DATA QUALITY & KNOWLEDGE CREATION

A QUALITATIVE STUDY OF DATA-USAGE IN HEALTHCARE

MASTER THESIS

ARIA HADAD: 93915

BUSINESS ADMINISTRATION & INFORMATION SYSTEMS
COPENHAGEN BUSINESS SCHOOL 2019
SUPERVISOR: WEIFANG WU

NUMBER OF CHARACTERS: 140.751

"It takes discernment to do this. Most problems don't require more data. They require more insight, more innovation and better eyes. Information is what we call it when a human being takes data and turns it into a useful truth."

- Seth Godin

ABSTRACT

This thesis explores the data quality and information transformation in the healthcare company Roche Diabetes Care. Through a case study the purpose of this thesis is to understand how data is transformed into meaningful information and how knowledge creation is positively influenced. The theoretical framework of data quality by Wang & Strong (1996) is used to measure the data usage and find patterns in the data-usage process. Furthermore, the DIKW pyramid by Ackoff is used to explain the transformation process from data to information to knowledge, and the linking between these elements.

In considering the data usage, four major findings were found to affect the data quality; change, mistake, poor structures and system design. These four incidents are contributing to the data not being properly employed by employees. Therefore, in order for Roche Diabetes Care to improve data quality and make data consumers' use the data they need to 1. Regularly provide feedback to the data by actively using it. 2. Invest in tools and processes to ensure accuracy. 3. Provide clear and definable data-usage structures to employees. 4. Improve the flexibility and design of the system.

Furthermore, the system is the middleman of transforming the data into information, and in order for Roche Diabetes Care to turn the data collected into meaningful information where knowledge creation is positively influenced, it has to encourage employees to share data, information and knowledge among one another. To increase knowledge creation the company should invest resources in tools and procedures that enhances knowledge sharing. This thesis argues that one of the most effective way to increase knowledge creation is to enhance knowledge sharing.

Keywords: Data Quality, Healthcare, Knowledge Management & Knowledge Creation.

ACKNOWLEDGEMENTS

I would like to send a huge and special thanks to my supervisor Weifang Wu, for competent feedback, for always being available and for supporting me throughout the whole process of writing this thesis.

I would also give a big thanks to the employees from Roche Diabetes Care, who provided me with their time to interview them. A special thanks goes to Rikke Fog, who allowed me to do the research in the company. Another big thanks go to Anna Madsen, who supported me understanding the company structures and functions from a compliant perspective. Without participation and support, I would not have been able to conduct this thesis.

Lastly, I would like to thank my family and friends for supporting and motivating me throughout the last six months. Thank you!

// Aria Hadad

Table of Contents

1. INTRODUCTION

I was on my first vacation in Berlin with my younger sister. We had just arrived at our hotel and ready to check in. We had planned to go visit Brandenburg Tor as soon as we had checked in and left our luggage in the room… preferably before it went dark outside. “We have a reservation and would like to check in, I said politely.” The women on the other side of the desk asked me for a name. “Aria, I said.” She used several minutes searching in the database and returned with a: “I am sorry, but I cannot find any reservation for Aria… But no worries, we do have rooms available.” I couldn’t understand as I had already paid for the room and showed her the booking confirmation sent to my e-mail. She used an additionally fifty minutes searching in the database and returned with a “well, there might have happened a mistake. But I think I have found your reservation. Is your surname Bakker?” She asked. No, it’s Hadad.” I answered. She continued looking in the screen to figure out how to move further. A forty-five minutes later we had a room but were exhausted, and as we had exceeded our time schedule, we decided to not go visit Brandenburg Tor and instead went to buy some dinner and then go to sleep.

1.1 Problem Statement

The incident described above kept made me think of the importance of data quality and why an organization should focus seriously in managing high-quality data as well as making the data readable and understandable for data consumers. High quality data in itself is not always enough, but the structures and representation of it plays a significant role on its usage too. In order for the data to provide value and support data consumers’ tasks at a planned time, it not only has to be accurate and timely, but also being represented in a clear, consistent and understandable format, easy for data consumers to interpret, comprehend and make use of it. For this to happen, it relies on the system’s structures and design. As reported in the Wall Street Journal, “Thanks to computers, huge databases brimming with information at our fingertips, just waiting to be tapped…. Just one problem: Those huge databases may be full of junk.” Poor data quality can have a severe impact on the effectiveness of an organization, however, high data quality can be useless if not represented in consistent and comprehensible manner, that is easy for data consumers to understand and use. Therefore, the problem of data quality is not considered to be only the data itself, but also the way it is represented by the system.

1.2 Literature Summary

Larry English (1999) describes the data-information relationship as information being data in context, with knowledge being information in context, knowing the meaning of the information. Similarly, Tayi & Ballou (1998) describes data as "raw material for the information age", meaning data support managerial and professional work and are critical for all levels of an enterprise. Data quality is in the literature defined as "fitness for use" and presented as a multidimensional concept. Often discussed dimensions include accuracy, completeness, consistency and timeliness. These dimensions are used to measure data quality and the choice of the dimensions is primary based on intuitive understanding, industrial experience or literature reviews. (Wand & Wang, 1996). However, literature reviews also show that there is no overall agreement on data quality dimensions. While Ken Orr, 1998, Strong, Wang & Lee, 1997 & Pipino, 2004, talk about change as an incident that decrease data quality in organizations, Agarwal & Yiliyasi, 2015 discuss the factor to be information overload. Strong, Wang & Lee, 1997, suggest that high quality data is a result of flexible systems, and proposes that the notion of data or information quality depends on the actual use of data, and what may be considered good data in one case, may not be sufficient in another case, therefore they propose that the quality of the data generated by an information system depends on the design and flexibility of the system. (p. 87).

1.3 Research Question

This thesis will be exploring the data quality in the global healthcare organization, Roche Diabetes Care, and aims to investigate the data usage from a data consumer-perspective. The paper will be studying the data quality degree and consider the prospective of data that is collected and aggregated in the organization. To provide the company with suggestions to utilize their data more efficiently it is needed to investigate their data usage, thus measure the data quality. In order to be able to investigate the aforementioned problem the following question is formed:

"How can Roche Diabetes Care turn the data that is collected into meaningful information where knowledge creation is positively influenced and managed in the company?"

As mentioned above, data quality and knowledge creation are seen as an important influence on decision-making, planning, execution and profit. By conducting an extensive literature review it was found that not much attention has been paid towards data quality in healthcare systems. Therefore, the overall aim of this project is to make an analytical- and empirical contribution to the existing data quality literature by conducting a qualitative research. Through the investigation of the employees of Roche Diabetes Care, it is aimed to understand how and what data is collected, used and maintained, and what challenges are related to the data utilization processes. As data quality in healthcare is a relative unexplored area, it is hoped to be able to make and empirical contribution concerning the relevant topic.

1.4 Theoretical Framework

The theoretical frameworks used in this thesis are Wang & Strong's data quality framework and the DIKW pyramid developed by Ackoff. Wang & Strong (1996) developed four categories of dimensions, such as, intrinsic, contextual, representational and accessibility, which data quality can be measured by. Additionally, the DIKW pyramid is intended to explain the transformation of data to information to knowledge. While Wang & Strong's framework is used to measure and explore the data-information issues, the DIKW model attempts to look more into the findings and deduce answers to those.

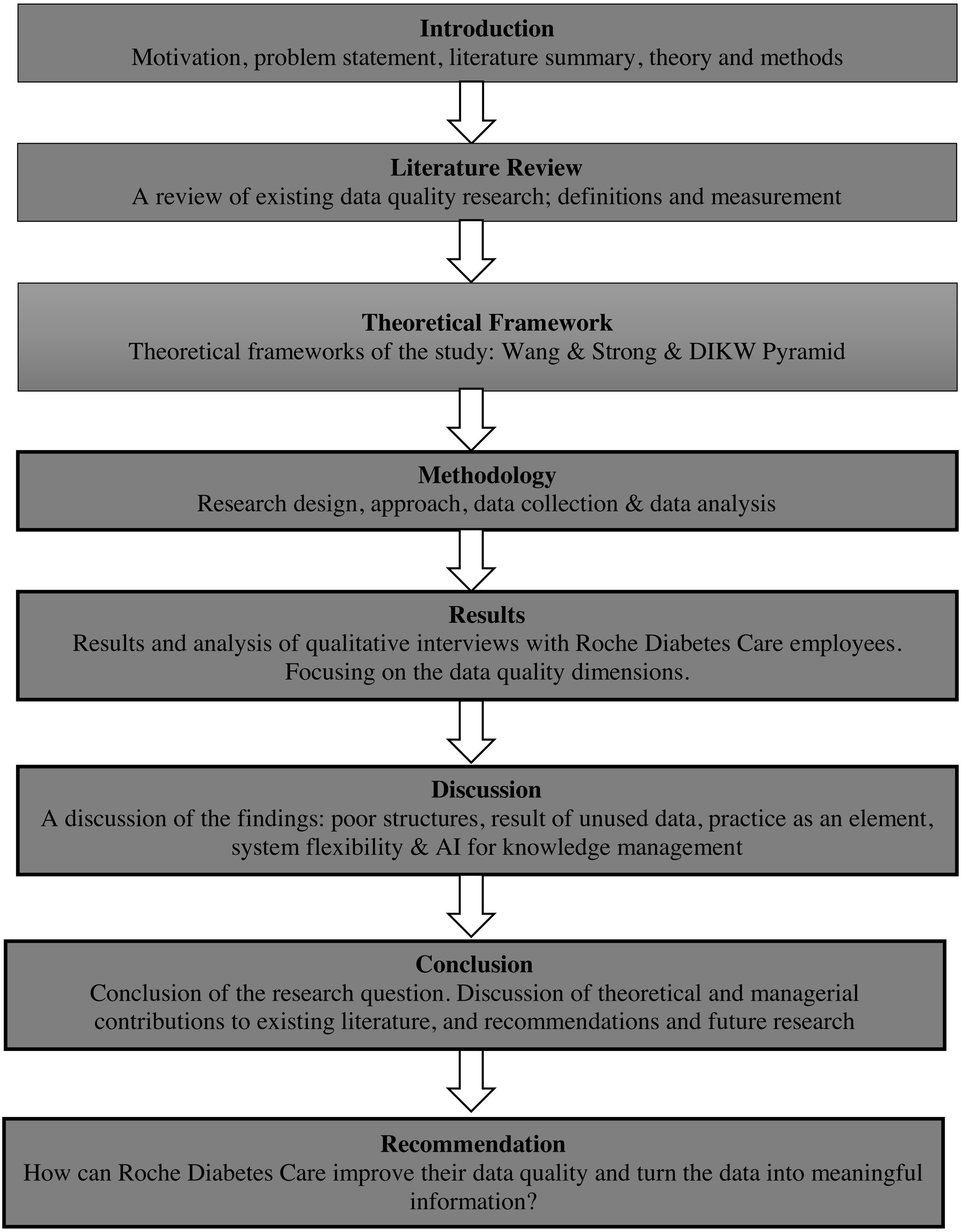
1.5 Methods

The research has employed qualitative data collection and analysis techniques. In order to gain insight into the concept of data quality the literature was searched for definitions, afterwards, an extensive literature review was conducted to summarize existing discussions about data quality. The data was collected by doing semi-structured interviews with employees of Roche Diabetes Care and analyzed through a color-coding scheme. By choosing to conduct a case study it is believed that the findings will be more trustworthy and focused towards exploring the data quality in the targeted organization. Further, each employee interviewed were able to provide own knowledge and experiences about the data usage, which indicates a good source of ideas about employee interaction with the data.

1.6 Thesis Structure

In order to provide the reader with a clear and easy overview of the thesis, following figure is created. The thesis is organized in eight chapters structured as following:

Figure 1: Thesis Outline



2 LITERATURE REVIEW

2.1 Defining Data and Information

In an attempt to define and measure data quality a number of lists of necessary data quality dimensions have been defined in the literature. However, in order to get insight into the research question it is needed to first make a clear distinguish between the terms ‘data’ and ‘information’. According to Tayi and Ballou (1998) data is defined as “the raw material for the information age”. Data support managerial and professional work and are critical to all decisions at all levels of an enterprise (Stockdale & Kerr, 2007; Fuller & Redman, 1994; Tayi & Ballou, 1998). Whilst information is useful data that has been interpreted and processed in such a way as to increase the knowledge of the person who uses the data. (McFadden, Hoffer, & Prescott, 1999). Larry English (1999) describes the data-information relationship as information being data in context, with knowledge being information in context, knowing the meaning of the information.

2.2 Previous Work in Data Quality

As mentioned, in order to define data quality a number of dimensions have been developed and provided for measurement. Typically, the list includes dimensions such as accuracy, reliability, timeliness, usefulness, consistency, precision, understandability and importance. Wang and Strong (1996) have developed a framework based on a survey conducted to understand data quality dimensions from the point of view of the data consumers. As stated by the authors in the very beginning of the research: “data consumers have a much broader data quality conceptualization than IS professionals realize” (p. 5). The survey that was conducted was grounded on the opinions of data quality experts and resulted in four categories of data quality dimensions, namely, the intrinsic, accessibility, contextual and representational. Consequently, these categories are used to group the sixteen most significant data quality dimensions, that are used to measure an organization’s data quality level. (Wang & Strong, 1996). Shanks and Darke (1998) propose a semiotics-based data quality framework grounded on four semiotic levels: the syntactic, semantic, pragmatic and social levels. These levels are considered as being the least developed topics, in the data quality literature, however, they are becoming gradually important for organizations in terms of implementing data warehouses and enterprise resource planning systems, global organizations, inter-organizational information systems, and trans-national consulting organizations. (Shanks & Corbitt, 1999; Shanks & Darke, 1998). When it comes to incidents and factors that can improve or decrease data quality a handful of key authors are contributing to the discussion as well as citing one other. Ken Orr, 1998, Strong, Wang & Lee, 1997 & Pipino, 2004 talk about change as a decrease on data quality in organizations. Since data in the databases are static, and the real world keeps changing the only way to ensure data to be correct is to keep using it. Ken Orr, 1998 presents the Feedback-Control-System (FCS) framework, which states that "if a system is intended to track the real world, there must be a mechanism to synchronize the data in the system with changes in the real world – feedback is necessary." (p. 67). Agarwal & Yiliyasi, 2015 discuss information overload as a factor, which decreases information quality. Agarwal & Yiliyasi, 2015 focus on social media platforms and discusses "how low barriers to publication and easy-to-use interactive interfaces have contributed to information quality problems." (p.5). With the rapid pace of content generation, it gets difficult to follow what is currently happening in social media. The information overwhelms individuals and can challenge the accuracy and reliability of the information. Strong, Wang & Lee, 1997, Wand & Wang, 1996 & Wang & Strong, 1996 propose that high quality data is a result of the flexibility of the system. Since the notion of data or information quality depend on the actual use of data, what may be considered good data in one case, may not be sufficient in another case. Hence, the quality of the data generated by an information system depends on the design and flexibility of the system. (p. 87). Ker & Stockdale, 2007 discusses the importance of high-quality data in the health sector. Their research identified the struggle of getting data users and managers at all organizational levels to understand the importance of data quality and accept responsibility for its improvement and maintenance.

Overall, data quality is subjective and the user who is using the data define whether it is of high quality within its context of use (Pringle, Wilson, & Grol, 2002; Strong, Lee, & Wang, 1997). Therefore, this paper's definition of data quality is:

"Data are of high quality if they are fit for their intended uses in operations, decision-making, and planning. Data are fit for use if they are free of defects and possess desired features."

(Redman, 2001, p. 241)

3 CASE COMPANY

3.1 Roche Diabetes Care

"Doing now what patients need next"

With its Accu-Chek brand, Roche Diabetes Care is one of the three independent divisions of the multinational Swizz healthcare company F- Hofmann La-Roche. As stated in their website the vision of the company is to help people with diabetes in the world think less about their daily diabetes routines, so they can get true relief every day and night. The Accu-Chek portfolio includes blood glucose meters, lancing devices, insulin delivery systems, and digital solutions for data management, advice, coaching and education. (Roche; Diabetes Care, para. 1). Over the last 10 years, the diabetes market has grown significantly and at the moment more than 415 million people are living with the disease globally, while the number is predicted to double within the next 25 years. (IDF Diabetes Atlas). Increased obesity rates and deskbound lifestyles are leading influences on the growing number of Type 2 Diabetes. And with an increased Type 2 Diabetes patients follow increased spending of the treatment of diabetes-related sequela and therefore, new healthcare models to tackle these challenges are needed.

Roche Diabetes Care seeks to utilize digital healthcare technologies in order to support the detection, diagnosis and therapy of the disease and aims to facilitate continuous diabetes self-management support. (Sturman, 2017). With more than 40 years of experience, Roche Diabetes Care aims to create innovative products and solutions to drive and advance the delivery of efficient and effective diabetes care all over the world. (Sturman, 2017). The company's commercial approach includes diversifying value creation around people with diabetes, HCP's and payers using digital solutions as the enabler and the vision is the symphony of four areas of focus, namely, communities, product solutions, digital solutions and services.

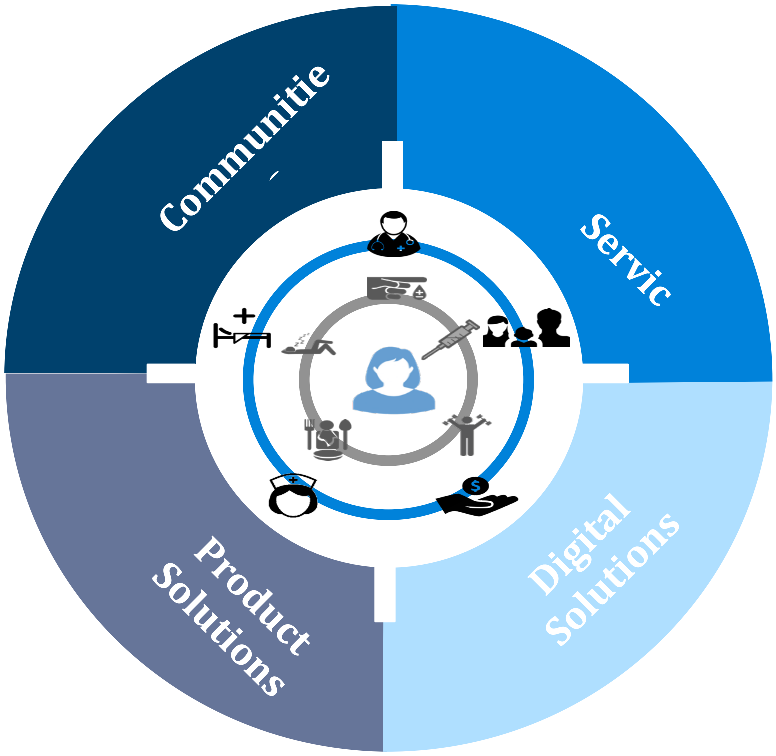


Figure 2: DC Agenda 2025: Roche Diabetes Care Strategy.

COMMUNITIES

Roche Diabetes Care focuses on building external communities and network among diabetics consisting of dialogue and come-together groups. Most countries have created online platforms for people with diabetes in order to give their patients a voice as well as to serve as a community where people with diabetes can meet one another, find relevant information and support. In 2010 Roche DC created a website named "Osmeddiabetes" meaning "us with diabetes", which were further expanded into a community platform where individuals can meet, connect and share stories. The intention of this online patient community in Roche DC, is to bring people into the community by providing content, stories and experiences to where they are already engaging. The website acts as both a content creator and content maintainer. In order to support the website impactful social media pages on Facebook and Instagram are established. Most often articles are created on the website and linked to both Facebook and Instagram. Furthermore, the purpose of strategy for these platforms differs. While Facebook is used to reach out to older people, mostly people with type two diabetes, Instagram is used to engage with younger people with diabetes. This differentiation derives from the idea that Facebook is used to engage by a senior audience, while Instagram by a junior audience. Articles, testimonials and diabetes hacks are published across all the channels in order to engage with the community and keep up the interaction on the platforms. The website and the social platforms are interconnected, and the platforms help keeping traffic to the website by linking to the content created on the site. Online external communities are considered as part of Roche DC's global strategy of helping people with diabetes staying connected and in range.

