7.3 How can the Quality of Data be Improved, and What are the Costs?

It is challenging to know when the company is investing optimally. The practice today is heavily dependent on experience with the business according to Observation 22. This study identifies different ways of increasing data quality as shown in Observations 23, 24, and 25 and all of them come with different costs. It is possible to increase quality by physically fixing the sensors and transponders creating and storing the data, improving the organization's maturity on data management, conducting data cleaning strategies, or training the user to use the uncertainty in data with low quality. Each of these approaches is carried out in different parts of the data value chain. The cost associated with each is divided into different price levels (high, medium, low) to say something about their magnitude, as shown in Section 6.3. The study says that the methods and costs are spread out in different parts of the data value chain. These methods are essential to understand how to invest in data quality.

7.4 Recommendations

Some of the findings of this study confirm findings from different researchers. However, the results also contribute to the literature and have relevance for contemporary practice. Furthermore, since the study is based on a small sample size and a limited use case for the data, further research would be required to strengthen the findings of this study.

We discovered that the economic principles such as supply and demand and the price-quality relationship evaluated to data needed more research. Therefore, it is recommended to conduct studies that can elaborate on the topic for a broader range of data sets. In addition, conducting studies that can quantify these principles would be necessary for a deeper understanding of the principles. Studies on which quality factors are most important (valuable to buyers) and how value correlates with changes in various quality factors would be a reasonable extension of the topic.

Further research should focus on conducting studies similar to this one evaluated on other data sets and other use cases than the focus of this research. It may also be required to supplement and further identify the demand drivers for data quality as other significant drivers may be driving the demand.

Research on investment strategies and methods for identifying the aspects one should invest in should be extended and tested in the industry and on several data sets to see the correlations. In addition, studies can be conducted to investigate whether this method is helpful in other data sets.

Further clarification on the cost of investments in data quality may also be a point of interest in future studies. For example, there may be ways to study how the different quality improvement actions influence the benefit of the data and how the investment in them should then be performed.

References

Agarwal, P. (2018). *Supply and demand*. https://www.intelligenteconomist.com/supply-and-demand/ (Accessed: 26.03.2021)

Austeng, K., Midtbø, J. T., Jordanger, I., Magnussen, O. M., & Torp, O. (2005). *Usikkerhet-sanalyse - kontekst og grunnlag*. Norges teknisk- naturvitenskapelige universitet.

Ayiro, L. P. (2012). *A functional approach to educational research methods and statistics : Qualitative, quantitative, and mixed methods approaches*. Lewiston, N.Y. : Edwin Mellen Press, 2012.

Balou, D. P., & Tayi, G. K. (1996). Managerial issues in data quality. http://mitiq.mit.edu

Belbin., R. M. (1981). *Management teams: Why they succeed or fail*. Heinemann.

Bowen, G. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*, 27–40. https://doi.org/10.3316/QRJ0902027

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*, 77–101. https://doi.org/10.1191/1478088706qp063oa

Breidert, C., Reutterer, T., & Hahsler, M. (2006). A review of methods for measuring willingness to pay.

Brinkmann, S., & Kvale, S. (2005). Confronting the ethics of qualitative research. *Journal of Constructivist Psychology*, *18*, 157–181. https://doi.org/10.1080/10720530590914789

Brookshier, K. (2018). *Method vs. methodology: Understanding the difference*. https://uxdesign.cc/method-vs-methodology-whats-the-difference-9cc755c2e69d

Bryman, A. (2012). *Social research methods*. Oxford University Press.

Campbell, P. (2021). Willingness To Pay: What Is It and How to Calculate Customer WTP. Retrieved August 28, 2021, from https://www.priceintelligently.com/blog/willingness-to-pay

Caulfield, J. (2019). *How to do thematic analysis*. https://www.scribbr.com/methodology/thematic-analysis/

Caulfield, J. R. (2019). *How to do thematic analysis*. https://www.scribbr.com/methodology/thematic-analysis/ (Accessed: 27.07.2021)

Choudhury, A. (n.d.). *Research design: 6 things to know about research design*. https://www.yourarticlelibrary.com/social-research/research-design/research-design-6-things-to-know-about-research-design/64496

Cooper, D., & Schindler, P. (2008). *Business Research Methods*. McGraw-Hill. https://books.google.no/books?id=AAR-PwAACAAJ

Costa, N., Svanberg, M., & Hörteborn, A. (2020). The use and usefulness of ais data. https://www.sspa.se/the-use-and-usefulness-of-AIS-data

Coventry University. (n.d.). *The research process*. https://www.futurelearn.com/info/courses/research-process/0/steps/71889

Creswell, J. W. (2008). Educational research: Planning, conducting, and evaluating quantitative and qualitative research (3rd ed.)