SERVICES

Roche DC commence a customer centric framework with the goal to help uncover new opportunities with meaningful and practical applications, reinforcing the quality and the patient at the core of their business. (diagnostics.roche.com). Therefore, Roche DC provides free healthcare consulting for all patients. The company aims to create a tailored approach to each patient depending on the patient's needs and condition. Only by customizing the healthcare services the company can be sure that the individual patient will receive the right and the most effective treatment. The company understands the importance of providing not only the right tools and monitors for patients, but also the right knowledge and guidance. While having focused extensively on developing the physical products and offering patients the most innovative products within the diabetes landscape, the company has now changed focus on offering at least at much professional service and knowledge so that patients are in center all the time. This also permits patients to stay connected with professionals all the time. Furthermore, Roche DC has a 27/7-hours Customer Service, that also delivers all kinds of knowledge-based service by answering questions and providing solutions to patients struggling with the monitors. Service is an essential component of the company's vision of helping people with diabetes everywhere in the world think less about their daily diabetes routines and get true relief. (diagnostics.roche.com). Lastly, the services provided by the company aim to enable better and more effective diabetes self-management. However, the services do not only include help to setup the monitors or achieving professional consulting, but rather it is considered as an incorporated mindset of providing relief in any aspect of the life of a person with diabetes.

PRODUCT SOLUTIONS

In terms of physical products, Roche DC is the creator of Accu-Chek blood glucose monitoring systems, insulin pumps, test meters and strips and is also considered to be the biggest supplier of diabetic products in Denmark. The existing Accu-Chek portfolio provides both people with diabetes and healthcare professionals innovative products with impactful solutions for convenient, efficient and effective diabetes management. (diagnostics.roche.com). However, as mentioned, the current business strategy of Roche DC has changed the focus of being product-centric to becoming more solution-oriented by paying more attention to services that provides long-lasting solutions to people living with diabetes. This means that the company does not invest extensive amount of resources on promoting the products as the vision of the company is no longer to provide people with diabetes products but supporting them throughout the whole journey with sparring, coaching and solutions. The product solutions can rather be seen as an enabler to the digital solutions, communities and service solutions, since products are the basis of what the company is selling.

DIGITAL SOLUTIONS

As mentioned, Roche DC focuses on personalization of diabetes management to suit the individual patient's needs and condition. The company focuses on a collaborative, integrated and personalized approach to decide the ideal therapy for each person with diabetes as well as the people in the risk zone of developing the condition. In partnership with the healthcare professionals, caregivers and payers the company endeavors to create value by providing integrated solutions to monitor glucose levels, deliver insulin and track relevant data. By driving digital health in an open ecosystem and offering integrated diabetes management solutions and services, Roche DC is aiming to shape the way diabetes care is being provided. Stay-in-the-Zone is the new healthcare programme created by Roche DC as part of their digital solutions. Through the programme the company wishes to move the focus from being a supplier of medical equipment to be a productive collaborator within the Danish healthcare system. (welfaretech.dk). By using a holistic approach to diabetes management, Stay-in-the-Zone involves healthcare professionals to reinforce the citizens' quality of life, self-management of treatment and sense of security. The expenses of the programme is value-driven and depends on the progress and satisfaction with the treatment; meaning that the municipalities only pay for the citizens' health progress and not the treatment itself. As stated, the vision of the programme is to help people with diabetes think less about their condition and stay more time in range, achieve better daily overview, improved quality of life and experience less late complications. (welfaretech.dk). The main focus of Stay-in-the-Zone are constituted through continuous dialogue with the municipalities and supports the patients through several initiatives such as:

- An online-based app, where health data is collected and managed by the individual patient and can be shared with both doctors and nurses. The app helps the patient to gain control over its blood sugar to "stay in range".
- Frequent support from coach and dietitians – can both be individual and group-based sessions
- Communities and network groups containing dialogue and exercise groups.
- Hardware tools including blood sugar measuring equipment, test and finger sticks.

The four areas of focus are what constitute the business strategy of Roche DC and these aspects aim to envision a holistic approach to the diabetes challenge and enabling people with diabetes to experience more control and overview and a higher quality of life. The shift of focus, from being product-centric to becoming solution-oriented, is also as part of a strategic movement towards creating value to patients and not only providing them with single product. Roche DC believes in the power of digital health solutions and integrated diabetes management solutions to move the needle, bringing true relief to people living with the condition and developing a transparent system for healthcare professionals.

4. THEORETICAL FRAMEWORK

In previous chapters several theoretical frameworks presented in the data quality literature were discussed with the intention of providing the readers with an overview of existing research on the topic. In this thesis the framework by Wang & Strong (1996) will be the main contributor of theories. However, in order to explain the certain phenomena, it is necessary to include other supporting theories as well. The data quality framework by Wang & Strong (1996) is used to measure the data quality in the company, while the Data-information-knowledge-wisdom (DIKW) model is used to describe and discuss the transformation of the data collected, into useful and meaningful information.

4.1 A Conceptual Framework of Data Quality (1996)

Richard Wang's and Diana Strong's research identifies characteristics of data quality that are important to data consumers. The authors' starting assumption is that "data consumers have a much broader data quality conceptualization that IS professionals realize" (p.5). For the construction of the framework is followed the methods developed in the marketing research for determining the quality characteristics of products. The approach treat data as a product, which is in this thesis seen as an appropriate way "because information systems can be viewed as a data manufacturing system acting on raw data input to produce output data or data products." (p. 8). Even though data consumers are not necessary purchasing data, they are choosing to use or not to use data in their different tasks. Within the marketing discipline there are found several approaches for assessing product quality attributes that are important to consumers. Following the marketing research, the research by Wang & Strong (1996) identifies the attributes of data quality that are important to data consumers. The framework is developed through several steps.

They start out by collecting a survey of in total 118 data attributes identified by data consumers. Next, they collect importance ratings for these attributes and structure them into a hierarchical representation of data consumers' data quality needs. When grouping the intermediate dimensions into families a preliminary conceptual framework developed from the experience with data consumers were used. The framework consisted of four 'ideal' categories and the intention was to evaluate the extent to which the intermediate dimensions matched these categories. Afterwards, a follow-up study was conducted consisted of two phases. The first phase intended to sort the dimensions into categories and label the categories. In the second phase, a different set of subjects was instructed to sort the dimensions into the categories discovered from first phase to confirm the findings. The key result of Wang & Strong's research is a comprehensive framework of data quality from data consumers' perspectives. The framework performs as a foundation for improving the data quality dimensions that are seen as important to data consumers and, as mentioned, will in this thesis be used to measure the data quality in Roche DC. Following the four categories used to measure the data quality is defined.

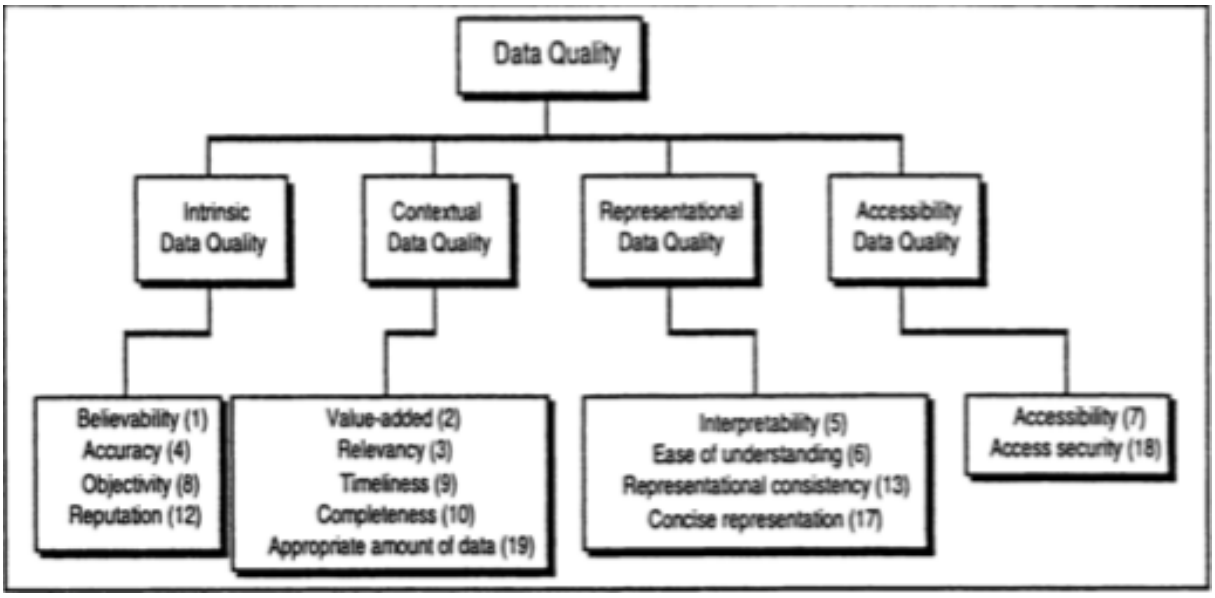


Figure 3: A Conceptual Framework of Data Quality

Intrinsic Data Quality

Intrinsic DQ suggests that data have quality in their own right, meaning that the data values presented are consistent with the actual or true values. Intrinsically good data is accurate, correct and error-free and comes from a valid source. A common cause of intrinsic DQ concerns are mismatches among sources of the same data. Dimensions that include intrinsic DQ are: accuracy, objectivity, believability, and reputation. (Wang & Strong, 1996).

Contextual Data Quality

The contextual category indicates how data quality must be considered within the context of the task at hand, meaning, that data must be applicable to the task of the data user to be considered as high quality. However, the focus of contextual DQ is the data consumer's task, not the context itself. For instance, it is required that contextual appropriate data must be relevant to the data consumer, in terms of timeliness and completeness. Dimensions that include intrinsic DQ are: value-added, relevancy, timeliness, completeness, and appropriate amount of data. (Wang & Strong, 1996).

Representational Data Quality

Representational DQ include aspects related to the format of the data as well as the meaning of the data. Representational DQ suggests that the system must present the data in such a way that is easy to understand and interpret by data consumers. This happens when data is presented in an intelligible and clear way. Dimensions that include the representational DQ are: interpretability, ease of understanding, representational consistency, and concise representation. (Wang & Strong, 1995).

Accessibility Data Quality

Accessibility DQ indicates the importance role of systems, referring to as the extent to which data is available to or easy to access by the data consumers. On the other hand, it is also important that the system is secure to protect the data from external and non-relevant people. Data-representation issues interpreted by data consumers can be seen as accessibility concerns. Dimensions of accessibility DQ are: accessibility and access security. (Wang & Strong, 1996).

When researching the literature, the framework described above by Wang & Strong was the most discussed and used in the work of data quality. The hierarchy is a useful classification and the categories highlight surfaces of data collection, storage, and use that have a direct impact on data consumers' perceptions of quality. (Coleman, 2013). For a data consumer to have a perception of intrinsic qualities of data, he or she must understand what the data is intended to represent and how the data is in conformance with that representation. If the data is not aligned with a consumer's assumptions about these things, the consumer will perceive it as inaccurate or unbelievable. Data can be seen as incorrect if the consumer does not understand the conventions through which a system presents it. Contextual DQ emphasizes the idea that the quality of data is defined to a large extent by the intended uses of data. Last, accessibility DG emphasizes the aspect of system design. The consumer can only judge the quality if he or she has access to the data. (Coleman, 2013).

4.2 The DIKW Pyramid

Where is the Life we have lost in living?
Where is the wisdom we have lost in knowledge?
Where is the knowledge we have lost in the information?

T.S. Eliot, “The Rock”, Faber & Faber 1934.

Beside using Wang & Strong’s data quality frameworks to assess the quality and usefulness of the data in Roche DC, this thesis will also consider the “Data, Information, Knowledge & Wisdom” hierarchy in order to explain the transformation of data into useful information and knowledge within the company. The origin of the hierarchy is first presented in Sharma (2004) where the first entrance of the hierarchy is discussed both in the Knowledge Management and Information Science domains. The hierarchy is later been further-developed by Ackoff (1989) and revised again by Rowley (2007). This thesis will mainly follow Ackoff’s description and proposal of the hierarchy. The DIKW model highlights a linear input and output process, in which raw data inputs, digital storage and information organization, and the end-products, such as knowledge, are arranged and processed to support knowledge creation and decision-making. (Kelly, 2018). According to the DIWK hierarchy, data are the “most fundamental elements of the digital age” (Kelly, 2018; Walsham, 2001: 26). Ackoff defines data, Information, knowledge, understanding, intelligence and wisdom and explores the processes associated with the transformation between the elements (Rowley, 2007). Following a definition of the different elements will be given referring to the original developer of the model. However, in this thesis, only elements such as data, information and knowledge are comprehensively considered in the analysis, but all four elements are described in the theoretical framework.

Figure 4: The DIKW Pyramid (Adapted from Rowley, 2007: 164)

4.2.1 Definitions and usage of data, information, knowledge & wisdom

Data are defined as symbols representing possessions of objects, events and their environments and are only of use if they are in a useable, interpretable or relevant form for the task at hand. The difference between data and information is not structural but depends on the functionality. (Ackoff, 1989). The first and fundamental step of the DIKW hierarchy is data. In order to achieve a meaningful result, it is necessary to start out by collecting raw data. All measurements such as logging, tracking and recordings are considered as data.

"Information is a distinction that makes a difference." (Mackay, 1969). Information is the collection and interpretation of data. Data is collected into groups and provided with meaning. The main role of information systems is to store, retrieve and process data and turn them into useful information. (Ackoff, 1989). Only when data is placed in a context, where meaning can be derived and attached to the data, information emerge. Also, information has properties; meaning that information can be incomplete, useful or inaccurate. What makes information significant and valuable is that it is semantic and has a meaning, as it is an interpretation of data with an appropriate context.

"Knowledge is the capacity for effective action" for use of any information. (Senge, 2000). A suitable description of knowledge is that data and information become knowledge when it gets "organized and processed" to provide understanding, experience, learning and expertise to a current activity. Knowledge can be gained either by transmission from someone who has it, by instructions, or by obtaining it from experience. (Ackoff, 1989). Therefore, knowledge is the capacity to act, supported by useful information, to attain intended goals and objectives. While information is semantic, knowledge is pragmatic and intended for action and to be put into use.

Wisdom is the last element in the hierarchy and is defined as the ability to increase effectiveness. Wisdom adds value to information and requires the mental function that is called judgement to come into play. The ethical and aesthetic values are intrinsic to the actor and are unique and personal. (Ackoff, 1989).

The DIKW Pyramid is used to contextualize data, information, knowledge, and wisdom, with respect to one another as well as to identify and describe the processes involved in the transformation of an entity at a lower level (e.g. data) to an entity at a higher level (e.g. information). The main assumption is that data can be used to create information; information can be used to create knowledge; and knowledge can be used to create wisdom. (Rowley, 2007). The model can be viewed from two different viewpoints: contextual and understanding. The phases of the **contextual** concept are as following: One moves from gathering data parts (data) to connecting the raw data parts (information), to forming whole meaningful contents (knowledge) and conceptualize those whole meaningful contents (wisdom). The perspective of **understanding** highlights the DIKW Pyramid as a process starting with researching & absorbing, doing, interacting, and reflecting.

The hierarchy model by Ackoff presents as a necessary tool for the second part of the analysis when explaining the data transformation into useful information, knowledge and decision-making. As each step of the pyramid answers questions about the initial data and adds value to it, the more questions one answers the higher one moves up the pyramid and the more knowledge and insight is achieved. At the top of the pyramid the knowledge and insight gained is turned into learning experience that guides the actions for improvement of data quality

5. METHODOLOGY

Throughout the process of conducting a scientific research many questions and approaches are considered. Many decisions are reflected upon and many choices are made. Scientists are studying multiple options to be sure to choose the right scientific method that ensures to effectively answer the research question. For the reader it would be most effectual to only be provided with the results and conclusion. However, this is not a realistic case as the practice of business research does not exists completely sealed off from the social sciences and the knowledgeable commitments that their specialists hold. (Bryman & Bell, 2017). Therefore, it is necessary to explain the scientific choices made in the process of conducting this qualitative research. Consequently, following, a description of choices and selections, the philosophy of science, data collection and methods for analyzing the findings will be presented. In order to provide the readers with an overview of the methodology section, following figure can be used as a section guide.

Figure 5: Methodology Guideline

5.1 Research Design

5.1.1 Analytical and Empirical Contributions

The aim of this thesis is to contribute with analytical- and empirical knowledge to the already existing data quality literature by doing a single case study. A case study is an empirical inquiry that investigates a contemporary phenomenon in depth and within its real-life context. (Yin, 2009. Further it is a great way to test theories (Gustafsson, 2012; Anderson, 1983), to render description (Kidder, 1982) and to develop theory about topics. (Gustafsson, 2012; Eisenhardt & Graebner, 2007). The topics for case studies can be internal organizations, group processes and strategy (Gustafsson, 2012). As the goal of this research is to explore phenomena such as data quality and knowledge creation in a single organization, and develop a deeper understanding of the subject, a single case study is the strategic choice to go for. The main component of a case study, relevant for this research, is the formation of the research question, theoretical suggestions, units of analysis, the linking of data to the right theoretical suggestions and the evaluation of the suggestions. (Lee, 1999).

5.1.2 An Iterative Approach

When conducting a scientific research paper, it is relevant to figure out the best possible approach, which provides the researcher with enough data and theory needed for completing the project. Two most common approaches used among researchers are called deductive and inductive approaches. (Bryman & Bell, 2017). This thesis is based on a deductive approach, since, Wang & Strong's theoretical framework on measuring data quality was the point of departure of the research. However, elements of an iterative approach are used too. The iterative approach, also called the abductive approach among researchers, made perfect sense to use, as it provided something in between.

The iterative process involves working back and forth between data and theory, as each new finding will influence the next step of the process. (Bryman & Bell, 2017). The process of jumping back and forth continues until all the data and theory is found and related. As mentioned, this research had its starting point on the theoretical framework by Wang & Strong. Using the dimensions presented in the framework as a guideline, the qualitative interviews were conducted. The data collected through the interviews with the first five employees highlighted a certain pattern, which led the project to change the research question to focus more on the transformation on data into meaningful information. Further, the first couple of interviews also signified that more theories were needed in order to gain a fully perception of the data-information usage in the company.

Theory regarding the iterative approach suggests that once the phase of theoretical reflection on the data has been carried out, the researcher might need to collect more data to confirm the theory or follow up on the hypothesis. This strategy looks like part of the deductive and inductive approaches and, as mentioned, it primary specifies that researchers go back and forth between data and theory, just like this thesis has done throughout the data collection process.

5.2 Social Science

5.2.1 Ontological Concerns

Questions of social ontology attempt to define what reality is. It is the study of the reality of the world. (Jordansen & Madsen, 2010). The fundamental idea is the question of whether social entities should be considered objective entities with a reality external to social actors, or they should be considered as social constructions constructed from the perceptions and actors of social individual actors. These two different views are also referred to as objectivism and subjectivism. (Bryman & Bell, 2007). While quantitative research often views the world as being one single objective, qualitative research assumes that many subjectively realities can exist in the same time. (Lee, 1999).

The ontological beliefs of this project are based on the interpretivism paradigm. It is believed that reality, as well as our knowledge of it, are socially products and incapable of being understood independently of the social actors that construct the reality. The aim of the interpretivist research is to understand how members of a social group, through their participation in social processes, enact their realities and give them meaning, and to show how these meanings, beliefs and intentions of members help to constitute their social action. (Orlikowski & Baroudi, 1990). The reason that this research goes with an interpretivist paradigm is because a case study has been conducted and the data obtained comes from different employees, each comes with different experiences regarding the data usages in the company, hence, having a unique view on the topic.

5.2.2 Epistemological Concerns

While ontology attempts to define reality, epistemology attempts to find methods to figure out the truth about this reality. (Bryman & Bell, 2007). It focuses on understanding how people recognizes the reality, and how human knowledge is produced. (Jordansen & Madsen, 2010). The substance of this can either be seen as objective: realism, or subjective: interpretivism. As mentioned earlier, this thesis goes with the interpretivist paradigm, as each qualitative interview that is conducted reflects a unique interpretation of the topic. Therefore, the research goes with the subjective side of the epistemology. While qualitative researchers highlight the fact that one should interact with the studied phenomena, qualitative researchers often assume their independence from the variables studied. (Lee, 1999). After finishing the interviews, the author of this project continued the interaction with the company through e-mail and telephone communication. Whenever, the authors needed any further information, the company was contacted. Every single interaction made during both the interviews and the e-correspondence affected both the author and the employees. Therefore, it can be carefully argued that the thesis project has been conducted with a subjective and interpretivist view.