Crossman, A. (2019). *What is a snowball sample in sociology?* https://www.thoughtco.com/snowball-sampling-3026730 (Accessed: 23.05.2021)

Curry, E. (2016). *New horizons for a data-driven economy*. Springer Open.

DAMA. (2017). *Dama-dmbok: Data management body of knowledge (2nd edition)*. Technics Publications, LLC.

DeCarlo, M. (2018). *Scientific inquiry in social work*. Open Social Work Education.

Deloitte. (2020). Data valuation: Understanding the value of your data assets. https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Finance/Valuation-Data-Digital.pdf

Det kongelige kommunal- og moderniseringsdepartementet. (2021). Meld. st. 22 - data som ressurs. https://www.regjeringen.no/no/dokumenter/meld.-st.-22-20202021/id2841118/

Diane Coyle and Stephanie Diepeveen and Julia Wdowin and Jeni Tennison and Lawrence Kay. (2020). The value of data. https://www.bennettinstitute.cam.ac.uk/media/uploads/files/Value_of_data_summary_report_26_Feb.pdf

Digital Norway. (2021). *Mener for få forstår verdien av data: faktisk ganske kritisk*. https://digitalnorway.com/mener-for-fa-forstar-verdien-av-data-faktisk-ganske-kritisk/ (Accessed: 12.06.2021)

Digitaliseringsdirektoratet. (2021). *Hva er risiko?* https://internkontroll-infosikkerhet.difi.no/risikostyring/hva-er-risiko (Accessed: 23.05.2021)

DNV. (2017). Data quality assessment framework.

Eisenhardt, K. M., & Graebner, M. E. (2007). *Theory building from cases: Opportunities and challenges*. https://doi.org/10.5465/amj.2007.24160888 (Accessed: 07.08.2021)

Ellis, T., & Levy, Y. (2008). Framework of problem-based research: A guide for novice researchers on the development of a research-worthy problem. *Informing Science: The International Journal of an Emerging Transdiscipline*, *11*. https://doi.org/10.28945/438

Experian information solutions. (2014). Meld. st. 22 - data som ressurs.

Fernando, J., & Mansa, J. (2020). *Return on investment (roi)*. https://www.investopedia.com/terms/r/returnoninvestment.asp

Ferreira, C., Andrade, P., & Almeida, F. (2020). How to improve the validity and reliability of a case study approach. *Journal of Interdisciplinary Studies in Education*, *9*, 273–284. https://doi.org/10.32674/jise.v9i2.2026

Gard. (2019). *going dark' is a red flag – ais tracking and sanctions compliance*. https://www.gard.no/web/updates/content/27716479/going-dark-is-a-red-flag-ais-tracking-and-sanctions-compliance (accessed: 20.05.2021)

GoCardless. (n.d.). *What does opex (operating expenses) mean for my business?* https://gocardless.com/en-us/guides/posts/opex-operating-expenses/

Grier, K. (2019). *Three ways you can turn uncertainty into opportunity.* https://www.ey.com/en_gl/wef/three-ways-you-can-turn-uncertainty-into-opportunity (Accessed: 12.06.2021)

Grønmo, S. (2020). *Case-studie.* https://snl.no/case-studie

Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, *26*(3). https://doi.org/https://doi.org/10.1016/j.jsis.2017.07.003

Gupta, A. (2021). *The 6 dimensions of data quality.* https://www.collibra.com/blog/the-6-dimensions-of-data-quality (Last updated: 06.04.2021)

Haug, A., Zachariassen, F., & Liempd, D. (2011). The costs of poor data quality. *Journal of Industrial Engineering and Management*, *4*. https://doi.org/10.3926/jiem..v4n2.p168-193

Hayes, A. (n.d.). *Law of demand.* https://www.investopedia.com/terms/l/lawofdemand.asp (Accessed: 12.06.2021)

Hoepfl, M. C. (1997). Choosing qualitative research: A primer for technology education researchers.

Hoffmann, F., van Hettinga, E., & Nowak, M. (2020). Price elasticity – the way to optimal prices. https://7learnings.com/blog/price-elasticity/ (Accessed: 01.08.2021)

Husby, O. (1999). *Usikkerhet som gevinst : Styring av usikkerhet i prosjekter : Mulighet - risiko, beslutning, handling.* Norsk senter for prosjektledelse. https://doi.org/https://urn.nb.no/URN:NBN:no-nb_digibok_2018042048113

IALA. (2016). Guideline no. 1082 on an overview of ais. https://www.navcen.uscg.gov/pdf/IALA_Guideline_1082_An_Overview_of_AIS.pdf

IAS. (2004). *Ias 38 — intangible assets.* https://www.iasplus.com/en/standards/ias/ias38 (accessed: 27.03.2021)

IBM. (2016). *Data quality score.* https://www.ibm.com/docs/en/iis/11.5?topic=results-data-quality-score (Accessed: 27.07.2021)

iedunote. (n.d.). *Research process: 8 steps in research process.* https://www.iedunote.com/research-process (Accessed: 07.07.2021)

International Maritime Organization. (2019). *Ais transponders.* https://www.imo.org/en/OurWork/Safety/Pages/AIS.aspx

Investopedia. (2020). *Introduction to supply and demand.* https://www.investopedia.com/articles/economics/11/intro-supply-demand.asp (accessed: 26.03.2021)

Investopedia. (2021). *Whatsapp: The best facebook purchase ever?* https://www.investopedia.com/articles/investing/032515/whatsapp-best-facebook-purchase-ever.asp (accessed: 26.03.2021)

ISO Central Secretary. (2015). *Data quality — information and data quality: Concepts and measuring* (Standard ISO 8000-8:2015(E)). International Organization for Standardization. Geneva, CH.

ISO Central Secretary. (2008). *Software engineering — software product quality requirements and evaluation (square) — data quality model* (Standard ISO/IEC 25012:2008). International Organization for Standardization. Geneva, CH.

Jacobsen, D. I. (2005a). *Hvordan gjennomføre undersøkelser?* Cappelen Damm akademisk.

Jacobsen, D. I. (2005b). *Hvordan gjennomføre undersøkelser?* https://www.uio.no/studier/emner/jus/afin/FINF4002/v12/Metode1.pdf

Jankowicz, D. (2002). Research methods for business and management. *Einsburg Business School. Watt University*. https://ebs.online.hw.ac.uk/documents/course-tasters/english/pdf/h17bm-bk-taster.pdf

Jansen, D., & Warren, K. (2020). *What is research methodology?* https://gradcoach.com/what-is-research-methodology/

Jiang, M. (2020). *The reason zoom calls drain your energy*. https://www.bbc.com/worklife/article/20200421-why-zoom-video-chats-are-so-exhausting (Published: 22.04.2020)

Johannessen, A., Christoffersen, L., & Tufte, P. A. (2011). *Forskningsmetode for økonomisk administrative fag*. Oslo: Abstrakt Forlag.

Johannessen, A., Tufte, P. A., & Christoffersen, L. (2010). *Introduksjon til samfunnsvitenskapelig metode*. Oslo:Abstrakt, 2010.

Kahn, B., Strong, D., & Wang, R. (1997). A model for delivering quality information as product and service., 80–94.

Kenton, S. (n.d.). *Demand curve*. https://www.investopedia.com/terms/d/demand-curve.asp (Accessed: 12.06.2021)

Kjertsad, N. (2019). *Ais*. Store norske leksikon. https://snl.no/AIS

KPMG. (2019). Data as an asset. https://home.kpmg/xx/en/home/insights/2019/10/data-as-an-asset.html

Krumsvik, R. J. (2014). *Forskingsdesign og kvalitativ metode: Ei innføring*. Vigmostad Bjørke AS.