5.3 Data Collection

5.3.1 The Qualitative Research

This thesis is conducted through a qualitative research as it is concerned more about words and relations rather than numbers and statistical matters. What most often characterizes a qualitative research study is a subjective epistemological position and an interpretivist ontological position as described in the previous part. According to Kvale qualitative research mostly focuses on the identification of meaningful categories, or parts of organizational phenomena (Lee, 1999). Similarly, the organizational phenomenon, which this thesis is concerned about and tries to identify is data quality. Furthermore, Kvale proposes that the number of tools to conduct qualitative research depends upon the analytic situation. (Lee, 1999). The tools needed for this qualitative research was a semi-structured interview guide, which will be further explained in the next section.

The topic of this thesis was found in a very early stage of the process. And after doing an extensive literature review the research question was also formulated. The literature review allowed the author to discover a gap in the healthcare-data literature that could also benefit the company. Since the author had already contacts within the company, it made it less difficult to choose the most relevant interviewees with sufficient knowledge about the data usage in the company. In order to collect the empirical data a semi-structured interview guide was used and seven employees in total were interviewed; five of them were sitting in local functions and two of them were sitting in global functions, but all had different years of work experience, different title and different knowledge about the data usage. One of them were working with the SAP system, while the rest only worked with Salesforce. Since the project has a local focus on Roche Diabetes Care Denmark, it made most sense to conduct majority of the interviews with people within the Danish affiliate. However, it is also valuable to gain a global view on the topic, since most of the data guidelines and structures are provided from global.

The qualitative research approach is less structured than the quantitative approach is. In qualitative research the main interest is the interviewees point-of-view, and the analysis is constructed from the interviewee's interpretations. (Bryman & Bell, 2007). The process of conducting the interviews were therefore also very well-prepared. The interview guidelines were made and approved a week before starting the interview conduction. Two different guidelines were made; one made for local employees and another targeted the global employees. While the questions asked for the global employees were mostly concerned about the global visions of the data, the questions asked for the local employees were based on the data quality dimensions. It is important to understand the interview guidelines as a flexible guideline rather than a fixed structure when doing qualitative research. All the questions asked were open-ended and it was important to encourage the employees to elaborate on definitions, stories and examples. By allowing them to speak generally instead of focusing deeply on one question at the time, rich and detailed answers were achieved. Further, when something interesting in regard to the topic was mentioned, the interviewee was encouraged to elaborate and speak further about this. This way, it was made sure that each interview was unique and all the knowledge the employee had about the topic was reached.

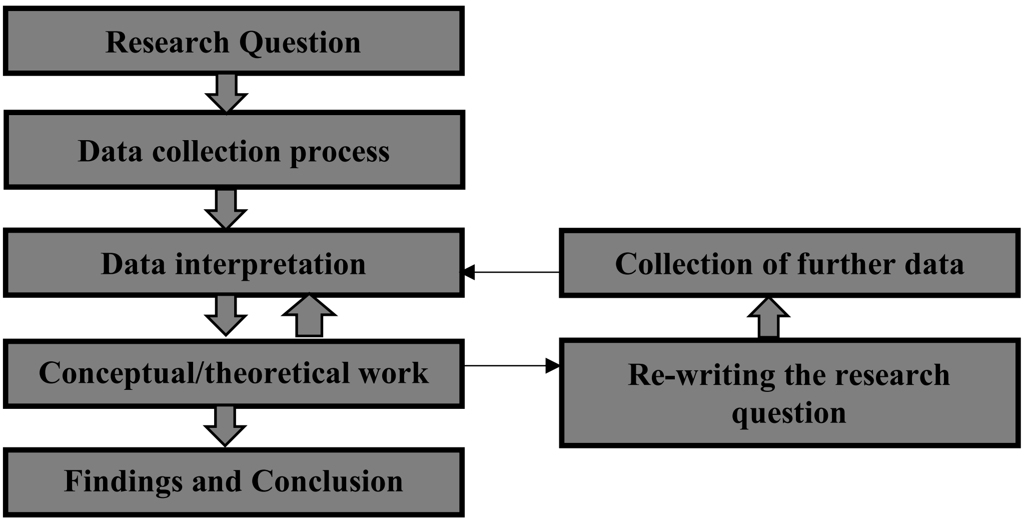


Figure 6: Steps of qualitative research

5.3.2 Formulating a Research Question

This process started by only knowing the topic. At the beginning, the only thing that was certain was that this thesis was going to look into the study of data quality in healthcare. As mentioned earlier, since healthcare data is a hot topic at the moment and the author had access to the company, it was an excellent opportunity to explore the topic of data quality in the context of healthcare. After researching about the topic, it was found that many researchers including Wang & Strong had tried to explain data quality, however, there were not much research done about data quality in a healthcare context. Inspired by Wang & Strong's data quality framework, which is designed to assess the data quality in an organization, and which is fully explained in the theoretical framework's section, the research question was defined.

This thesis aims to investigate how data is transformed into meaningful information and how knowledge creation is positively influenced in the company among employees. Therefore, the data quality framework is an excellent choice for understanding data-consumers data concerns. This project has decided to go with a how-question, since how-research questions is mostly asked by qualitative researchers who observe trends in the analysis and demonstrate flows in the topic of interest (Bryman & Bell, 2007). Additionally, the topic of data quality and data transformation is focused on movement and tries to understand how human as well as systems interact with the data and affect the transformation process. Therefore, instead of studying the static characteristics of data and human being, this thesis is concerned about the *hows* of the process. How data is processed, *how* data is transformed, and *how* data employees interact with the data. (Dervin, Foreman-Wernet, & Lauterbach, 2003). Furthermore, only one research question is developed, since it is believed that it is broad enough to cover the field of interest, and narrow enough to remain a precise and concrete focus.

5.3.3 The Interview Guide Structure

In order to collect the empirical data in a structured way and remain a focus on what is desired to achieve, a semi-structured interview guide is used for the process. Basically, a semi-structured guide refers to a list of memory prompts of an area to be discussed with the participants. According to Bryman & Bell (2007), when preparing a qualitative interview, the interviewer should ask himself: "what about this is puzzling me?" This is also imitated in this thesis. In the beginning of this project the author asked herself, what in regard to data quality and information transformation exactly made her wonder? The reflections and questions resulted in the semi structured interview guide that can be found in appendix 1.

Researchers suggest keeping in mind different elements when crafting an interview guide. Such as, remembering to create a certain amount of order on the topic areas, so questions can flow well, use a language understandable to the interviewees and try to formulate questions that can help answer the research question. (Bryman & Bell, 2007). All of these elements were also kept in mind throughout this process. However, the formulation of the research question was still not definite at the time the interviews were conducted, which resulted that alternative enquiry appeared after the questions were studied further in the analysis. Furthermore, the questions were all related to Wang & Strong's theoretical framework. One question was asked to each of the twenty dimensions, however, the interviewees were also encouraged to explain and talk about alternative discussions and topics related to the question.

5.3.4 Conducting the Qualitative Interview

As mentioned earlier in this paper, qualitative research is much less structured than quantitative research. The focus in qualitative research is the interviewees' point of view. Therefore, going away from the specific question is encouraged as it gives insight into what the interviewee considers as relevant in regard to the topic. (Bryman & Bell, 2007). This thesis has also made use of these generalities throughout the interview process. When conducting the interviews, the author had prepared for being flexible towards the answers and encouraged the interviewees to bring up and discuss any topic or issue they might find relevant in regard to the topic.

When the interviewee shared a story, the author tried to encourage the interviewee to go more in-depth by asking to follow up and ask questions such as: can you tell more about this process? Or, why do you handle this process this certain way? Are there alternative ways you could handle it? All the interviewees were asked the exact same questions, however, since each interviewee came from different departments and had different roles, they gave each a different insight and perspective to the questions asked. The flexibility and improvisations made each interview unique from one another, which is important when conducting a semi-structured interview. As stated by (Bryman & Bell, 2007: p. 475): "Questions may not follow on exactly in the way outlined on the schedule. Questions that are not included in the guide may be asked as the interviewer picks up on things said by interviewees".

Basically, a semi-structured interview needs to keep a balance between a free flowing and a direct conversation (Lee, 1999), which this thesis managed to follow through. All of the interviews were recorded, and the interviewer had time to take notes, make observations, and right down follow-up questions, that could be discussed in the end of the interview. Recording the interviews was extremely helpful, since, the author could focus more on other formalities such as focus on eye contact, listen, observe and make the interviewee feel comfortable. Further, recording the interviews is very effective and valuable for achieving fully details. After rehearing the recordings and starting to transcribe them a lot of new insight was realized, and new patterns were discovered.

Finally, all of the interviews were conducted in the company. An office was borrowed and the settings and context where the interviews took place was quiet, which was the reason that the interviews went well without any interruptions from the outside world. The fact that the interviews were conducted in the company made employees feel more comfortable and created a familiar atmosphere, where each interviewee could recall any experience and occurrence made.

5.3.5 Secondary Data

In this thesis, the primary source of data will be the data collected throughout the semi-structured interviews. The secondary data sources will come from the company's home page. Most of the company information that was needed to write about the organization was found at their website. However, figures, models and guidelines about the company strategy was received from one employee in the company. Further, articles about the company were also found online and used for the analysis.

5.4 Data Analysis

5.4.1 Transcribing

In order to reflect upon and understand the interviews researchers usually transcribe them. As mentioned, in total seven interviews were conducted, and all lasted approximately half an hour. After conducting the interviews, an intensive and time-consuming transcription process began. One single interview could last 3-4 hours to transcribe and resulted in around 5 pages of text. This resulted in nearly 30 pages of empirical data to process and analyze. (Appendix 2). Since the transcribed empirical data act as a foundation for the analysis, it is important that the process is done carefully and consistent. In order to check the reliability of the transcriptions, all of the interviews were listening to while reading the final transcribed text. The texts were further processed by applying data analysis coding discussed in the following section.

5.4.2 Data Analysis Coding

As mentioned above, a great amount of empirical data is collected throughout the data collection process. Researchers spend a great deal of time and effort to categorize the data, and code the categorizations in order to gain an overview and discover valuable insights. Coding is the process in which data is organized into meaningful structures and is most often done through three phases: open coding, axial coding and selective coding. (Lee, 1999). This thesis has completed the data analysis by going through all three phases right after one another other.

First in line is the open coding method. This process refers to an open approach, where the researcher can create as many categories as needed in order to organize the material. In this process all of the empirical data was read carefully several times and anything important and relevant to the research question was highlighted. When all the data was highlighted the coding process was continued by moving to axial coding. Axial coding is the second process and involves identifying relationships and connections between the open codes. Last phase is the selective coding process, where all the relationships are integrated into analytical frameworks. The data analysis work was considered as one of the most important processes of this thesis, since it provided clear interpretation, which supported the analysis. Following, an example of the data coding process provided:

Open	Axial	Selective
Q: To what extent do you experience the data to be correct, error-free and accurate? When they call us directly they just answer the questions that we ask them if they don't understand a question they are able to ask for clarifications. By calling us directly there is less chance for misunderstanding and higher chances that the data is accurate.	Two data streams: - Online system form - via call Data accuracy depends on the medium used to collect	Data is received either by call or survey. Data accuracy is higher when it is by call, meaning data accuracy depends on the medium used to collect it

Table 1: Data Coding Analysis. For a more detailed overview, see Appendix X.

5.5 Limitations and Credibility

5.5.1 Limitations

One of the major limitations faced throughout this process was getting access to the right people in the organization. In order to get a wholly picture of the data and the usage of it, the project required interviews from both local employees, the management team and global employees. However, since people are busy with own agendas it was time-demanding to organize and collect the interviews in time. By being dependent on people to offer their time for the 30 minutes interviews resulted in some delays in the project plan. However, the delay was not too long, and every employee during interviews showed commitment and took their time to answer each question very thorough- and systematically.

The second limitation considered is the number of participants who contributed with empirical data. Basing the analysis and findings on seven participants may be considered as not sufficient for gaining a general understanding of data quality and usage. However, since this thesis follow a qualitative approach the low number of participants can be uplifted by creating thick descriptions and rich accounts of details of the employees understanding of data quality in Roche DC. Therefore, even though the thesis conducts an intensive study of a small group of people, the data that is collected, represents a broad part of the studied topic. Qualitative research examines the quality of each answer provided and is measured by the richness of the detailed answers given.

5.5.2 Credibility

In regard to the quantitative research, both reliability and validity are important measure in assessing the quality of the research. However, researchers have discussed their relevance for qualitative research for being low, since measurement is not a major call among qualitative researchers (Bryman & Bell, 2007). Because this thesis operates from a subjective and qualitative standpoint, they are not considered important.

However, since credibility proposes something about how believable the findings are, it is important to be addressed. When dealing with different kinds of social realities it is important that the collected data is trustworthy and reliable. This indicates that the research is credible. It means both that the researcher has carried out the research according to right principles of qualitative research, but also, that the researcher has not mislead the interviewees and affected their interpretation. (Bryman & Bell, 2007). Furthermore, according to Lee (1999), in order to consider a qualitative study's credibility, the researcher is supposed to be able to answer yes to this question: "has the participants' cognitive schema or worldview been successfully captured by the researcher?". This thesis can answer yes to that question; therefore, the findings of this thesis are considered as being credible.

5. RESULTS

In order to answer the research question correctly, the analysis part is divided into two sections. The first section will be mainly focusing on Wang & Strong's data quality framework. The data quality in Roche DC is measured and assessed from the four categories and twenty dimensions presented in the theoretical framework. Wang & Strong's framework is intended to provide an understanding in terms of accuracy, timeliness and usefulness of the data. The second phase of the analysis will consider the DIKW model in order to analyze the transformation of data into useful information, and the information into useful knowledge. Furthermore, concepts like knowledge creation and knowledge sharing will be discussed as discharged from the analysis.

5.1 PART ONE: Assessing the Data Quality

When analyzing the empirical data in regard to the theoretical framework, the analysis will be structured in the same order as the dimensions were asked about throughout the interview. The structure of the first part of the analysis will be intrinsic, contextual, representational and accessible data quality. It is also worth knowing that not all dimensions were discussed throughout the interview, since not all dimensions were relevant for all employees. For instance, employees in global functions were able to better answer questions about the system structure and visions of the data, while employees in local functions were better at answering questions about the data usage and storage in the databases.

5.1.1 Intrinsic DQ Pattern

The first category of dimensions is namely intrinsic, which, as mentioned, refers to the extent in which data presents itself as accurate or non-conflicting, hence are able to add value to the task and project at hand. The dimensions that are related to the intrinsic category and which define the exactitude of the data are; *accuracy*, *objectivity*, *believability* and *reputation*. An intrinsic DQ concern could, for instance, arrive by mismatches among sources of the same data or when the sources are poor or not trustworthy. It could also arrive when data is not handled and worked carefully. All Roche DC employees within the Danish affiliate, who were asked to describe the correctness of the data defined the data itself as accurate and reliable majority of the time, however there were three certain ways the correctness of the data was negatively influenced. One way was when patients failed to provide the right answers by mistake. This could be data such as an address, e-mail or age. Here, external forces such as patient mistakes, is the reason for inaccuracy. The second way is when employees make mistakes when logging the data into the system. The data itself comes mostly from patients, hence derives from reliable and original sources, however, the process of inputting the data into the systems by an external force is when the inaccuracy appears. As stated by one employee "So, the data is correct if the people entering the data enters it correctly. I mean it is always a human error that is the cause for the data not to be correct." (Interview 3: Linda, 2019).

When such a thing as above mentioned happens, data consumers, which are Roche DC employees, often do not know the reasons for why the quality problems arise, and most often they don't even know that data is conflicting when they start handling the process. The first step for the data consumers to realize that data is conflicting is when, for instance, e-mails are bouncing back within the system. As mentioned by another employee: "The error or inaccuracy occurs when people change e-mail address or phone numbers without informing us. Whenever we send out something there are some e-mail addresses bouncing." (Interview 4: Anne, 2019). This occurrence can be both a result of patient or employee typo but can also be a result of change. Another main reasons for the data inaccuracy within the systems is the static/dynamic issue. The concern with data quality is to ensure not that the data quality is perfect, but that the quality of data is accurate enough, timely, enough, and consistent enough for the organization to survive and make reasonable decisions. (Ken 1998). Since, the data in the systems are static and the outside world keeps changing, companies who do not use resources on keeping the data in databases up to date face a higher extent of data-inaccuracy. If the data has to maintain a certain level of accuracy feedback is necessary in a regular and consistent basis. The importance for accurate data is especially critical in the healthcare sector, since inaccurate data can have great consequences for patients, e.g. if they don't receive the right product at the right time. However, Roche DC does not have direct interaction with patients, so the inaccuracy of data does not have immediately impact on patients.

Yet, the abovementioned concerns and reasons for data inaccuracy can over time appear as *believability* problems not only for data consumers, but also for stakeholders, which can lead to poor *reputation*. If a significant large amount of e-mails are bouncing due to incorrect data stored in the systems it indicates that only a little value is added for the organization, resulting in higher cost compared to the actual effect.

The system that is used for data management in the company is Salesforce. Between the system and the data acts the data consumer as a middleman. Therefore, judgement and subjectivity in the data production process is also a common challenge for the data quality. Interpreted data is considered to be of lower quality than raw, uninterpreted data. Though, the data comes directly from patients the fact that it is entered in the systems by employees can challenge the *objectivity* and cause that poor-quality data becomes common knowledge. One more aspect to consider is related to the marketing efforts that can result in the data stored in the systems not always representing the reality. As stated by one of the employees from the customer service department:

"Well, the data itself is correct. The aspect that could be challenging is that when we run a campaign and offer free devices, people can tend to choose the wrong devices. For instance, a type two diabetic who has just started monitoring chooses a product which is more suitable for a type one. Here, the data will be correct, but the product type or suitability won't. This leads to the fact that we have data in our system about people having a product not suitable for them… and they will throw the product out only after realizing this." (Interview 1: Casper, 2019).

This incident does not only cause that the data in the databases fail to represent the real life, and become considered of poor quality, but also leads to believability concerns in regard to the products that the company is offering. Since, diabetes is a very restricted disease, it can require customized treatment based on the needs of the patient. When the company is offering devices and the patients order the product that is not suitable for them, but for other patients, only to realize this and throw it out later, the company loses first, credibility and second, money. The patients will find the company reliable if the products offered provides value to the life of the patient and is in accordance to the promises. However, in order for this to happen more targeted and customized product treatments are needed. Hence, not only is the data in the systems different from reality, but also side effects are experienced in a regular basis.

5.1.2 Contextual DQ Pattern

The second category of dimensions is contextual, which refers to the requirement that data quality must be considered within the context of the task at hand, refers to the extent in which data are applicable (pertinent) to the task or project at hand. (Coleman, 2013). As mentioned previously, the dimensions include: value-added, relevancy, timeliness, completeness, and appropriate amount of data. The data accuracy discussed above can be considered further in this section. The empirical data showed two causes for data consumers' complaints that available data did not support their task at hand: missing (*incomplete*) data and inadequately defined structure for the data usage resulted from the new GDPR rules implemented in 2018. The two causes are contextual as both issues present as problems within a certain context and not as a general problem experienced by all employee in Roche DC.

As mentioned in the intrinsic section, data consumers in Roche DC oftentimes experience the data to be incomplete. As defined by one employee: "Sometimes the data can be incomplete if the customer does not answer all the questions in the formula; for instance, miss a birthdate or phone number." (Interview 3: Linda, 2019). The incompleteness of the data of course contributes to data accuracy problems, and therefore often times fails to add value to the task that is processed at the time. Another employee describes the transitional period when a system change was made changing from MS Dynamics into Salesforce and also explains how changing to Salesforce has increased the chances for data to present as incomplete: "Before we started operating through Salesforce we worked with Dynamics, and Dynamics was connected to the Post Office Center. When you entered two or three of the first letters of the street name it provided you with suggestions, hereafter you could enter a road number. This way you could not enter a wrong or non-existing number." (Interview 3: Casper, 2019).

This gives indications of how the data can be prevented from performing as incomplete by help of the system design. In order for Roche DC to gain fully advantage of the data collected it would be an idea to make the system help them maintain and ensure the data completeness. However, the challenge occurs when Roche DC operates under the procedures of global that offers the structures, guidelines and tools. Salesforce is a huge system used throughout the whole Roche DC global landscape, and because of Roche DC DK's size it cannot request system changes by its own. Rather it is required to adjust the structures that global is operating under. Another cause of data incompleteness in Roche DC is, as mentioned earlier, when employees or patients fail to fill out the form correctly, or when data changes occur outside the system. Since the system cannot delete any data, but only change patient status "from being active user to non-active…" (Interview 3: Linda, 2019) it can be a challenge as the data might not always be timely and up to date. Because transaction data is incomplete, the database cannot be updated, which, in turn results in inaccuracy. A further constraint, which the system's inability to delete data can lead to, is data overload within the systems, which also has a side-effect of leading to poor data quality, which does not support the data consumer's task or projects at hand.

An additional cause for data consumers' complaints about data not supporting their task at hand was the inadequately defined structures for data usage. While some employees complained about the data itself a majority also complained about not having enough knowledge about how to use the data. In order to be able to use the data correctly it is important that a certain structure is provided. For the data to add value to task and company, clear and detailed data procedures are essentially. The empirical data collected through interviews showed a significant lack of awareness relative to the compliance side of the data usage. GDPR seems to act as a great limitation for the company to make efficiently use of the data. Five out of seven employees emphasized their lack of knowledge of the GDPR rules and explained it as the biggest hurdle when using the data: "The data that we collect has a huge potential for use. But once again we are limited by the GDPR rules. We still don't know what we are allowed to an what we are not allowed to. Once we know these guidelines we can become more targeted in so many aspects." (Interview 3: Linda, 2019).

However, this contradicts the statement that was reached from the global interview, who emphasized that the GDPR rules and guidelines have been already provided for the countries. As stated by the global Roche DC employee: "What I know is that they of course provide global policies and guidelines and work with local legal counsel and ensure that the global guidelines are broken down into local guidelines." (Interview 6: Julia, 2019). Basically, global policy guidelines are communicated to local compliance directors, which makes sure to communicate them internally in the local organizations. Nonetheless, there are still a confusion among employees in Denmark on the role of GDPR and the data usage structures, which results that the data is not utilized due to careful consideration and handling. Roche DC’s assessment of data in the context of its relevance and value to data consumers goes beyond missing data, and the misconception of the data usage certainly result in the data not being used or providing enough value to the tasks of the data consumers. This is stated explicitly by one of the data consumers: “I believe we can always become better at utilizing the data more efficiently. But with the new GDPR rules we are limited and also very careful on how we use the data.” (Interview 4: Anne, 2019).

Hence, from the empirical data, it can be deduced that an *appropriate amount of data* is received by the company, however, it is not aggregated further, and therefore does not provide sufficient *value* to the tasks of the data consumers. Incomplete data and inadequate structure are the main reasons for the data consumers’ complaints about the data quality not supporting their tasks. In order for these to problems to become solved efforts are needed in regard to communicating clear data usage guidelines to the company as well as using resources on maintaining complete and accurate data in systems. However, to solve these issues do not only mean to communicate appropriate and clear structures and guidelines for data usage, but also to understand what data is collected and categorize this data into meaningful categorize. When data is organized and labeled, data consumers will also be able to better create an overview of relevant data supporting their particular tasks and missions. To support the contextual data quality patterns, the data that is collected by the company and processed by data consumers needs to suit the company’s visions and missions, but also be aligned with the circumstances, rules and procedures practiced in the organization.

5.1.3 Representational DQ Pattern

The third category of dimensions is called representational and suggests that the system must present data in an intelligible and clear manner. (Coleman, 2013). For the data consumers to conclude that the data are well represented, data must not only be concise and consistently represented, but also interpretable and easy to understand. (Wang & Strong, 1996). The dimensions in this category, therefore, include: interpretability, ease of understanding, representational consistency, and concise representation. The empirical data collected throughout the interviews, highlights certain complaints about system structure and usage by data consumers. However, it was also indicated that the complaints about the system structures and usage depended on the person's role. As mentioned earlier, while the system facilitates ease of use for employees in sales roles, it presents itself as more complex for employees handling orders and providing service. This concern is stated by one employee from market access, who use the data to keep track of the connections with the municipalities and companies:

"I would like the CRM system to be more flexible and user-friendly. Both in terms of how to extract data, but also in terms of how fast it works. We assume that the system is used both by me and my colleagues on a daily basis, therefore it should include a high usability and speed, otherwise one is not motivated to use it as a data collecting system. So, the data itself does not need improvements, but the system does in terms of better usability and higher speed." (Interview 7: Karsten, 2019).

The data is represented correctly, however, the system is too complex for users to interpret and make use of the data within timely frames, due to complex system structures. Consequently, the system affects the concise representation of the data, which affects the usability of the data, leading to the data consumers not being able to benefit sufficient value of the data, and therefore, once again, considering the data of being of poor quality and not providing enough value for their tasks. As Ken Orr (1998) suggests, if data is not represented in a clear manner, chances that data consumers will make use of it will decrease. Therefore, if data consumers are facing challenge when working with the data due to complex system structures, the data quality will also be considered low, because the data will not be used in a consistent and effective way.

As mentioned earlier, the ease of use of the system depends on the role and task of the data consumers, however, the ease of understanding depends on the data sources: "The data that we receive from patients is easy to interpret and understand. However, the data we get from hospitals and doctors can be more difficult as they are using another terminology than we do here in Customer Care. (Interview 1: Casper, 2019). Roche DC's visions for data representation is data self-management, referring to patients being able to manage their own data within the systems. The company is working toward a solution where patients in connection with the free choice[1] arrangement is able to administer and manage own data visibility and accessibility independent from the organization or the municipality. The idea is to construct the company's e-commerce site in a way so that patients who have the free choice could order own blood glucose monitor and strips without first going through customer service. This way they can "manage their own products and orders as well as decide what data should be able to be visible and what data should be private." (Interview 2: Mette, 2019). For this to happen it is important that the company construct the system or website in such a way that it guarantees the data for always being represented in the same format and are compatible with previous data, subsequently, it means that it ensures *representational consistency*. In order for patients to successfully manage their own data it is important that the system provides clear structures and navigations, otherwise patients will not be able to construct concise representation of the data – and once again, not gain sufficiently value from the data. The problems of the data representational issue will be of little-added value to the data consumers as well as the overall company. If the self-management data vision becomes reality, the patients will be considered as the data producers, while the company will be considered as the data consumers. Therefore, in order for both parts to successfully interact with the data and make it support their tasks, a strong navigation as well as clear system structures are important.

5.1.4 Accessibility DQ Pattern

The last category of dimensions represents accessibility pattern and emphasizes the importance of the role of system-accessibility, understood as the extent to which data is available to or obtainable by the data consumer. (Coleman, 2013). It also indicates that data must be secure and not accessed by external. The dimensions include: accessibility and access security. In Roche DC accessibility DQ problems is presented in two ways, namely, confidential

[1] In Denmark all People with Diabetes (PwD) have the right to free choice of meter/strips paid by the municipality. However, the municipalities more and more often limit the choice of meters/strips available via tenders. In spite of this the PwD still has the right to choose the meter/strips he/she wants, but the municipality is only obligated to pay the amount of the meters/strips available in the tender. If the desired meter/strip is more expensive, the PwD must pay the difference out-of-pocket.

accessibility and data misinterpretations. First accessibility issue is characterized by underlying concerns about confidential accessibility. Secondly, data-representation issues, are sometimes interpreted by data consumers as accessibility problems. Roche DC's accessibility DQ concerns are mostly related to the confidential nature of patient records. Since GDPR requires consent in order to use any kind of patient data collected, data consumers tend to describe it as accessibility problems. The amount of data received is huge, however, since consent is needed in order to process it any further, it is described as accessibility problems. When data cannot be used, employees interpret it as not having access to it. "We receive a lot of data and I think the data is of a high value, but I would like to be able to send out more newsletters. We cannot access or use the data if we don't have consent." (Interview 4: Anna, 2019). In a healthcare organization; consents mean customers. As many consent declarations the company has as many customers or patients it can reach out to with marketing materials and campaigns. With the implementation of GDPR and the huge cleaning of the databases Roche DC has lost several thousands of consent declarations and as stated: "by losing so many consent agreements we have lost many customers as well." (Interview 4: Anne, 2019). Consequently, data consumers in marketing, find compliance to be a restriction to data accessibility and usage. Even though data consumers realize the importance of access security for patient data, they still perceive the permissions as barriers to accessibility. "And at the moment the biggest challenge is to figure out a way to collect consent agreements documenting that we are allowed to reach out to patients." (Interview 4: Anne, 2019). This in turn, has an effect on the overall reputation and value of the data.

Additionally, the representational DQ pattern can be discussed further in this section. As mentioned, when Roche DC migrated from Dynamics into Salesforce, a mixed attitude towards the change occurred. While for some, Salesforce facilitated many advantages, for others, it was presented as a complex and time-consuming system with vague structures. The employees who handle data about orders and market-shares find the data represented in the system difficult to interpret and understand. Most of the data, which comes from doctors are necessary to code into systems for product and service activities. However, the expertise required to interpret codes acts as barriers to accessibility, since the codes are not comprehensible to employees who are not familiar with the system. Analysis and interpretation of data, therefore, present as data interpretability problems. As stated by one employee: "… Second, you need to know what the symbols or numbers are representing. For instance, what does the number 0-7877-23 mean? In order to be able to use this data you need to know the hidden message or real meaning of it, which many in the organization don't." (Interview 2: Mette, 2019). Due to lack of communication and knowledge sharing, many employees do not know of the instructions that are making the data interpretable and understandable. Hence, the data becomes inaccessible to data consumers, because it is not in a representational form that permits insight for employees who are not aware of the coding procedures and meaning of it. In order to solve this problem clear communication and knowledge sharing is essential and will help employees to become aware of organizational structures and practices.

5.1.5 DQ Patterns in Roche DC

Within the intrinsic category three motives can cause data accuracy challenges. One is change; since data in databases often are static, while the real-world keeps changing, it leads the data to become incorrect if it's not used and given feedback to. Second is external factors, such as human mistakes, when processing the data. The third cause is patients who be failing to provide the right data. These three factors are what can cause data quality concerns and needs to be considered in order to improve the data quality from the intrinsic point of view. The contextual category discusses data incompleteness and inadequately defined structures. Data can present itself as incomplete as a result of the system design. Roche employees find the data to be missing or incomplete due to the systems size and complexity. Meaning that, even though the data is there, it can be difficult to use it properly. The second issue is inadequately data-usage structures provided by Roche global. Since, the GDPR has been implemented employees have been missing instructive guidelines that provides clear structures for data-usage. The lack of knowledge on how to use the data in line with compliance consequences that employees avoid making fully use of the data. Within the category of representational data quality, the design of the system can challenge the employees to find and process data. Especially, employees who handle data about orders face it to be complex and time-demanding. The last category of dimensions is accessibility patterns. Roche DC accessibility concerns are caused by two ways. Confidential accessibility and data-representation issues interpreted by employees as accessibility problems. In an extension of the above GDPR discussion, employees find the compliance site of the data-utilization as a limitation to accessibility. Another cause is data representation issues, which are interpreted as accessibility problems by employees. Due to a lack of shared understanding of the data coding and symbols, not all employees are aware of the real meaning of the data within the system, hence, interpret the data to be inaccessible.

When measuring data quality each of the twenty dimensions are considered and reflected upon. In order to be able to define the data quality it is necessary to analyze the data processes in line with the dimensions. Accordingly, the aspects of data quality and the finding hereof are reviewed further in the discussion part.

5.2 PART TWO: Transforming Data into Knowledge

In the first section of the analysis the data quality in Roche DC was assessed by help of dimensions such as, accuracy, timeliness, interpretability and accessibility. This was done in order to provide a more complete perspective on the company's data production and utilization processes. In order to understand the research question further, this second section of the analysis aims to take a look deeper into the data-information relationship by exploring how data and information are transformed into knowledge for knowledge creation, decision-making and patient support. For this purpose, the DIKW pyramid is used to analyze and discuss Roche DC's information and knowledge management processes. The theoretical discussion will have two major branches; information philosophy, focusing on the nature of information and knowledge management, which contributes to the notions of knowledge. Data and information will be thoroughly considered to understand the knowledge production and sharing.

As mentioned in the theory section the original model defines data, information, knowledge, understanding, intelligence and wisdom and explores the processes associated with the transformation between these elements. This is also the structure, this following analysis will go through, though, some minor exclusion of elements and changes are made in order to make it most relevant for the research question. Elements like, understanding and intelligence will be eliminated since they are not considered relevant for the case and the research question. Wisdom is considered but is not considered as the element providing the most insight to this case. Wisdom is located at the top of the hierarchy. Down from wisdom there are knowledge, information, and, at the bottom, we have data. Each is dependent on the category that fall below it – for example, there can be no wisdom without knowledge and no knowledge without information and vice versa. Following there will be provided an analysis of Roche DC in accordance to the elements that are described. Each element will describe the data-information case in Roche DC.

5.2.1 Data in Roche DC

It is stated that data are not structured, they do not convey any meaning and there are no built relationships between them. Every point of contact with the company and the health care system generates information and data about the patient. Data is collected from multiple streams and managed differently in Roche DC. Whilst most valuable data are interpreted to indicate the patient condition, a lot of the data is also interpreted to represent stakeholder aspects, such as hospitals and municipalities. While some of the data is of primary nature (raw data) some of it is also secondary data that comes from the registers. Table 2 presents a sample of patient data extracted from the system.

Table 2: Patient Data Sample /Roche Diabetes Care

S	T	U
I	**Asset Name**	**Product Name**
0	ACCU-CHEK Mobile (U1) meter mmol/L	Accu-Chek Mobile (U1) meter mmol/L
0	ACCU-CHEK Aviva Nano meter mmol/L	Accu-Chek Aviva Nano meter mmol/L
0	ACCU-CHEK Aviva Nano meter mmol/L	Accu-Chek Aviva Nano meter mmol/L
0	ACCU-CHEK Insight insulin pump	Accu-Chek Insight insulin pump
0	ACCU-CHEK FastClix II lancing device	Accu-Chek FastClix II lancing device
0	ACCU-CHEK Spirit Combo insulin pump	Accu-Chek Spirit Combo insulin pump
0	ACCU-CHEK Spirit Combo insulin pump	Accu-Chek Spirit Combo insulin pump
0	Accu-Chek Spirit Pump Kit 6yrs INTL/en	Accu-Chek Spirit Pump Kit 6yrs INTL/en
0	A-C LinkAssist	A-C LinkAssist
0	Accu-Chek Spirit Starter Kit hu	Accu-Chek Spirit Starter Kit hu
0	ACCU-CHEK Insight insulin pump	Accu-Chek Insight insulin pump

The data that is collected in Roche DC is interpreted to information and presents as essential for marketing tasks, product development, traceability and patient treatment. Therefore, they are collected from various sources across the Danish healthcare landscape and commonly stored in organizational databases, in electronic documents as well as hard-copy documents. Data is quickly made interpretable, since it is immediately entered to the system and given structures to. Majority of the times, customer service is the first point of contact for data. Different methods are used in customer service for data collection, however, all methods collect the same kind of data, namely, patient data.

5.2.2 From Data to Information in Roche DC

As defined in the historic DIKW model, information represents structured data. The role of information is to inform someone about something, hence, satisfy an information need. Information systems generate, store, retrieve and process data. The generation of data in Roche

DC requires the interpretation and analysis of raw data. Many different methods are used during this process, depending on the choice and aim of the information need. Often procedures specified for statistical data include the stages of checking, contextualizing, categorizing, standardizing and harmonizing the data. Table 1 above, which provides an overview of the data streams, in fact, provides an overview of structured data, which conveys a meaning for data consumers.

In order to provide an overview of the data generated by the company and the purposes of it, a table is created below.

Data Stream	Data Type
ICC – Call Center (Customer Service)	**Patient data:** Name, address, email, number, type 1 or 2, treatment information, medical device information: pumps/ BS monitor, hospital assigned, all interaction related to product complaints. **Used for:** • Marketing activities • Traceability • Product complaints
HOSPITALS	**Data Type:** Hospital name, address, number, contact person, segmenting on title (doctor or nurse) **Used for:** • Payment: Who pays for the product? • Responsible for the patient treatment.
PRIVATE COMPANIES	**Data Type**: Payer; in connection to our healthcare solutions only and supply; contact person, mail address, phone. **Used for:** • Deciding which solutions should be included in the supply of the healthcare solutions.

PATIENT EQUIPMENT	**Data Type:** Date for receiving the tools, expiration date, complications in regard to the tools, ordering. **Used for:** • Marketing purposes.

Table 3: What types of data is generated and used for?

The definition of data in the DIKW model fits well with Roche DC's data too. The data that is collected in Roche DC are often found in three ways:

- Patient data is all data about the patients. Within patient data personal or sensitive data is presented, which signifies confidential facts about patients, such as disease, sexuality, religion, political beliefs etc. Patient data is often multi-dimensional and include location, numbers and types, and is often collected through surveys. The synthesis and interpretation of patient data results in the production and support of marketing materials, traceability and product complaints.
- Public data relates to hospital agreements. This data represents everything about the hospitals that Roche DC has a collaboration with. It includes the hospital name, address, number, contact person (segmenting on title e.g. doctors or nurses) etc. The data is received in a statistical form and interpreted to describe the patterns and create market entrance strategies.
- Company data is external data about the companies, which Roche DC is dealing with. This data is also represented in a survey form and interpreted can reveal facts about external stakeholders.

5.2.3 Knowledge Sharing in Roche DC

As mentioned in the theoretical framework knowledge as defined by the DIKW model is contextualized information and is produced by combining available information with expertise, insight and intuition. (Petavratzi, 2017). Knowledge in Roche DC is know-how and is what makes possible the transformation of healthcare information into instructions and strategies for outcomes. In order to generate useful knowledge, it is necessary to combine information with human thinking. Roche Global employs the Enterprise Social Network, Yammar to foster conversations across the Roche DC global boarders. Employees from all around the world sparks health related conversations through different communities in different self-created Yammer-groups. These groups can also be called communities of practices. Similarly, in Roche DC DK, knowledge is created in different social and formal settings, such as through meetings, interactions, discussions and shared experiences. However, the empirical data show that despite the company's wish to encourage knowledge-sharing employees in different departments tend to not always receive all relevant information. An example of this could be that often employees in the department of customer service is handling information, which could provide huge value for people in marketing. However, this information is not always transferred, communicated or shared to other departments than the department of customer service. This observation derives from the fact that all the people interviewed had diverse level of knowledge and interpretations of the GDPR rules as well as described the data sharing and using quite differently. While some described the process and access of data in their roles as very strict and confidential, some people, with access to the same data, had a more lose approach to it.

5.2.4 Wisdom in Roche DC

The top of the DIKW proposed model is wisdom, which is said to represent the ability to increase effectiveness. Wisdom is only reached after greatly processing data, information and knowledge, and as mentioned the process starts in the bottom of the hierarchy with data. (Rowley, 2007). The higher elements in the hierarchy can be explained in terms of the lower elements by identifying an appropriate transformation process and in order to increase wisdom it is important to consider and process the data and information correctly. By building and promoting communities of practice that can enhance knowledge sharing, Roche DC will create structures that increase effectivity and influence good patient decision-making. Even though wisdom does not act as the critical element, it is still considered in the analysis. The discussion of wisdom in management literature is in the context of leadership and is seen as an important characteristic to business leaders. In order to generate wisdom, Roche DC must encourage and foster an open information culture where any information need is satisfied among employees.

The importance of a good leadership that encourage employees to contribute and inspire is what supports wisdom creation.

5.2.5 From data to information to knowledge

Having provided an overview of the concepts and related them to Roche DC, this section aims to examine the transitions between each element to gain a better understanding of the transformation of data into useable knowledge in the company.