Kudva, S. (2020). *Cost of poor data quality*. https://www.linkedin.com/pulse/cost-poor-data-quality-santosh-kudva/ (Accessed: 12.06.2021)

Kvale, S. (1997). *Det kvalitative forskningsintervju*. Oslo: Ad Notam Gyldendal.

Kystverket. (2021). *Ais - automatisk identifikasjonssystem*. https://www.kystverket.no/AIS

Kystverket. (n.d.). *Ais norway*. https://www.kystverket.no/en/navigation-and-monitoring/ais/ais-norge/

Laney, D. B. (2017). *Infonomics*. Bibliomotion Inc.

Laney, D. B. (2020). Your Company's Data May Be Worth More Than Your Company [Section: CIO Network]. Retrieved August 28, 2021, from https://www.forbes.com/

sites/douglaslaney/2020/07/22/your-companys-data-may-be-worth-more-than-your-company/

Lekanger, K. (2019). *Data brukes i stadig mer avanserte it-løsninger, men det er urovekkende liten fokus på kvaliteten.* https://www.digi.no/artikler/data-brukes-i-stadig-mer-avanserte-it-losninger-men-det-er-urovekkende-liten-fokus-pa-kvaliteten/456586 (Accessed: 12.06.2021)

Lexico. (2021). *Quality.* https://www.lexico.com/definition/quality

Lexico. (n.d.). *Research.* https://www.lexico.com/en/definition/research

Li, W. C., Makoto, N., & Kazufumi, Y. (2019). *Value of Data: There's No Such Thing as a Free Lunch in the Digital Economy* (Discussion papers No. 19022). Research Institute of Economy, Trade and Industry (RIETI). https://ideas.repec.org/p/eti/dpaper/19022.html

Malt, U., & Grønmo, S. (2020). *Strukturert intervju.* https://snl.no/strukturert_intervju

Mantell, C., Benson, R., Stopford, M., Crowe, T., & Gordon, S. (2014). Shipping intelligence weekly clarkson research services limited.

Manyika, J., Chui, M., Farrell, D., Kuiken, S. V., Groves, P., & Doshi, E. A. (2013). *Open data: Unlocking innovation and performance with liquid information.* https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information

Mar, A. (2013). *7 definitions of quality.* https://business.simplicable.com/business/new/7-definitions-of-quality (Last updated: 14.03.2013)

Maxwell, J. (2013). *Qualitative Research Design: An Interactive Approach: An Interactive Approach.* SAGE Publications.

Maxwell, J. A. (2012). *Qualitative research design.* SAGE Publications Inc.

McCombes., S. (2021). *How to create a research design.* https://www.scribbr.com/research-process/research-design/?fbclid=IwAR1hrwh1RlsZ0pnrweISKwAHUTfLyhPd4sV64kkiooBugfVcndd (Accessed: 12.06.2021)

Melanie. (2019). *5 major consequences of poor-quality data and how to avoid it.* https://www.unleashedsoftware.com/blog/5-major-consequences-poor-quality-data-how-avoid (Accessed: 12.06.2021)

Mestl, T. (2020). Ais 101. [PowerPoint presentation].

Mestl, T., & Ruud, S. (1999). Information quality exchange model.

Middleton, F. (2019). *Reliability vs validity: What's the difference?* https://www.scribbr.com/methodology/reliability-vs-validity/

Moore, J. (2019). *How to make the right data-driven decisions.* https://blog.stratasan.com/data-driven-decisions-dikw-pyramid

Moore, S. (2018). *Follow these 5 steps to effectively design a compelling data quality improvement business case.* https://www.gartner.com/smarterwithgartner/how-to-create-a-business-case-for-data-quality-improvement/ (accessed: 26.03.2021)

Oceana. (2018). *Avoiding detection: Global case studies of possible ais avoidance.* https://oceana.org/publications/reports/avoiding-detection-global-case-studies-possible-ais-avoidance (accessed: 20.05.2021)

Ontotext. (n.d.). *Knowledge pyramid.* https://www.ontotext.com/knowledgehub/fundamentals/dikw-pyramid/ (accessed: 27.07.2021)

Patroumpas, K., Alevizos, E., Artikis, A., Vodas, M., Pelekis, N., & Theodoridis, Y. (2017). Online event recognition from moving vessel trajectories. *GeoInformatica*, *21*. https://doi.org/10.1007/s10707-016-0266-x

Patton, M. (1999). *Enhancing the quality and credibility of qualitative analysis.* Health Sciences Research.