Majority of the sources that has been studied discuss information as “Data with meaning”, and this definition is also adopted in this analysis. Pure data that is collected in the customer service department is mostly with no structure, and for this data to become information it is shaped and structured by the receiver and employee. “We receive patient data such as patient names, addresses, e-mails, numbers, age etc.” (Interview 1: Casper, 2019). This data that is received is processed, interpreted and entered in a structured database, which makes it to useful information for data consumers to gain insights from. However, since most data is not utilized efficiently: “At the moment we don’t really use it for a lot beside formalities” (Interview 1: Casper, 2019), another concept like data mining can also be discussed in this section. Data mining refers to the application of specific algorithms for extracting patterns from data and discover knowledge in databases. (Fayyad et al., 1996). In order to transform the data into useful information, Roche DC can use the steps involved in the data mining process. Whether it is data about sales, marketing or it is product related Roche DC can extract and transform data into a data warehouse. This data can be stored and managed in multidimensional databases and analyzed by using application software. The result of this process is analyzed data in an understandable form, which can be further used to analyze current trends and changes for decision-making and patient support.

So how is information further converted to become knowledge in Roche DC? Communities of practices has been discussed as one way to increase knowledge creation and sharing. In recent years, the progress of artificial intelligence and knowledge-based systems has inspired knowledge revolution in many fields and domains. (Petrides, 2002). However, the revolution of knowledge still needs human systems to understand and make use of it. An intuitive way for

Roche DC to make the information to useable knowledge could be by investing in building and promoting communities of practice within the organization. Communities of practice will not only allow members of Roche DC to work together on data analysis to generate new information, they also help community members think together to generate new knowledge, on both the individual and the community level. According to Davenport & Prusak (1998) knowledge is created through activities taking place within and between humans. While data is found in records and transactions, and information found in messages, knowledge is obtained from individuals or groups of knowers and in organizational routines. The empirical data shows that departments in Roche DC are not always aware of the activities going on in the different sections, therefore, low information and knowledge sharing across the departments are assumed. However, the empirical data also shows that global, in turn, invests more resources in information-sharing and knowledge-creation. As stated by the Global Business Process Manager: "We see a big value in having a global database of providing insight across borders of what are trending and what are the hot topics in the market and what patients and healthcare professionals talking with us about, so we can recognize these patterns and turn them into actions of matters in, for example, product development." (Interview 5: Julia, 2019). The abovementioned quote describes an excellent example of knowledge creation resulting from information sharing across the different divisions. As mentioned previously, the data that is collected in Roche DC and turned into information is not always processed to reach their full potential of use. Employees in one department has not always access to data in another department. This is due the compliant aspect of a healthcare company. Therefore, information sharing, and distribution of knowledge could happen by fostering communities of practice. People from different fields and with access to different data could come together to develop members' capabilities by building and exchanging knowledge with one another. Since the strength of communities of practice is self-perpetuating. As employees in Roche DC will generate knowledge, they reinforce and renew themselves. (Wenger & Snyder, 1999).

Within the healthcare industry Knowledge Management is becoming established as a core organizational element to contribute in the delivery of better patient care. (Baskaran et al., 2006). In a knowledge aware healthcare situation, knowledge creation should be a fundamental activity. In order for a service driven business such as healthcare can benefit, new opportunities to create knowledge have to be properly initiated. Artificial Intelligence could also be considered in the light of knowledge creation. The task of AI would make use of intelligent assistants to build knowledge-based systems that can solve data/information problems in their own and provide data consumers with accurate and timely information. Advantages of AI have been greatly discussed in the healthcare domain. By use of sophisticated algorithms AI can 'learn' features from large volume of healthcare data and use the insight to assist in knowledge production to inform proper patient care.

Data is interpreted to information and information is transformed into knowledge. Knowledge is used to make wiser decisions about strategy, customers, competitors and products. To increase efficiency and foster a strong culture where knowledge is used and shared among peers management staff in Roche DC must invest time and money in helping communities reach their full potential and the best way to do so is to assess the value of a community of practice by regularly listening to members' experiences in an organized way. Data mining and artificial intelligence are also efficient ways of transforming the data and discovering insight and knowledge. A further assessment of the concepts and methods is provided in the discussion part.

5.3 Final conclusion on analysis

The analysis is divided into two parts. The first part aimed to evaluate the data quality patterns in Roche DC and understand to what extent the data that is collected is error-free, timely and used properly for tasks and projects. The data quality was assessed throughout the twenty dimensions provided in Wang & Strong's framework and major findings that can affect the data-usage include, **1.** lack of clear data-usage structure and GDPR guidelines. **2.** Data in databases are static, while the world is in constant change. **3.** The system-design challenges the data consumers to process and use the right data to their tasks. The second part of the analysis intends to explain the transformation process from data to information to knowledge to wisdom. Through the DIKW pyramid, the main focus is to analyze and understand how data in Roche DC is processed and useful knowledge is created supporting patient care decisions. The last phase of the analysis leads up to the discussion part by reviewing concepts like communities of practice, data mining and artificial intelligence for knowledge creation and knowledge management. The following table creates an overview of the data quality problems discovered in part one and further evaluated in part two. The observed quality problems perform as a starting point to the following discussion.

Table 4: Generic data quality problems

D.Q. Category	Observed Data Problem
Intrinsic	Change Human mistakes Patient mistakes
Contextual	Data incompleteness Inadequately structures
Representational	Systemdesign
Accessibility	Confidential accessibility Data representation accessibility

6 DISCUSSION

The purpose of this chapter is to discuss the findings from the analysis, in the light of the managerial implications and potential data quality improvements. The discussion is organized around five themes, which all emerged from the analysis. The themes in this section derives in a chronological order as they were reviewed in the analysis.

The themes derived from the interviews and analysis are:

1. Inadequately data-usage structures provided
2. Unused data cannot remain correct for a long time
3. Practice as an element for knowledge sharing
4. Flexible systems result in high quality data
5. Artificial Intelligence for knowledge management

6.1 Inadequately data-usage structures

Five out of seven employees mentioned lack of data guidelines and structures as a reason for not using data sufficiently. Afraid of utilizing data in an incompliant way the data is rather not used for more than what is on the safe side. Even though Global proposes that guidelines and strategies are communicated to all countries on a regular basis, employees within Roche DC DK do not share the same conception and still seem to lack knowledge on what and how to use most data. However, the data-usage confusion is also reinforced by the implementation of the GDPR rules. The GDPR rules has created fuzzy data-usage boundaries and decreases data quality, because most data become redundant as a result of not becoming exploited properly. The data also indicated that most of the concerns about not using the data in a compliant manner was in relation to marketing and customer activities. The problem is that GDPR acts as more than only a set of rules to follow. Rather it is a comprehensive approach to consumer privacy and data security. Here, Roche DC should consider looking deeper into Article 5[2] of GDPR, which addresses how contact information is handled and how you can use a data quality approach, which involves both tools and processes as part of the company's compliance efforts. (Article 5: Privacyplan). The main requirements of GDPR's Article 5 contain appropriate

[2] GDPR Article 5 refers to the principles relating to processing of personal data

usage, accuracy and data security. It commands that steps must be taken to ensure that personal data that are inaccurate are erased or rectified without any delay (Article 5: Privacyplan). A further assessment of the tools and processes provided in Article 5 is discussed in the next section.

Kerr & Stockdale (2007) argues that high data quality is a result of an effective decision-making in organizations and the other way around; raising the level of data quality within an organization contributes to improving the quality of decision-making, enabling the reduction of uncertainty and the production of more timely and accurate decision outcomes. (Kerr & Stockdale, 2007 p. 2). However, data quality is not a widely discussed topic in Roche DC among the leaders and limited attention and resources are paid towards assessing data-usage or providing employees with clear data guidelines and definitions. Furthermore, the authors also point out the context of the data collection for contributing to data quality problems. The context of data collection is the intended purpose for which data are captured and the policies and procedures that govern acquisition, storage and usage. Since quality requirements, such as the demands for accuracy and timeliness may differ from one context to another, or change over time, problems arise when data consumers are not informed of the context, and therefore, make incorrect assumptions. This is especially significant in Roche DC where data are collected and interpreted from multiple sources. The empirical data and the analysis of this thesis agree upon the statement derived from Kerr & Stockdale's research that lack of knowledge on data-usage as well as clear GDPR guidelines prevent employees to use the data. Therefore, data quality assessment and data usage must be first considered at the management level and management staff must deliver employees clear defined procedures for data handling and utilization.

6.2 Unused data cannot remain correct for a long time

Four out of five local based employees discussed data in databases as not always being accurate and timely enough due to constant change of jobs, addresses or e-mails. As Ken Orr (1998) suggest; in order to maintain high data quality using the data is necessary. It is common practice to collect large amount of unused data based on the idea that someday it might be used. However, the data in databases becomes static, and over time, real-world changes, which leads to the quality of the data in the system decreases. This proposal is reflected in Roche DC's data collection and usage as well. The data accuracy is affected when patients change personal information without notifying the company. And this is earliest noticed by the company when it sends out marketing materials such as newsletters and experiences that e-mails are bouncing. Therefore, constant feedback and updates are needed in order to have accurate data. As mentioned in the previous section, Article 5 discusses these patterns and provides tools and processes that ensure to maintain data that is updated in a compliant way.

Automated validation of contact data is essential in the GDPR Article 5 compliance. The tools that Roche DC can use include address validation, which compare contact data against international postal and contact databases. Lead validation can also be used to compare several criteria to produce a total lead quality score to evaluate data accuracy. As mentioned earlier, the data seems often to be expired and is first realized when the company experience the e-mails' bouncing. Consequently, also e-mail validation tools can be useful for Roche DC as it can check for legitimate addresses as well as common misspellings (such as 'gmial.com' versus gmail.com). (Article 5: Privacyplan). The abovementioned tools can be implemented by Roche DC to ensure that data is timely and accurate for fast and continuous usage. The processes of GDPR include doing validation at both the time of data entry and at time of use. The time of use is especially essential as more than half of the company's contacts may become incorrect over time as a result of change. The interviews revealed that Roche DC does not use resources on processes to maintain data accuracy and timeliness due to lack of time and resources. One employee in marketing suggested that change of personal data could be updated when customer care did phone calls to existing customers. Since, this suggestion seems to demand more time and resources the validation tools and processes would be of higher beneficial for Roche DC to implement as they will guarantee data to be timely without putting in too much effort or time.

Ken Orr (1998) further discusses the primary role of information systems as being to present views of the real world so that people in organizations can create products or make decisions based on it. If the views presented by the system does not reflect the real world for any extended period of time, then the system is considered as poor and the organization will act irrationally. Therefore, constant feedback and update are needed as the quality of the data gets measured by the accuracy of the representation. This proposal reflects quite well the data representational issue in Roche DC also. While the company consider the data maintenance for being too expensive in terms of time and money, the cost of not having accurate data is higher. Oftentimes, the patient data and consents are used for marketing purposes, such as newsletters. If efforts are not made for keeping this data up to date the cost and resources used on the marketing activities are wasted since the message won't reach out to as many patients as intentioned. However, it is worth emphasizing again, that the concern with data quality is not to ensure that the quality is perfect without any kind of errors, but rather that it is accurate enough, timely enough as well as consistent enough for the organization and the data consumers to make sensible decisions. Therefore, the concern with the timeliness of the data in Roche DC is not only that it needs constant feedback, but rather to make sure that the efforts made in terms of marketing campaigns and newsletters reaches out to a sufficient amount of relevant people and creates a reasonable impact.

When it is stated that unused data cannot remain correct for a long time it means that data needs to be used in order to be accurate. Accuracy depends on a high extent on the usage. The reason for this statement is that when employees interact with data, they interpret and process it to develop it to match their tasks and needs. All interaction with data is data-processing and during this processing feedback is given to the data and the data can remain relevant to the organizational needs. However, if employees don't have much interaction with the data, because of the lack of guidelines/structures discussed in the first section, neither gets the data further-developed to support Roche DC's missions and strategies. The analysis also indicates that the department of compliance in Roche DC has not paid extensive attention on the solutions of some of the challenges that GDPR has brought, for instance, the data-usage concerns. Therefore, the validation tools and processes indicated in GDPR Article 5 compliance is, in this thesis, considered as a support to the data-usage and change challenges.

6.3 Practice as an element for knowledge sharing

The second part of the analysis endeavored to explore the transformation process of data into knowledge and a pre-discussion was started on how knowledge should be held and processed in order for employees to gain insight and the company to support patient care. Data is collected and given structure and meaning when entered in databases, and knowledge is created by using and interacting with the information. The DIKW model which was used to analyze the data streams and transformation in Roche DC defines data, information, and knowledge as terms directly build on one another. Data is supplemented with meaning and information is achieved. Information is interlinked; hence knowledge is created. Wisdom comes on the top of the pyramid and is not always achieved by knowledge. Wisdom adds value to information and requires judgement. However, in this thesis, wisdom is not considered as being relevant relative to what this research wants to look into, therefore, this element could be excluded. The analysis only focuses on the steps of data-information-knowledge, since these elements are what constitute the answer this project is looking for. Though the analysis also revealed that the data-information-knowledge was not just enough, but practice should also be considered. Therefore, if the DIKW model should be reconstructed to fit into this analysis, the element of wisdom would be replaced with the element of practice. Practice of information and knowledge sharing within the organization. Apart from data mining tools to categorize information and search for related patterns, Roche DC should create and encourage communities of practice for knowledge creation purposes. Data is transformed into knowledge at some point, however, knowledge without interaction won't produce great outcomes. A fundamental activity of knowledge should be to drive products and results in the organization, and it positively happens by practicing and sharing knowledge across departments.

By fostering communities of practice Roche DC will encourage knowledge sharing across the different departments. The thing with a community of practice is that its members stand for the connection of the organization and the network. The members lie within two stands of social relations; one, bound together by organizational membership, another bound together by practice. From the perspective of practice, it is possible to understand the flow of new knowledge into and out of organizations. Members within a community come from different departments, handle different kinds of data and possess different kinds of knowledge. In order for employees to share and create new knowledge interaction and practice are needed elements. High-quality data and derived information create institutional knowledge and reasoning processes that enable the organization to extract the maximum benefit from the resources. This approach is called 'knowledge management' and draws together the tangible and intangible elements of data and shares them amongst all workers in the company. (Davenport, 1998: Kerr

& Stockdale, 2007). Beyond knowledge sharing and knowledge creation, communities of practice can provide Roche DC with several other benefits, such as transferring best practices among workers, developing professional skills and help Roche DC retain talented workers.

Most of the data challenges, such as data interpretations, accessibility problems, and system difficulty are all challenges, that Roche DC could overcome by letting employees make use of one another's expertise and skillsets. Beside sharing knowledge, a self-made community within Roche DC will also allow members to share best practices across the company. Derived from the analysis; most workers in Roche DC do not know much about one another's fields and work tasks. Therefore, a community within Roche would not only make people engage, but also make them use one another's knowledge and capabilities. While one employee found the system to be without trouble another employee find the same routines in the system to bring confusion and be time-consuming.

Exchanging best practices will also lead to employees in Roche DC developing additional professional skills. Studies have shown that beginners learn as much from advanced learners as they do from their masters. (Wanger & Snyder, 1999). Therefore, meeting on a monthly basis and coaching one another to learn from different fields will increase knowledge sharing and either their skills will be further-developed, or new skills will be developed. The best IT, management or sales people don't rely on their own brilliance, but read peer-reviews and discuss research and knowledge with colleagues. Consequently, communities of practice will act as effective arenas for fostering professional development in Roche DC.

Apart from sharing best practices and developing professional skills by learning from one another, a community of practice will also allow Roche DC members to solve complex organizational problems using one another's expertise. Since, Roche DC is a multinational company operating in more than 40 countries, communities of practice could be made across the different markets. Meaning that instead of only creating local communities, Roche DC members could build global communities containing members from all around the world. The aim would be to use one another's skills and knowledge to create answers to local organizational problems and uncertainties.

6.4 Flexible systems result in high quality data

As shortly described earlier, majority of the employees within the Danish market expressed a certain concern with the flexibility of the system. Wand & Wang (1996) analyze data quality in terms that are not data-centric but oriented towards system-design. According to these authors the quality of data depends on the design and production processes involved in generating and processing the data. The notion of data or information quality depends on the actual use of data. What is considered as good data for one situation may not be appropriate for another situation. Hence, this relativity of quality presents a problem (Wand & Wang, 1996), since the quality of the data generated by the system depends on the design of the system. This claim reflects well the findings of this thesis. Most data-interpretation issues are results of the system's inability to present the data in an appropriate format. As mentioned, most employees consider system construction and design as inflexible and time-consuming. However, the main problem is that because Roche DC exists as a small unit as part of a gigantic corporation, they are subject to certain specific structures and preferences, which they, as a company, cannot change, but has to adapt to. Flexible system design result in high data quality because the system allows data consumers to access, interpret and use the right data for the right task at the right moment. Inflexibility within systems also creates data-interpretation problems as found in the analysis. When data consumers find it difficult to navigate or use the system properly they also tend to access the wrong data for their tasks. This leads to misinterpretation of the data. According to Strong, Wang & Lee (1997) system flexibility contributes to easy data accessibility and understanding, which result in better data utilization. The analysis indicates that coded medical data is technically accessible, however, data consumers often view it as inaccessible because they cannot interpret it easily due to complex system structure. In order for Roche DC to utilize the Salesforce system better, despite not having it customized into their local needs, the company should invest in creating clear and understandable structures for the system, so that employees can easier get access to the right data as well as interpreting it correctly. It is important to stress that in order for Roche DC to design information systems that deliver high-quality data, the view of data quality must be well understood. High-quality data must be defined clearly and understood well among all employee. As mentioned, in order to design for better quality, it is necessary to understand what quality means and how it is measured in the company among people.

6.5 Artificial Intelligence for knowledge management

According to Agarwal & Yiliyasi (2010) information overload has an impact on the freshness of information, while information overload and information freshness are both incidents that can impact data quality. With a rapid pace of content generation, it gets difficult to follow what information is real and what is outdated. The churn rate of information is high, and this is also realized in Roche DC. Throughout the analysis it was found that huge amount of data was collected, whether it was patient data or data about municipalities and business partners, however, not much of the data was actually used further. The reasons for not using the data counted several motives such as lack of knowledge about data-usage, structures and the system design. Because of the huge amount of data and information collected from the different streams, the information quickly overwhelms the employees and often the employees can face the dilemma of choosing freshness of results over accuracy. As data and information continuous to enlarge in the company the management of information and knowledge gets critical for Roche DC in order to keep up with relevant and useful data. The challenge of information and data overload on data quality is a well discussed topic especially within the global domain. Roche DC employees need to understand the structure and dynamics of data in order to identify the relevant content, and more importantly, to tackle the above-mentioned challenging issues. Therefore, the hot topic at the moment is Artificial Intelligence. Roche global has a future vision of using artificial intelligence to manage and handle the data that is collected, but not necessarily used. Were Roche DC does not want to invest in human resources sift data they endeavor in applying machine learning to help look into dark data[3]. The idea is that if the company has not visited certain accounts in their territory or not heard from customers for longer than six months artificial intelligence can sift to this data without any manual intervention needed. Thus, AI can propose action for a set of data, for instanc reaching out to patient by a simple call or sending an email to customers asking whether they still use their

[3] **Dark data** is data which is acquired through various network operations but not used in any ways to derive insights or for decision making

product. Roche DC aims to turn the dark data into data that can be processed for marketing campaigns and product development.

Beside using artificial intelligence for utilization of dark data it can also be used for information management and knowledge creation. Knowledge management plays an important role towards the success of transforming individual knowledge into organizational knowledge. (Liebowitz, 2001). As discussed briefly in the analysis section; knowledge management is more than tools and technology. In reality, it is people and culture that needs to be considered in order to develop a knowledge sharing environment. Similarly, as discussed previously, knowledge sharing is knowledge creation. If people are not willing to share their knowledge, knowledge management efforts will eventually fail. In looking at ways for sharing knowledge, transforming individual knowledge into collective, organizational knowledge, and transforming organizations into "knowledge organizations", the field of artificial intelligence can help push these principles of knowledge management. Roche DC could use knowledge discovery and data mining approaches, which are both AI-related methods, to determine relationships and trends in knowledge sources for creating new knowledge. The analysis of the company described one of the major issues to be lack of knowledge-sharing among employees. One of the basic functions of knowledge management is considered to be knowledge distribution and involves sending knowledge internally and externally to those who could benefit from the use and application of the knowledge. For Roche DC to benefit from the knowledge collected by individuals, an infrastructure within the company could be implemented, responsible to disseminate the knowledge to relevant individuals and departments. Instead of having a passive distribution mode where each individual worker has to access the organization's knowledge repository, it would work out more efficiently to have a knowledge management team in charge of analyzing the knowledge and distributing it to the relevant employees, management, customers, and stakeholders. To assist in this process, techniques such as intelligent agents could be applied to analyze the knowledge, email, personal information and web pages to disseminate appropriate summaries or individual pieces of information and knowledge to those who should make best use of it. Furthermore, data mining and knowledge discovery have already been discussed earlier and are also methods of AI, which could be applied to look for trends, relationships and new knowledge from the company's knowledge and data warehouses. Previously was discussed communities of practice as an effective way of sharing and creating new knowledge. Similarly, online communities are also considered successful and cost-effective ways of sharing and distributing knowledge across the organization. In order for online communities to become successful, workers have to engage and contribute by regularly interacting and sharing different types of knowledge by one another. Considering the size of Roche DC and the small number of employees located in the Copenhagen offices it may be a challenge to establish and manage self-created groups by workers themselves. Nevertheless, the problem of knowledge sharing has been considered by the local management team and an approach to solve it has been provided by introducing WhatsApp for informal knowledge sharing content. All workers as well as the management team are connected to the group and different professional and personal news are shared by everyone in the company. However, as mentioned, this knowledge is most often informal and does not contribute much to workers' daily tasks. The information shared in the WhatsApp forum is mostly enough to get a general idea of what is going on in the company. Therefore, online communities would be more effective to create relevant knowledge supporting people's professional assignments. Due to the local size of Roche DC, the online communities could apply for all global countries and collect valuable knowledge not only from workers in Copenhagen, but across the world. For instance, an online community for the marketing department could be created which could involve all the marketing people in Roche DC across the whole world. The same for teams such as Sales and Product Development could be created with the purpose of sharing knowledge and expertise. As mentioned, this idea is an extension of the idea of communities of practices discussed in the earlier section and could operate as part it. The methods and tools for knowledge management are many to be considered. What is important is that knowledge management should be treated as a strategic goal in organizations for better capturing and leveraging knowledge internally and externally.

7. CONCLUSION

This thesis has considered the concepts of data quality and knowledge creation and attempted to explain how data is transformed into meaningful information in Roche Diabetes Care, and how knowledge is created. The analysis is divided into two parts, whereas the first part measures the data quality in the company and the second part explores the data-information relationship and discloses the data, information and knowledge sharing, which leads to the discussion. In the first part of the analysis, four major data quality concerns were found. Factors such as, change, human mistakes, data incompleteness, inadequately structures and system design, all lead to data quality issues, which affect data consumer's usage of the data.

Change is one important incident that degrades data quality. Since data in databases are static and the outside world keeps changing it is important that data is used and regularly provided feedback to. The empirical data highlighted that a great amount of data was collected in Roche DC, however, often it was not used, and therefore, over time, the validity of this data expired. Secondly, factors such as both human and patient mistakes also lead to data quality issues in the company. Employees described one of the most often data-inaccuracy moments to be when patients fail to provide with the right answer or employees fail to enter the exact answer. This leads to data consumers experiencing the data to be incomplete when using it for their tasks. Furthermore, a major reason for data consumers not using the data is inadequately structures. As a result of the new GDPR rules, employees are not fully clear about how to use the data. Poorly data structures are provided, which results that employees choose rather not to fully use the data because they are afraid of using it incompliantly. Last, the system design also affects the data-usage, which affects the data quality. The Salesforce system can be complex and time-consuming for some employees, while acting as convenient for others. The empirical data showed that data representational issues were most often related to the system being incapable of presenting the right data to the right people at the right time. The discussed factors are considered to influence the quality of the data in Roche DC, however, the analysis also showed that they are not assessed or explicitly conversed among employees and management.

The second part of the analysis focused on the DIKW model and attempted to explain the transformation from collecting data to processing information and creating knowledge. With the DIKW pyramid, the main focus was to analyze and understand how data in Roche DC is processed and useful knowledge is created supporting patient care and management decisions. The findings in the second part contributed to the discussion of this thesis by reviewing concepts like communities of practice, data mining and artificial intelligence for knowledge creation, sharing and management. One of the major findings the overall analysis revealed was that the company did not utilize fully the data that they collected, meaning a lot of data was wasted. Furthermore, the analysis also revealed that information sharing in the company is low. Therefore, it is discussed that the company should invest and practice knowledge sharing in self-created communities. It is suggested that the strongest way to enhance knowledge creation is to encourage knowledge sharing across the different departments. Employees in Roche DC, each have access to different kinds of valuable data and knowledge, however, most often this information and knowledge is not shared among peers. Therefore, this thesis discusses the element of wisdom in the DIKW model to be replaced with an element of practice. Data is transformed into information, which is further processed to knowledge, and in order for the company to get insight from this knowledge, it is discussed that it has to be practiced, shared and communicated among employees.

Conclusively, it can be argued that, in order for Roche DC to turn the data that is collected into meaningful information where knowledge creation in the company is positively influenced, it has to encourage employees to share data, information and knowledge – not only locally, but also globally. The company should invest resources in tools and procedures that enhances knowledge sharing among employees and management. This thesis argues that one of the most effective way to increase knowledge creation is to enhance knowledge sharing.

7.1 Suggestions for future research

This thesis mostly has considered the data quality and information usage internal the organization. All the interviews that were conducted and observations and research made had an internal focus within the organization. It would be interesting to look more into external conditions and explore how the data quality and information distribution can improve, to better support patient care decisions. For instance, it would be interesting to change the research question to become "how can Roche Diabetes Care turn the data into meaningful information where patient care decisions are positively influenced?”. The research will then focus more on data quality supporting patient care decisions.

Lastly, shifting paradigm, it could be interesting to test the information quality challenges in Roche DC by implying quantitative analysis. By narrowing down the focus and only consider patients, the study could look into patient data samples and use machine learning to discover different trends, patterns and behavior for marketing purposes.

7.2 Recommendations for Roche Diabetes Care

Throughout this thesis it has been discovered that the data that is collected in Roche DC has a much bigger potential than realized and utilized by the company. Therefore, this section aims to provide suggestions for how the company should handle their data strategy to gain the most advantage.

1. **Definable structures must be provided by management**

   One of the major findings discussed in this thesis is the employees’ lack of knowledge of how to use the data and for what purposes. Each and every employee emphasized the huge and actual potential of the data, that Roche DC collects from the various channels. Nevertheless, each and every local employee also stressed their concerns about the compliance side of the data-usage. From the analysis it is deduced that even though global has communicated clear structures and guidelines to local entities, Roche DC DK has not yet managed to transfer this information and knowledge to the whole Danish organization. Therefore, in order to successfully use the data and information, the first step must be for management to focus on constructing clear and understandable guidelines for the local employees to make use of. The paper addresses the quality of data to be dependable on the data consumer’s usage of the data. And in order for employees to make use of the data to support their tasks, they need reliable and trusted data structures. Therefore, based on the findings, the first suggestion for the company is to create a sustainable and compliant data-usage strategy and communicate it clearly to the entire local organization. Only when data consumers are aware of the strategy and guidelines, they actively act to discover more about the data potentials.

2. **Using resources on keeping data up to date**

   Another major finding which was discovered in this thesis is the timeliness of the data in the databases. The analysis revealed that because of constant change in the dynamic world the data in the databases needs constant feedback. When a solid and sustainable data strategy is created and clearly communicated to the whole organization, the next step would be to use resources on filtering out the amount of data, so it is up to date and timely enough for supporting data consumer's tasks. The system allows Roche DC to log, extract and change data, but not to delete anything. This of course leads the company to have huge amounts of unused and inaccurate data maintained in the system. Therefore, the second suggestion for the company is to use resources on managing the data so it's constantly up to date and timely and the information about patients and stakeholders are consistent and true. When the marketing team sends out newsletters to patients, a big amount of the e-mails is bouncing, which means that the cost of the marketing initiatives is higher than the actual affect. Management of Roche DC must therefore not only create understandable data structures but also allocate resources on data maintenance. As considered in the discussion part Artificial Intelligence could be one affective solution. To avoid investing in human resources sift data, Roche DC can apply machine learning to manage and sort relevant data that is up to date. In order for Roche DC to use the data effectively, it is essential that the data is trustworthy and accurate. Furthermore, AI can enhance direct marketing initiatives by proposing action for a set of data to automatically reach out to patients after a certain amount of time.

REFERENCES

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

Agarwal, N. & Yiliyasi, Y. (2010) Information Quality Challenges in Social Media. Department of Information Science, The University of Arkansas at Little Rock.

Baskaran, V., Hisbel, A., Naguib, R. & Walls, M. (2006). Knowledge Creation Using Artificial Intelligence: A Twin Approach to Improve Breast Screening Attendance. EMBS Annual International Conference. New York City, USA, Aug 30-Sept 3, 2006.

Bryman, A., & Bell, E. (2007). Business research methods (Second Ed.). Oxford: Oxford University Press.

Clampitt, P.G. (2012). Communicating for Managerial Effectiveness: Problems, Strategies, Solutions. Thousand Oaks, CA: SAGE Publications.

Davenport, T. & Prusak, L. (1998) Working knowledge: how organizations manage what they know. Ubiquity: An ACM IT Magazin.

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. The Academy of Management Journal, 50(1), 25-32.

English, L. 1999a. *Improving Data Warehouse and Business Information* Quality. New York: John Wiley & Sons.

Gustafsson, J. (2012): Single case studies vs. multiple case studies: A comparative study. Academy of Business Engineering and Science. Halmstad University.

Jordansen, B., & Madsen, J.,Astrup. (2010). Introduktion til videnskabsteori. In P. Darmer (Ed.), Paradigmer i praksis: Anvendelse af metoder til studier af organiserings- og ledelsesprocesser (pp. 37-73). København: Handelshøjskolens Forlag.

Kahn, B., Strong, D.M. and Wang, R.Y. (1997) A Model for Delivering Quality Information as Product and Service, Proceedings of the 1997 Conference on Information Quality, Massachusetts Institute of Technology, 80-94

Kenn Orr. (1998) Data Quality and Systems. Communications of the ACM. Vol. 41, Nr 2.

Kerr, K. & Stockdale, R. (2007) Data quality information and decision making: a healthcare case study. 18th Australasian Conference on Information Systems. 5-7 Dec, 2007, Toowoomba.

Lee, T. W. (1999). In Williams L. (Ed.), Using qualitative methods in organizational research. Thousand Oaks, California: SAGE Publications.

Lee, Y. W., Pipino, L., Strong, M. D., & Wang, Y. R. (2004) Process-Embedded Data Integrity. Idea Group Publishing.

Liebowitz, J. (2001) Knowledge management and its link to artificial intelligence. Elsevier Science Ltd.

Ma, X. Data-Information-Knowledge-Action Model. In: Schintler, L.A., McNeely, C.L. (eds.) Encyclopedia of Big Data. Springer, Cham, Switzerland

MacKay, D.M. (1969), *Information, Mechanism and Meaning*, MIT Press, Cambridge.

McFadden, F. R., Hoffer, J. A., & Prescott, M. B. 1999. *Modern Database Management* (Fifth ed.). Reading: Addison-Wesley.

Orlikowski, W., & Baroudi, J. (1990) Studying information technology in organizations: research approaches and assumptions. Center for Research on Information Systems. New York University.

Petrides, L.A. (2002). Organizational learning and the case for knowledge-based systems. *New Directions for Institutional Research*, 2002 (113), 69-84.

Pipino, L., Lee, Y. & Wang, R. (2002) Data Quality Assessment. Communications of the ACM. Vol. 45 No. 4v.

Prusak, L and Davenport, T., Who are the gurus' gurus? *Harvard Business Review*, reproduced in *The Weekend Australian Financial Review* (10–11) (2004) 25.

Redman, T. C. 2001. *Data Quality. The Field Guide*. Boston: Digital Press

Rowley, J. E. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science*, 33, 163– 180. https://doi.org/10.1177/0165551506070706

Rowley, J. (2007) The wisdom hierarchy: representations of the DIWK hierarchy. Journal of Information Science, 33 (2) 2007, pp. 163-180

Shanks, G. & Corbitt, B. (1999). Understanding Data Quality: Social and Cultural Aspects: Department of Information Systems. The University of Menbourne.

Shanks, G. and Darke, P. (1998) Understanding Data Quality in a Data Warehouse, The Australian Computer Journal, 30, 122-128.

Sharma, N. (2004) http://www-personal.si.umich.edu/~nsharma/dikw_origin.htm accessed December 2004

Strong, D.M., Lee, Y.W. and Wang, R.Y. (1997) Data Quality in Context, Communications of the ACM, 40:5 (May), 103-110

Stockdale, R., Norris, T., Kerr, K., (2007). Data Quality Information and Decision Making: A Healthcare Case Study. Australasian Conference on Information Systems. Research Gate.

Tayi, G. K., & Ballou, D. P. 1998. Examining data quality, *Communications of the ACM*, vol 41, pp. 54-57.

Tilly, R & Posegga, O. & Fischbach, K. & Schoder, D. (2015) What is Quality of Data and Information in Social Information Systems? Towards a Definition and Ontology. The Sixth International Conference on Information Systems, Forth Worth.

Valone, T. V & Templeton, J. J. (2002) Public Information for the assessment of quality: a widespread social phenomenon. The Royal Society.

Wand, Y. and Wang, R. (1996) Anchoring Data Quality Dimensions in Ontological Foundations, Communications of the ACM, 39:11 (November), 86-95

Wang, R. and Strong, D. (1996) Beyond Accuracy: What Data Quality Means to Data Consumers, Journal of Management Information Systems, 12:4, 5-34

Yin, R. K. (2009). Case study research: design and methods. essential guide to qualitative methods in organizational research. fourth.

LINKS

Accu-Chek Product descriptions: Retrived from the homepage of Accu-Chek on April 20th, 2019: https://www.accu-chek.dk

International Diabetes Federation. *IDF Diabetes Atlas, 8th edn*. Brussels, Belgium: International Diabetes Federation, 2017. Retrieved from: http://www.diabetesatlas.org

Roche Privacy Policy: Retrieved from the hompage of Roche Diabetes care on April 8th, 2019: https://www.roche.dk/da/OmRoche/Rocheidanmark/Diabetescare.html

Roche Digital Solutions: Retrieved from the webpage of Welfaretech on April 16, 2019: https://en.welfaretech.dk/members/roche-diabetes-care

Sturman, C. (2017, September). Roche & Accenture: Pioneering solution within diabetes management. Gigabitmagazine. Retrieved from: www.gigabitmagazine.com on March 19, 2019.

APPENDIXES

Appendix 1: The Interview Guide

Aspect	Tentative Questions
Role	Would you briefly describe your position and role at Roche?
Data Brainstorm	which system(s) are you working with? Do you collect any data? What kind of data is that? What kind of data do you use? How do you use the data? To what extent do you agree that you are using the data very well? Strongly agree – Agree - Not sure - Strongly Disagree - disagree What kind of improvement would you like to make? Are you doing anything to improve it? Do you allocate any resources for that? Do you face any challenges in doing that?
Intrinsic DQ: - Accuracy - Objectivity - Believability - Reputation	To what extent do you experience the data to be correct, error-free and accurate? Is the data interpreted or processed in any way before you receive it? How raw do you consider it to be? To what extent do you accept the data as being true and credible? In terms of its sources; how much do you trust the data?
Contextual DQ: - Value-added - Relevancy - Timeliness - Completeness - Appropriate amount of data	To what extent are the data beneficial and provide advantages from its use? To what extent are the data applicable and helpful for your tasks? To what extent is the age of the data appropriate for the task at hand?
Representational DQ: - Interpretability	

- Ease of understanding - Representational Consistenc. - Concise representation	To what extent is the data in an appropriate language and definitions clear? How easy is the data to understand? Do you face any challenges interpreting the data?
Accessibility DQ: - Accessibility - Access security	How easy is it for patients to access their own data? Do they know what kind of data is stored in our systems? How restricted and secure is the data to access outside?

Appendix 2: Transcribed Interviews & Coding Process

Open Codes	Axial Codes	Selective Codes
Casper Elkjaer **Q: Would you briefly describe your position and role at Roche DC?** **M:** My title is Customer Care Advisor and I am point of contact for Denmark in Mannheim. Then I have a team of six people I am managing in the Call Center in Mannheim where my role is to ensure that they solve the tasks, which Denmark is assigning us. **Q: Which system do you work with?** **C:** Salesforce. **Q: What kind of data do you collect?** **C:** We get patient data such as patient names, addresses, e-mails, numbers, age etc.… And that's about it. We have been faced with GDPR just like everyone else. Before there were different kinds of data that we actively collected, but this is nothing we keep any longer because of GDPR. Data that we also register in our system is the different product each patient is receiving or has received. **Q: What is this data used for?** **C:** At the moment we don't really use it for a lot. We create some campaigns from this data. We did it a lot before GDPR, but now we need patients to actively provide consent, which documents that we are allowed to use the data, or we are allowed to contact them. Beside more targeted campaigns we don't really use the data for anything. **Q: Why do you collect it then?** **C:** It's formalities. But also, we need to know to some extent who is using our products, so if something is wrong with one of our products we need to be able to recall it. We need to be able to contact the people having this product and inform them about the issue. This is done by a serial number that each product has. But to be honest, I don't think we have half of the real customers registered in our database.	Receives sensitive data GDPR rules resulted in less types of data collected and kept. = turtle Information about the products that customers are receiving is registered. = Patterns Data collected, limited use. Consent declaration for data usage has made it difficult to use the data collected. Data collected use for marketing campaigns. Data collection in CS → Formalities Most customers using the products not registered in the database.	Data is collected, but not used The GDPR rules makes it difficult for the company to use the data. Data collection in Customer Care is considered more as formalities rather than insight.