Peredaryenko, M. S., & Krauss, S. (2013). Calibrating the human instrument: Understanding the interviewing experience of novice qualitative researchers. *The Qualitative Report*, *18*, 1–17.

Pitney Bowes Software. (n.d.). The roi of data quality. https://www.pb.com/docs/US/PDF/software/customer-information-management/wp-roi-data-quality-93734-amer-1401.pdf

Price, R., Shanks, G., Burstein, F., & Holsapple, C. (2008). Data quality and decision making. https://doi.org/10.1007/978-3-540-48713-5_4

Prosjekt Norge. (n.d.). *Usikkerhet, risiko og muligheter.* http://v1.prosjektnorge.no/index.php?subsite=pus&pageId=430 (Accessed: 12.06.2021)

PWC. (2019). Creating value from data. *Strategy*, *9*(1), 1–9. https://doi.org/https://www.strategyand.pwc.com/gx/en/insights/2019/creating-value-from-data/creating-value-from-data.pdf

Rea, N., & Sutton, A. (2019). Putting a value on data. https://www.pwc.co.uk/data-analytics/documents/putting-value-on-data.pdf

Redman, T. (2008). *Data driven: Profiting from your most important business asset.*

Rolstadås, A. (2020). *Usikkerhet - (prosjektledelse).* https://snl.no/usikkerhet_-_prosjektledelse

Ross, S. (2021). *How does price elasticity affect supply?* https://www.investopedia.com/ask/answers/040615/how-does-price-elasticity-affect-supply.asp (Accessed: 09.06.2021)

Rugg, D. (2010). An introduction to triangulation. *United Nations Programme on HIV/AIDS: Switzerland.* https://www.unaids.org/sites/default/files/sub_landing/files/10_4-Intro-to-triangulation-MEF.pdf

Rummel, J. F. (1963). *Research methodology in business.* Harper Row.

Sander, K. (2020). *Forskningsdesign.* https://estudie.no/hva-er-forskningsdesign/

Saunders, M., Thornhill, A., & Lewis, P. (2016). *Research methods for business students.* Harlow : Pearson Education.

Saxena, A. (2019). *What is data value and should it be viewed as a corporate asset?* https://www.dataversity.net/what-is-data-value-and-should-it-be-viewed-as-a-corporate-asset/ (Accessed: 12.06.2021)

Shanks, G., & Darke, P. (1998). Understanding data quality in a data warehouse. *Aust. Comput. J.*, *30*.

Shen, J. (n.d.). *Risk*. https://www.investopedia.com/terms/r/risk.asp (Accessed: 12.06.2021)

Shim, J., French, A., Guo, C., & Jablonski, J. (2015). Big data and analytics: Issues, solutions, and roi. *Communications of the Association for Information Systems*, *Vol. 36*. https://doi.org/10.17705/1CAIS.03739

Short, J. E. (2017). *What's your data worth?* https://sloanreview.mit.edu/article/whats-your-data-worth/ (accessed: 26.03.2021)

Skattedirektoratet. (n.d.). *Nox tax*. https://www.skatteetaten.no/en/business-and-organisation/vat-and-duties/excise-duties/about-the-excise-duties/nox/ (accessed: 26.03.2021)

Smestad, B. B., Asbjørnslett, B. E., & Rødseth, Ø. J. (2015). *A study of satellite ais data and the global ship traffic through the singapore strait* (Doctoral dissertation). https://doi.org/10.13140/RG.2.2.15529.90726

Spacey, J. (2015). *What is quality risk?* https://simplicable.com/new/quality-risk (Accessed: 12.06.2021)

Steenkamp, J.-B. (1988). The relationship between price and quality in the marketplace. *De Economist*, *136*. https://doi.org/10.1007/BF01803598

Steinar Kvale, S. B. (2019). *Det kvalitative forskningsintervju*. Gyldendal Akademisk.