Q: What kind of improvements would you like to make to improve the data quality? **C:** Here in Mannheim we handle a lot of product complaints. I have some side projects going on where I am in charge of signing product complaints and the reasons into the system. All of this is done manually, and I cannot extract data only sign data. It would be valuable if one could extract data and use it to discover patterns related to the complaints we receive. **Q: Are you doing anything to improve the data quality? Do you allocate any resources for that?** **C:** No and yes. At the moment we are running campaigns on Facebook and Instagram, and people order a lot of devices. Within the order forms that the patients are filling out there is a lot of data they can either opt in or opt out. We do use resources to update the forms and the data that are put in or out of the systems by creating time span and reasons for how the data is collected and for how long time. This is done in order for us to be able to document for the authorities that the data we collect is up to speed relative to GDPR. So yes, it is prioritized. **Q: To what extent do you experience the data to be correct, error-free and accurate?** **C:** Well, the data itself is correct. The aspect that could be challenging is that when we run a campaign and offer free devices people can tend to choose the wrong devices. For instance, a type 2 diabetic who has just started monitoring chooses a product which is more suitable for a type 1. Here, the data will be correct, or we can correct it but the product type or suitability wont. This leads to the fact that we have data in our system about people having a product not suitable for them. Here the data will be correct, it is correct that a certain person has received a certain product, but after this person has received it and tested it and found out it's not suitable it	Handling product complaints by signing the claims and the reasons into a system. = Data collection done manually. No data extraction of product complaints – only signing. Recommendation → extraction of data using to discover patterns in complaints. Resources used to update forms and creating time span and reasons for how the data is collected. → GDPR documentation/reasons. = Not data quality reasons. Error-free & Correct data collected. Wrong devices are used. Problem is not data-related, but human related. Data in databases do not represent the real life.	Data complaints cannot be extracted. → Recommendation: extraction of data could be used to discover patterns in complaints. Resources used to update forms and creating time spans are GDPR related and not data quality related. The data collected is error-free & correct but does not always represent the real life.

will throw it away all while we still have registered the person to be user of the product in our system. Otherwise, I believe the data we collect is correct and error-free. **Q: Is the data interpreted or processed in any way before you receive it? How raw do you consider it to be?** **C:** The data is interpreted compared to the questions that are asked when collecting the data. But the questions are relatively harmless such as, are you an insulin user? Yes or no? How many times do you test? This kind of data we are not allowed to sign into the system. But it can help us knowing if they have ordered the right devices. So, in that sense the data is quite raw as it comes directly from the patients. Also, we receive data by two different ways; one is through the online systems where the patients fill out the form. The second is when they call us directly. And when they call us directly they just answer the questions that we ask them and if they don't understand a question they are able to ask for clarifications. By calling us directly there is less chance for misunderstanding and higher chances that the data is accurate. **Q: To what extent are the data beneficial and provide advantages from its use? How much value does the data provide?** **C:** Unfortunately, it doesn't take up much and provides only a little value. This is a fight that I have fought for almost three years trying to convince that we should become better at using the data. The data is useful in that respect that if patients are calling with a problem I can see all the interactions that has been with the patient. I can follow a track record with the patient and discover patterns. For instance, if one patient has called twenty times and complained about all the products it has received I can conclude that nothing might be wrong with the products but the patient. These kinds of patterns I can track in Salesforce and this kind of tracking of data is	Unrealistic information registered in databases Data comes directly from patients. Person sensitive data not allowed in the system Two data streams: - Online system form - Call Data accuracy depends on the medium used to collect RDC uses limited efforts on structuring and making use of the data collected More usage potential: Following a track record with patients and discovering patterns	Since the data comes directly from patients, it is not biased. The bias happens when the data is entered in the system. Data is received either by call or survey. Data accuracy is higher when it is by call, meaning data accuracy depends on the medium used to collect it Recommendation: Following a trach record with patients and discovering patterns.

very useful. But in relation to big data we don't really make use of it. Overall, I don't think the data quality is high. There is a huge potential of using data better and more efficiently especially because we collect so much data, but I wouldn't say we use it. At the moment our biggest struggle is the new GDPR rules. There are certain vague rules about what kind of data we are able to sign into the systems and in order to be on the safe site we don't really touch or do anything with the data. At the moment we are waiting for global legal to provide us with some guidelines for data usage. At the moment I don't even think we can register any gender for our patients. It is kind of silly, as you can find out the gender by looking at the name, but this strict it has become with the new GDPR rules. If we could register the treatments that the patients are doing it would be easier to target our campaigns even more and touch more patients. **Q: How long do you have the data stored in the systems? Any rules?** **C:** I have worked here for three years now. Once in a while we have had what is called "data washes" where we just delete a lot of patient data that we meant were no longer applicable or useful. In this process we consider all the patients and those with limited interactivity will get their data deleted. We strive to keep the databases down. People who have not had a lot of interactivity for a year get their data deleted from our system. A couple of years ago we had a shift from Dynamics into Salesforce where we completed one of the biggest data washes and deleted almost half of the data. The initiative of deleting almost half of the data came from the consideration of 'less is more', so we focused more on quality rather than quantities of data. We would rather have less data from active users rather than a lot of data from nonactive users. We want real people who say something.	No use of big data – low data quality GDPR is the main reason for the little interest in data usage Employees not clear about the data usage rules. Potential result: targeted marketing campaigns. Data washes happens once in a while to protect data overload. = Less is more Focus on quality and not quantity.	Employees are little clear about the data usage rules, as global has not provided clear structures. Timeliness & Accuracy are considered by handling data washes, which improves data quality.

Q: How easy is the data to understand? Do you face any challenges interpreting the data? **C:** The data we receive from patients is easy to interpret and understand. However, the data we get from hospitals and doctors can be more difficult as they are using other terminology than we do here in Customer Care. **Q: Do you believe that there is a potential that the data can be used in a different way than it is done now? And how?** **C:** Within what is legally acceptable in terms of GDPR I am not sure how much it can be used or how big a potential the data can have. But if we don't consider the GDPR rules there is a huge potential for us to utilize the data more efficiently. Mostly because we really don't make use of it. We could start more campaigns and doing more direct marketing, which would be a new path for us. Using the patient data to target our campaigns would be a great deal. For instance, by looking at the age differentiation and discovering some patterns we can make more targeted campaigns and shoot accurately. At the moment we more or less just hope to hit the right ones with our campaigns, by using the data efficiently we can become more targeted. There is a huge financial gain for Roche DC if we can become better to targeting our campaigns to affect the right ones.	Quality data are the active users. Ease of understanding depends on the data sources GDPR creates data usage confusion Big potential for data, however not realistic Data usage potentials include: - Direct marketing opportunities - Targeted marketing	There is a big potential for data, however GDPR makes it difficult to realize
Open Coding	**Axial Coding**	**Selective Coding**
Q: Would you briefly describe your position and role in Roche DC? **M:** My title is Business Supporter which is covers widely. I am servicing our salespeople if they have troubles in the market. I am of service for Customer Service in Mannheim if they have questions related data in our SAP		

systems. I am also working to launch our new web shop with our product manager. I transfer data from Salesforce into SAP on a daily basis. So, all the data that customer service register in Salesforce I transfer to SAP via a macro. **Q: Which system do you work with?** **M:** SAP. **Q: Do you collect data? What kind of data?** **M:** I don't collect data from SAP, I put data used in connection with our outcome-based solutions into SAP. But this data is not data that you can pull out and use as this is all patient data used to make deal masters and billing plans. **Q: What kind of data do you use?** **M:** The data we get in SAP is mostly patient data, which I cannot use. And I also work with our solution-based sale where we make a deal master and a billing plan for each patient, and here we have collected data from all of the patient who are registered in our solutions-based program. This is all kind of data about the patients' healthcare and their progress in the program. This data also comes into SAP, however, neither used for other purposes than formalities. As mentioned before, with our solution-based health program we make a deal master and a billing plan for each patient. They are all registered with a number as we are not allowed to have visible patient data in the system. All patient data is in our Sharepoint system in a closed file which only our Market Access Director and I have access to. Every time we are going to send a bill to the municipality - as they are the ones who are paying depending on how much progress the patients have made and how much long they have "stayed in the zone" – the invoice has to include their CPR numbers as well as their case identification number. This is the data I have on a closed file in Sharepoint.	Data about outcome-based solutions used to create deal master and billing plans. (patient data) Patient data for billing plans and deal masters include data about the person's healthcare and the progress made in the program. Most of the data in SAP not used for other purposes than formalities Patient data in SAP is anonymized by a number Municipalities are the one paying for the health program depending on how much progress the patients have made. = All data is considered as proof for municipalities.	Data mostly not used for other purposes than formalities (SAP) Data from the healthcare program is used to measure the progress of patients who participate.

Q: To what extent do you experience the data to be correct, error-free and accurate? **M:** At the moment we only have data on the patient level including a CPR number and a case identification number. This data is used by agreement with the municipality and the patient, so I believe all this data is correct and accurate as the sources are the patient himself and the doctors. Also, we will increasingly be using this data as we will be working more on our solution-based healthcare with municipalities and companies and the free choice concept, where patients have the right to change their appropriation from the municipalities and to us. By this way we will be using all of the patient data more. So, my answer to your question is that the data that is put into SAP is all correct, but what could go wrong is outside factors like the person who is entering the data. So basically, the data is correct if the person who enters the data has entered it correctly. In this sense accuracy is defined by the people who interpret or process the data first.	Since data sources are patients and doctors it is highly trusted. Human mistakes can affect data accuracy. → External factors.	The data itself is of high accuracy, but human mistakes can affect it. → External factors can play into the accuracy.
Q: If you could change or improve anything related to the data you receive. What should it be? **M:** I don't think we can improve the data just like that. But we could focus on our e-commerce and construct it in a way so patients who have this free choice could order their own blood glucose monitor and strips through the e-commerce site, so they don't have to go through customer service or me. And they could have access to all of the data about themselves collected by the site. This way they could kind of manage their own products and orders as well as decide what data should be visible for us and what data shouldn't. This is something we are working towards. So basically, that patients have their appropriation numbers registered in this webshop and they are able to order and act independent from the municipality.	New e-commerce site intends to make it easier for patients to get access to all data and act independently from customer service when ordering. Future vision: Self-management of data	Representational DQ pattern

Q: Are you doing anything to improve the data quality? Do you allocate resources for that? **M:** Not in my field and area. As mentioned before, I can't really interpret or actively do much about the data that I receive because it's all patient data. You can always pull out data from the systems. I do that all the time, but this is related to sale. This can be done on a customer level. This data could be improved or used for marketing purposes. I am also sure that they use it in the marketing department for direct marketing. If we want to know what a specific customer in a specific region has bought, we pull out this data from SAP.	Data in the systems can be used for marketing purposes. = Direct marketing	
Q: To what extent is the data interpreted or processed before you receive it? **M:** It's interpreted or processed in a sense that whatever I put into our SAP system is data that I receive from someone else. This is for instance an order that I receive from a hospital or a company. And they receive it from the patients. So yes, if the hospital provides me with 100 % correct data it is registered in our system with 100 % accuracy. But if the data is faced with a human error, for instance the company or hospital has made a mistake when entering the data, it will affect the accuracy in our system also. So, it's more external factors that plays into the accuracy rather than internal. So, we may assume that the data in our SAP system is interpreted or processed in a high extent.	The data is processed before received = Doctors & hospitals The data is as accurate as the hospitals register it to be External factors play into the accuracy rather than internal. Data in SAP highly interpreted and affected	
Q: Could you tell me more about the SAP system in general? **M:** The SAP system has many functionalities. It includes a purchase system, a sale system, a finance system, material master and a production system. The modules I am using I SAP are the sales, purchase, finance and material master modules. All the products we sell here in Diabetes Care are created in the material master, which are master data in the SAP system that includes		

product name, order number, product date etc. All of this data is structured and maintained in a certain way. You can keep oceans of data in the material master all product related. When we then sell the product in the sales module the order number helps the system to receive all the product information. The order number is unique for one product so when we make an order the system receive the product information or relevant data from all of the modules. So, product number and patient numbers are unique in SAP. All our patients have a unique number. **Q: Is there something you miss in SAP? Anything you would like to improve in the system?** **M:** This is something we can consider on a long-term. Just to go back to our solution-based healthcare program. As more municipalities we convince as bigger data amount we will have and as more difficult it would be to maintain or register this data. In terms of this it would be nice if we could get global to help us with creating better and more efficient systems. Right now, it is all done manually. This takes time and resources as we expand and include more municipalities in our program as I am the only one working with SAP at the moment. **Q: Can you tell me more about the solution-based healthcare program?** **M:** The purpose with the program is to improve the patient's everyday life through solutions and not only products. We don't aim to sell insulin pumps and glucose monitor but trying to improve the life of our patients by solutions like offering them consulting, a dietitian or a coach. So, each municipality can sign up patients for the program lasting a certain period and within the period we need to create results to show the municipality, which we will then receive a bonus for. In this process it is very important that we collect patient data	Each product has a unique order number, which presents the product information Patients & products have unique numbers = Representational Consc. Progress in healthcare program means bigger amount of data Bigger data requires better and more efficient systems Result-based bonus is offered to the company Patient data is collected throughout the whole process for assessment of the progress	The unique number that presents each product acts as a form for data security. Challenge 1: The system is not suitable for big amounts of data → Bigger data requires better and efficient systems

throughout the whole process. Only this way we can assess the effect of the treatment. **Q: How long do you have the patient data in the systems?** **M:** The challenge with SAP is that we are not able to delete the data ourselves. This is a challenge because it contrary the new GDPR rules. There is no solution on this yet, but it is a topic with a lot of attention.	SAP functions contradict GDPR rules	Challenge 2: Once you enter data into SAP you cannot delete it → GDPR issue
Q: How easy is it for you to interpret or understand the data structured in SAP? **M:** It is easy to understand and interpret. I don't really face any challenges in regard to this. The language is English in SAP, however, the data I enter is all in Danish.	Interpretability = Easy	
Q: How restricted and secure is the data to access? **M:** Everyone with access to SAP can find this data. But first you need to know that they data is collected here. Second you need to know what the symbols or numbers are representing. For instance, what does the number 0-7877-23 mean? In order to be able to use this data you need to know the hidden message or real meaning of each number, which many in the organization does not. There are no names entered in SAP every person has a person number which the person is identified by. This is how we use data. All the patient data we receive are sent to a secured mailbox.	In order to find the data, you need to know about the hidden messages = Representational Data representation issues is seen as accessibility issues Patient data sent to a secured mailbox = Accessibility	
Q: Do you believe there is a potential for Roche DC to use the data in a different way? **M:** In order to use any data, we need patient consent. We use the data as we are allowed to and that's pretty much it. I think there is a potential to use data in a different way, but you need to remember this is a healthcare company with very strict rules and regulations regarding data usage. At the moment we only use data as we know we are allowed to. For further use we can end up getting outside the compliance zone. We	Patient consent needed for any use of data Strict compliance rules	

need to tell patients the exact thing we are going to use the data to and they need to provide consent before we can use it. Our customer service in Mannheim puts information into our Salesforce system. This information is all patient information. Everyone calling in and addressing any concerns or issues with the products will become registered in Salesforce. And the information that can be created and maintained in Salesforce could go from being information about the product the patient requires to any kind of issues with a certain product realized by the patient. All of this data is in Salesforce. Once a day there will be created a report by Salesforce which is uploaded to SAP through a macro.		
Open	**Axial**	**Selective**
Q: Would you briefly describe your position and role at Roche Diabetes Care? **L:** I am Customer Support Coordinator at ICC, which is our Customer Service Department in Mannheim. I am in charge of all the administrative tasks related to patient data such as creating new customers in the systems, ensuring standing orders, that all the information is correctly, and customers receive their products at the right time. When I say creating new customers in the databases I mean entering data like, name, address, e-mails, birthday, whether they are patient, kin or staff. We have a field called BGM, which include type one and two people with diabetes. Then we have IDS, which involve the people who are using our insulin pumps. Then we have some fields where they can inform whether they want to receive newsletters. It's important that we are able to document this. When we receive an inquiry from a user, they have already ticked off whether they allow us to contact them or not. This e-mail I attach to their profile, so we are always able to go back and see their consent. Previously, we also noted down whether they had type one or two diabetes, how often they	Patient data in the systems include: name, address, e-mail, birthday, role, etc. Distinguish between BGM & IDS Documentation is highly prioritized → GDPR	Most of the data is collected, structured and maintained for documentation purposes. → GDPR rules

measured their blood sugar, and which therapy they were in. But today [after the GDPR rules] we are not sure that we are allowed to log this information anywhere. This is very sad, as I believe this information could be valuable for segmentation and more targeted marketing. At the moment we do this by considering the products that are registered for the customer and send it out, but the challenge is that you cannot know whether the customer has type one or two diabetes. **Q: Which system do you work with?** **L:** Salesforce, only. **Q: And you only collect patient data?** **L:** We have information about our business customers such as municipalities, regions and companies in the systems as well. Typically, a contact person is connected to each business partner, which include the person's name, e-mail, phone number and company name. So, when a patient is created in the system with all necessary data about them, we then link a contact person and enter their role such as they are politicians, buyers etc. We have an app where they can give us their consent on whether they want to receive any kind of materials. **Q: So, what is the data used for in general?** **L:** Mostly marketing campaigns, but for our professional customers such as municipalities, regions and HCP's we also use the data to create an overview of who we have contacted and what our deal is with them. So, the data is also used for market segmentation. **Q: To what extent do you believe that you are using the data very well?** **L:** I think we use the data correctly, but I think we could use it more than we do. For the professional customers such as municipalities and regions there have not been done any segmentation or campaigns. More have been done for the patients. This is something we are focusing more on now and trying to do more campaigns and	Person-sensitive information stricter to use. → GDPR Person-sensitive data could be valuable for segmentation and target marketing purposes. Consent collection through an app. Data is used to create an overview of the people involved in the supply chain Next step: Doing segmentation and	Person-sensitive data could add value in terms of segmentation and target marketing purposes, however the company needs to find ways around GDPR rules. Recommendation: Segmentation and campaign targeting municipalities

segmentation targeting municipalities, regions, hospitals and payers. With our new strategy we try to focus more on the professionals and will be using data about these more than we have so far. **Q: Are there something you miss from the data?** **L:** I don't miss any kind of data, because we do get the data we need. What we need is some compliance guidelines of the data. We need more information on how the new GDPR rules are allowing us to use the data. Salesforce provides with so many opportunities and flexibility, but we get limited as we don't know how much and for what purposes we are allowed to use the data. **Q: To what extent do you experience the data to be correct, error-free and accurate?** **L:** That's a good question. The data is as correct as it's entered. Sometimes the data can be incomplete if the customer does not answer all the question in a formula, for instance, miss a birthdate or phone number. So, the data is correct if the people entering the data enters it correctly. I mean it is always a human error that is the cause for the data not to be correct. **Q: Is the data interpreted or processed before you receive it?** **L:** No, it's not. Because it comes directly from the patients. The patients are the sources to the data that we collect. **Q: To what extent do you accept the data as being true and credible?** **L:** Well, I assume that they are trustworthy… we have to assume that. But as mentioned before we cannot control the people who enter the data. Patients who enter the information can miss a birthdate or a correct e-mail. Before we operated through Salesforce we worked with Dynamics and Dynamics was connected to the post office. When you entered two or three of the first letters of the street name it provided you with suggestions, hereafter you could enter a road	campaigns targeted municipalities. Data usage guidelines not clear for data consumers → GDPR Salesforce = many opportunities and high flexibility Data incompleteness Data accuracy depends on people Integration from Dynamics into Salesforce required a change of mindset	There are contradicting views on the flexibility of the system In order for data to create value for the company it is important that it is accurate and timely. The accuracy depends on people

number. This way you could not enter a wrong or non-existing number. Now it is all free text, so the patients can more easily provide us with wrong data and information. Further, the system cannot delete any data. It can go from being active to non-active, but we cannot delete any data. This can be a challenge as the data might not always be timely and up to date. **Q: To what extent are the data beneficial and provide advantages from its use?** **L:** The data is essential for my tasks. If a customer is calling and want to order something it is super helpful and relevant that we have all of the data and can go through it to figure out, we have all the right information needed. In terms of our patients using insulin it is also very important that we have the correct address and hospital as otherwise the patient won't get the right devices and we will use time and resources sending invoices to the wrong people. **Q: To what extent is the age of the data appropriate for the task at hand?** **L:** Well, the data will never become too old, but we can go in and create reports to find out when a customer has last been seen active and if it's five years ago we assume the customer is no longer active. A couple of years ago we have had an 'Accu-Chek Club' where everyone could sign in for free and receive a products and devices and just becoming updated on what was happening within Accu-Chek. But many has moved on or got new devices, so after five years we assume they are inactive if they haven't ordered anything. However, if we hear from them again we are able to go in and make them active. As mentioned before we are not able to delete any data, but we can change them to be active user again. **Q: To what extent is the data easy to understand? Do you face any challenges interpreting the data?**	The system features can encourage inaccurate data In order for the data to be valuable for the company it is important that it is accurate and timely. No data can be deleted (active or inactive) → Data overload within the system → Contradicts data washes	Since data cannot be deleted in the system, but only change status it can encourage data overload = contradicts the intention of data washes & the less is more principle

L: I find it very easy to understand and interpret. It's all patient data and comes from the patients, so the format and language are straightforward. Salesforce is a very easy system, whether you pull out data or register new data it is very user friendly. **Q: How easy is it for you to access the data?** **L:** For me it's easy because I am Business Administrator, so I can easily get access to whatever data needed. But when one without business admin rights needs access they have to ask for permission and go through some courses and trainings, then define which role they have; for instance, whether they are in marketing, sales or service or business administrator. And if you have a profile for instance in marketing you can get access to Salesforce, but still not access to the patient data. So, the patient data is quite secured. **Q: Do you think that the data has more potential for use and how can Roche DC use the data in another way more focused patient care?** **L:** The data that we collect has a huge potential for use. But once again we are limited by the GDPR rules. We still don't know what we are allowed to and what we are not allowed to. Once we know these guidelines we can become more targeted in so many aspects. In 2012 we migrated from Prisma into Dynamics. Prisma was a very old system, so it was great with a new and more user-friendly system operating through windows and with plenty of features and details. And in 2017 we went from Dynamics into Salesforce. There were many challenges related to the shift. England was the first country completing the integration and they complained a lot. They considered it to be too slow and not very user-friendly. The processes were very complex and slow and the setup in Salesforce was too complex and difficult compared to Dynamics. When we changed to Salesforce here in Denmark there were a mixed attitude towards it. The	Easy interpretable Data security: high No clear GDPR rules provided from global. Results in people being confused about the data usage	GDPR itself is not a threat on the data usage, but the lack of guidelines provided keeps employee from using data much

Customer Service Department were not happy about the system as it was time-demanding, but the sales people found the new system to be great as it provided them with so many features. For instance, we can make our own reports, dashboards and sales charts, relevant for the work they are doing every day. Overall, employees who are dealing with orders find it more complex and time-demanding than the rest of us. The system is less user-friendly and more time-demanding for people in the customer service, which is problematic as everything in the customer service department is measured by minutes and seconds, meaning that each call will be measured in minutes and seconds. So overall, it's not an order-system, but a sales-system.	Salesforce = too slow & not user-friendly. System flexibility = Low	Salesforce provides advantages for some, while being more complex and time consuming for others
Open	**Axial**	**Selective**
Q: Would you briefly describe your position and role at Roche Diabetes Care? **A:** I am a senior strategic project manager, working with marketing, and have been in Roche Diabetes Care for 8 years. I have a temporary role as product manager for our BGM product, which include all of our Blood Glucose Monitor devices. As project manager I am in charge of different marketing projects such as product launches. **Q: What kind of systems do you work with?** **A:** In my role as Product Manager I mostly work with Salesforce Marketing Cloud, which is the system that we use to send out newsletters. I also have access to Salesforce CRM system, but haven't really been introduced further for it, so it's not used. When I am sending out newsletters, I get Linda, who's Business Admin, to pull out data, such as consent declarations of people who has approved that we can contact them. I also use Excel for data storing. It provides a good overview of all the patients, municipalities and companies, which have	Marketing people only have access to Salesforce Marketing Cloud = Accessibility Excel is used to achieve overview of consent and customers	Data accessibility depends on the role of data consumer

approved on getting contacted. Last, I also use outlook e-mail. **Q: Do you have access to all kind of data in Salesforce?** **A:** Yes mostly. I don't have direct access to the system, but I can always ask our business admin to extract reports and data for me. **Q: What kind of data do you collect and what kind of data do you use?** **A:** I collect data like, names, addresses, e-mails, phone numbers etc. We also collect sensitive data such as whether the person has diabetes type one or two and which treatment they get. In relation to newsletters, I mostly use data like e-mail addresses. All the reports that our Business Admin creates and uploads in the Salesforce Marketing Cloud are anonymized, so I cannot see people's data. I can only see that a report is made from all the criteria I have provided our Business Admin with; for instance, all the people should live in a certain municipality and use this certain product. Then I 'use' the e-mail addresses, without really being able to see the email addresses to send out the newsletters. In relation to our Healthcare Program I have all general information about the patients, so I can contact and inform them when we are having meetings or starting off new sessions. This data is used for coordinating purposes, but it is also important that we, depending on which type of diabetes they have, can customize their treatment and provide them with the right service. **Q: You just distinguished between personal data and sensitive data; can you provide with examples on both types?** **A:** Yes. Personal data includes, name, address, phone number and e-mail and is treated one way. While, sensitive data are data representing their disease, sexuality, religion, political views etc. This data is handled in a different way. **Q: To what extent do you believe that you use the data correctly?**	Person data & Sensitive data collected with different intentions and purposes. Anonymization of data in the Salesforce Marketing Cloud highlights accessibility discussion Personal data used for coordination purposes. Sensitive data used for customization of treatmens Difference between personal & sensitive data. Handled in different ways with different levels of strictness.	Sensitive data can provide the most value to the company if used correctly in line with GDPR

A: I believe we can always become better at utilizing the data more efficiently. But with the new GDPR rules we are limited and also very careful on how we use the data. However, all the people we have data about in the Healthcare Program have provided us with consent that we can use their data for marketing and other purposes. The part with the declaration of consent we use a lot of the time working with. Whenever we are having an activity we are using plenty of time making sure that we have the right to use the data. **Q: If you could improve anything with the data you collect, what should that be?** **A:** I don't think I would improve anything, because I think the data we receive is of a high value, but I would like to be able to send out more newsletters. I miss a way to collect more consents. We cannot use the data if we don't have consent. ´ When this whole GDPR process started a year ago, we had a huge cleaning in our databases, so we went from having around 20.000 people that we could contact to only having a couple of thousands today. By losing so many consent agreements we have lost many customers as well. And at the moment the biggest challenge is figuring a way to collect consent agreements documenting that we are allowed to reach out to patients. I cannot think of data that we miss, I feel like we do have the data we need, however, not always the right consents. **Q: To what extent do you experience the data to be correct, error-free and accurate?** **A:** The data is correct, until we are proved wrong. The error or inaccuracy occurs when people change e-mail address or phone numbers without informing us. We don't send out any physical material any longer, so it's all digital information that can be inaccurate regarding our use. Whenever we send out something there are some e-mail addresses bouncing. To be honest, I don't really do anything about this even though I	Lack of GDPR knowledge makes employees become less open to use the data Consent declaration is the biggest obstacle for data use Recommendation: How can we find a way to collect more consent declarations? Human touch plays into the data accuracy. When materials are sent out there will be emails bouncing. (Static/Dynamic issue on data and systems)	GDPR rules are the main discussion of the data usage. People are most concerned about HOW to use the data One of the main questions after the GDPR implementation that needs to be considered is HOW can we find a way to collect consent declarations for data usage? Human touch affects the data quality The world is dynamic and the data in the databases are static,

should. The system can create a list of all the emails that are bounced. It would be a good idea to provide our customer service with the emails that are no longer active, so they could update them, but this requires resources that we don't have or is just not prioritized. **Q: Is the data interpreted or processed before you receive it?** **A:** It should be raw. It's only about email addresses, so I assume it's raw unless the person who is entering the data has made a mistake. **Q: To what extent do you consider the data to be true and credible?** **A:** I do trust the data to be credible and true. I have to trust it that way. But we first know it once we have sent the newsletters and received a bounce. As mentioned before the data is as trustworthy as the thoroughness of the people entering the data. The data itself is credible, but what can affect the extent of its truthfulness is the human touch. **Q: To what extent are the data beneficial and provide advantages from its use?** **A:** If we consider the data to be fully correct, I would say that it provides us with a high value as I can send out a message to our users and make them act, such as ordering a product or subscribing newsletters. Then I can see in the system how many has clicked on the campaign, signed up for the newsletters, or made downloads. This is one of the most efficient ways to reach out to our users. So, the data is of huge value to our company, and this is the reason that it's important that we have as many consents as possible. Also, in terms of sending out surveys the data is very beneficial. **Q: To what extent is the age of the data appropriate for the task at hand? How long do you consider the data to be relevant?** **A:** Hmmm, I am not sure about this. I mean… the data is relevant as long as it is used. As mentioned, we don't use resources to manage	Recommendation: Putting in more resources on keeping the system up to date Human factors on accuracy Consent declarations has a big effect on the value of the data Data maintenance is not a big prioritization.	meaning this can affect the data accuracy. -Recommendation is to put in more resources on keeping the system up to date.

or keep the data up to date. It's not a prioritization. I don't know whether global use resources to manage or keep the data updated, but we don't here in Roche Diabetes Care, Denmark. What we could do is that whenever a customer was calling, we could go in and look whether all of the information about this particular customer was right. This would be the easiest way to keep the data updated. Alternatively, they should call everyone, but there are no resources for that. **Q: Do you think that the data has more potential for use and how can Roche DC use the data in another way more focused patient care?** **A:** There is a huge potential. But it's all about the compliance part; What are we allowed to and what are we not allowed concerning the data. I think the challenge is also that we have so few consents declarations. If we had consent on every patient in our database, we would be able to use the data much more efficiently. We would become much more targeted in our marketing. We could customize every campaign and become much more targeted in our services. I don't know if we are allowed to, but this could be a potential.	Recommendation for keeping data updated is checking the activity when patients are calling. Consent declaration for data usage	By considering the activity when patients are calling RDC puts in resources on the maintaining the data accuracy, so less bouncing's are faced.
Open	**Axial**	**Selective**
Q: Would you briefly describe your position and role in Roche DC? **J:** I work in a global team and by work stream I cover different processes. My role is to cover all the service functionalities. So, each country has a certain process with their customers. They would like to document in the CRM solution that all of the interactions are tracked and can be reported on. In the team we represent various modules that support these business processes. I am responsible for the service cloud, which is one of the modules in Salesforce. In the	Documentation for CRM solutions of high importance	

service cloud we have different forms for data entry where new customers and existing customers can be identified and saved and each of the interaction is logged as a service case with the summary of the conversation and the outcome. If it's a simple enquiry we can close the case quickly. If the customer's reporting an issue with one of our insulin pumps or the meters and the complaint cannot be solved by local I take care of the process of how to forward or escalate it to a global function. I have been working for Roche for 20 years in several functions such as in the service function, call-center and customer service. **Q: And you basically only work with the Salesforce System?** **J:** Yes, we decided to work with Salesforce because it is the biggest cloud-based Customer Relationship Management Solution System and was able to cover all of our requirements in sales and service. Further, Salesforce has many partners that they work with and a huge team of developers that continuously improves their solutions. We also looked at the costs of course. Last, Salesforce was also interesting because of their license model. Everyone around the world in Diabetes Care need a license to log into Salesforce, which we believe is important in terms of the privacy and capability. **Q: How do you work with data in your role? Or how does global work with data in general?** **J:** When we started rolling out the CRM solution we interviewed all countries asking them what kind of data about the customers they wanted to choose to store in the database. It starts with title, gender, address details, and various paramedic about classifying and segmentation. So, we collected a lot of fields from the countries and then we consolidated and harmonized them, so we had one harmonized page in which all	The service cloud provides an overview of customer tracking = Data representation. Salesforces covers all the requirements in both sales and service Salesforce license is needed to log in = Accessibility	Data representation accuracy depends on the systems' ability Accessibility depends on the system structure

of the customers are entered, whether they come from brazil, Malaysia or Germany all of the customers are saved with their paramedics in a global harmonized way. So, we are able to segment not only the end-users, our patients, but also the healthcare professionals in a similar way. There are some fields that are mandatory of course, like which country you live in and when the mandatory information the countries are able to create a new customer record in the solution. We then also have various ways of how the countries get their data and how the countries get their permission for their customers to save the data. When they call in or write in, the customers grant permission by either providing a signature or a consent through a voice recording. When customer service has obtained the permission and it's saved, they will create a new record in the system and then we are able to send their data virtually to the cloud so when our user in Salesforce clicks on save, the new record goes to a data center which is then operated by Salesforce. As you can imagine we have millions of records in Salesforce and when a country wants a new field for a new opt-in, if they have a special new purpose, like a marketing newsletter, they contact us and say: 'hey we have a new email and newsletter and would like to capture an opt-in for the customer and we make sure that this new field is available in the record. **Q: Do you face any challenges with the data that you work with?** **J:** If we consider the customer service department, it is important that when a customer calls in, that we can identify and find the customer in the database as well as all the interaction made with the customer. So, for them it is important that the system has a fast and performant search engine and we are also working on connecting our Salesforce with our future phone system, so that when a customer phones in we can identify the number and find all of the	Salesforce roll out included all the countries to come to a shared agreement on the data aspects stored. Rules and guidelines provided for countries for how to create new customer records, which makes it possible for global to keep an overview All interaction with customers is tracked	Even though global suggest data guidelines and structures are provided, local are not aware Global data visions include targeted marketing

information about the customer directly when we receive a call. **Q: What global visions does Roche DC have for the data? How does Roche global wish the data to be handled in general?** **J:** Our research development colleagues; people who think about new ways of helping patients, caretakers and healthcare professionals in managing diabetes will of course like to benefit from the insight we have in our CRM solution so with a global solution where every interaction is documented in the sales structured way global functions can get insight about the topic reasons why customer is calling us and suggestions for improvement. Here we see a big value in having a global database of providing insight across borders of what are trending and what are the hot topics in the market and what patients and healthcare professionals talking with us about, so we can recognize these patterns and turn them into actions of matters in for example product development.	The advantages of a global database: People across borders can use the insight collected in CS department. Value of global database is knowledge sharing across all landscapes.	Global desire to increase knowledge sharing across all boarders by making data visible The high value of a global database
Q: Can you describe the role of GDPR and how it has affected the data usage in Roche? **J:** Obviously, we store a lot of sensitive data, as we know a lot about our patients and their diabetes, so we already have a solution that protect them and their data. It's an encryption solution, so when I enter information whether the patient have a type one or two diabetes, this information is encrypted and nobody outside diabetes care is allowed to see this data. When it's stored in the datacenter it's safe and cannot be accessed in the cloud. So even before the GDPR came to effect we had this encryption solution in place that protect the data. What I implemented newly in regard to the GDPR across all of the countries, a template for if the customer wants to know what data we store. So now we have a template in which we can fast and easily extract all of the data we have stored about	Not much change after GDPR implementation: Encryption solution tool also existed before. The encryption solution tool for data protection	Encryption solution implementation in DK Encryption solution implementation in DK

the citizens. This change I implemented in service and not really needed in marketing, but of course I am only talking about the changes in Salesforce. If you are asking the marketing or product collogues they will tell you about the changes needed in different SOP's after the new GDPR rules came to affect. However, in the IT landscape we were not affected as much as we already operated through a good data encryption solution tool. **Q: Can you describe some of the guidelines for data usage provided by global? Does the global team use resources to ensure that local countries comply with the guidelines?** **J:** I am not sure how the global leader team works in that respect. What I know is that they of course provide global policies and guidelines and work with local legal counsel and ensure that the global guidelines are broken down into local guidelines. I know that the GDPR tasks for these are to make sure to store the global resources in a central repository and then of course on top of collaborating and communicating between the global functions and the local functions. There is also an instrument about the corporate audit; for instance, there was recently a social media presence in the countries. This is also interesting, like, how we interact with customers in our Facebook & Twitter pages. So that's a very formal process where the auditer visit the countries and ask questions and fill out an audit report and hopefully they have no critical findings. **Q: Do you use resources on keeping the data up to date?** **J:** The system admin can support the customer service team on duplicates for instance. When a service team find that the same customer is available in the system five times, they can create reports and send the ID's (the numbers of the record) to the system admin. Then can then merge the customer history and the record, so instead of having	Data storage solutions Global policy guidelines are communicated to local compliance directors, which makes sure to communicate it internal the organization. Resources used for issues like duplication and data update	The communication challenge happens in local entities

five duplicate the system admin can make sure that there's only one record. This can also have within the country. There is a data loader capability; how you can extract data from solution and the data is stored on a service and is protected of course and only the CRM and marketing manager has access to it. Then they can apply some filters and clean the data. Either the countries do it themselves by using this download and upload function, or they can ask the system admin to help them extract and clean the data. **Q: What data challenges were faced when you integrated from Dynamics into Salesforce?** **J:** The countries were using Dynamics for eight to ten years, so they had already used it for quite a while and already had many records in Dynamics. So, we then asked questions to the countries like 'how interesting are you in keeping the old customers?' so if a customer was not active and you did not hear about the customer for two years; does the record stay in the archive in Dynamics? Or do you also want to transfer the inactive customers? The countries then decided how much time they wanted to go back in history and maybe only take the customers with them that they have had an interaction with within the past twenty-four months. As you can imagine, once these rules got applied, only a subset of data was actually extracted from Dynamics. Let's say all the ones from the past twenty-four years where a case was created, a product was registered, and so on. And then we had mapping tables that all of the orders were extracted in an order table, all of the cases, all of the product. And these tables were uploaded by a migration team into Salesforce. This was a big resource effort from the country perspective and the IT to ensure that only clean data is uploaded into Salesforce, that you have a fresh start and you put quality data (a complete set of data that is accurate). That	Data loader capability for data extraction and storages. Change from Dynamics into Salesforce resulted in data update process. Patient records were only considered active if there was interaction within the past two years. Focus on quality and not quantity.	Tools and procedures provided for data storage, but not used in DK New systems have resulted in new ways of collecting and storing data

you don't start with a blank data set and the call center does not have to recreate every customer when they call from scratch. **Q: Do you think that there is more potential for the data collected and how?** **J:** Yes. The hot topic at the moment is Artificial Intelligence. Where we in Diabetes Care do not want to invest in human resources sift data we can look into applying machine learning to help us look into dark data. For instance, when you have not visited certain account in your territory or you have not heard anything from customers for longer than six months this artificial intelligence can sift to this data and no manual intervention is needed and the AI can propose action for a set of data for instance to reach out to patient to a simple call or send an email to customer asking whether they still using our products. So, in the future we can hopefully use AI to turn the dark data into data that we can process for our campaigns.	Less is more principle Artificial intelligence for dark data. Focus on marketing campaings.	Less data more value is considered AI can turn dark data to value
Open	**Axial**	**Selective**
Q: Would you start by briefly describing your position and role in Roche DC? **K:** I am a Market Access Advisor and work within supply and sale of Roche DC's solutions. **Q: What kind of systems do you work with?** **K:** I work with Salesforce, which is a Customer Relationship Management System. It is used to keep track of our connections with the municipalities and companies, both in terms of the	Data for relationship management	

appointments made with them, but also keeping track of all the information we have with the partners. **Q: What kind of data do you work with or handle in your position?** **K:** It's what I call prosa data. So, it's not hard data, but more data about what indication we have received from customers in relation to our offerings. All the data comes from phones, interactions, emails and meetings with the municipalities. Here in market access we only sit with data about municipalities and not patients. **Q: What are the data used for?** **K:** The data is used to qualify our services and offerings to the municipalities. But, also to qualify our offerings to the tendering process, where we will be given information to the wholesaler so that they can win the tendering process. Basically, we use the data to create a strategy for completing a sale. **Q: To what extent do you think you use the data correctly?** **K:** I think we use the data correctly in order to always be in process and come further. No doubt that we could become more efficient and more streamline on how to collect information and how to disseminate, or knowledge share that information. But I think there is a good structure on what is market access in terms of the supply part, which only concerns wholesalers, that part is a closed system. And then we have sale and promotion of our healthcare program, in which only market access and the sales team shares that information. **Q: What kind of improvement would you like to make to the data? Is there something you miss from the data?** K: I would like the CRM system to be more flexible and user-friendly. Both in terms of how to extract data, but also in terms of how fast it works daily. We say that the system is used both by me and my colleagues on a daily basis, therefore, it should include a high usability and speed, otherwise one is not that	Data for tracking patterns Soft data received used for indication about customers Data about municipalities and business partners processed Data used to create market access and sales strategy Wish for improvement of data dissemination and knowledge sharing. Requires the system to be more user friendly and flexible	Data used for different purposes: Tracking patterns and building relationship with stakeholders Main purpose of the data is to look at it and find patterns, hence build a sustainable market access and sales strategy System inflexibility result in low data quality

motivated using it and using it as a data collecting system. So, the data itself does not need improvement. It's more the system that needs improvements in terms of better usability and a higher speed. It could be valuable if the system was more targeted to B2B.	Data does not need improvements as the systems does.	
Q: Are you doing anything to improve it? Do you allocate any resources for that?		
K: Yes, I believe that. Many of my tasks are to ensure that we use Salesforce and the tools provided and try to expand both the sales and market access team to use it even more. This is both, so we can use it as a daily work tool for our extern collaborations but also so that we can collect information and use this information for registering of the customers and sending reports to global. The challenges are maybe the size of the system. Salesforce is a huge system, and when we are located in small Denmark it can be difficult for us to get individualized solutions to the systems. Hence, we become subject for a big system, meaning the system is customized for a big company trying to solve the solutions in the big countries. We don't really have many opportunities to change anything in the system and need to just use the tools provided. This can be seen as a challenge. Once again, the data is not the challenge, but the system is.	Salesforce and tools try to expand both the sales and market access team. System tools providing different solutions to different problems	Different approaches are needed to tackle different data concerns. The systems provide different tools for that
Q: What kind of problem can little DK not solve in this big system?		
K: As mentioned before it would be problems related to the speed and the usability. Sometimes our salespeople find it complex and too slow. It's a big system with many processes, which results that people on the road (like salespeople) find it very too demanding. But also, in the version we work with it is difficult to connect with activities and campaigns, as it is not really intuitive and not super easy