Stepanyan, L. (2020). *Body language in times of zoom*. https://lucystepanyan.medium.com/body-language-in-times-of-zoom-fe0807b736f (Published: 11.09.2020)

Svartdal, F. (2019). *Bekreftelsestendens*. https://snl.no/bekreftelsestendens (Accessed: 12.06.2021)

SVB. (n.d.). *Everything there is to know about ais*. https://www.svb24.com/en/guide/ais.html (accessed: 19.03.2021)

Talha, M., Kalam, A., & Elmarzouqi, N. (2019). Big data: Trade-off between data quality and data security. *Procedia Computer Science*, *151*, 916–922. https://doi.org/10.1016/j.procs.2019.04.127

Tambe, P. (2014). *Big data investment, skills, and firm value*. INFORMS.

Tata Consultancy Services. (2013). The emerging big returns on big data.

Tjora, A. (2012). *Kvalitative forskningsmetoder i praksis*. Oslo: Gyldendal akademisk.

Tjora, A. (2015). Kvalitative forskningsmetoder i praksis.

Trillium Software. (2010). Building a tangible roi for data quality. https://www.iconresources.com/Icon/assets/downloads/30001%20Building%20a%20Tangible%20ROI%20for%20Data%20Quality%202010.pdf

University, A. (2018). *Literature review tutorial: Why do a lit review?* https://subjectguides.library.american.edu/c.php?g=175218&p=1154280

University, C. (n.d.). Research instrument examples. https://www.tc.columbia.edu/media/administration/institutional-review-board-/irb-submission---documents/Published_Study-Material-Examples.pdf

Varian, H. R. (1992). *Microeconomic analysis, vol. 3*. W.W. Norton.

Visma. (n.d.). *Priselastisitet*. https://www.visma.no/eaccounting/regnskapsordbok/p/priselastisitet/ (Accessed: 08.06.2021)

Wang, R. Y., Reddy, M. P., & Kon, H. B. (1995). Toward quality data: An attribute-based approach. *Decis. Support Syst.*, *13*, 349–372.

White, A. (2020). *Methods for valuing data*. https://blogs.gartner.com/andrew_white/2020/03/09/methods-valuing-data/

Widding, L. Ø. (2005). *Case som metode : Hovedutfordringer knyttet til ulike forskningsdesign når hensikten er å generalisere*. Bodø : Handelshøgskolen i Bodø.

Winkler, D. (2012). *Ais data quality and the authoritative vessel identification service (avis)*. https://navcen.uscg.gov/

Wold, T. (n.d.). Tematisk analyse.

Worthy, B. (n.d.). *Filler words in legal transcription: Why they should be included*. https://www.gmrtranscription.com/blog/filler-words-in-legal-transcription (Published: 21.07.2016)

Yin, R. K. (2014). *Case study research design and methods (5th ed.)* SAGE Publications Inc.

Zhang, T., Dr Wang, W. Y. C., & Pauleen, D. (2017). Big data investments in knowledge and non-knowledge intensive firms: What the market tells us. *Journal of Knowledge Management*, *21*. https://doi.org/10.1108/JKM-12-2016-0522

Zola, A. (n.d.). *3 factors to consider when calculating roi for your data analytics project*. https://intersog.com/blog/3-factors-to-consider-when-calculating-roi-for-your-data-analytics-project/x (accessed: 26.03.2021)

Appendices

A Interview Guide

Opening:

What is your role in the company?

How long have you been involved with AIS data?

Main part:

Research question 1:

What are the main drivers for willingness to pay for data quality?

What is the motivation to increase the data quality?

Explain how the potential profit changes according to the data quality/willingness to pay?

Why is it essential to find out when the data is good enough for data science?

Research question nr 2:

What part of the data quality aspects should one invest in to maximize opportunities and minimize risk?

How do you know that the data quality is good enough?

You can invest in data quality in many different ways, but how do you decide what to invest in?

What is the risk of using the quality of raw data instead to increase it?

How can the contract with the data provider influence the value of the data?

How can you know you have invested enough in data quality?

Research question nr 3:

How can the quality of data be improved, and what are the costs?

What are the criteria to improve quality?

How do you decide how you should clean the data?

Which type of data cleaning do you perform, and which has proven to be successful?

Ending:

Is there anything you want to add?

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway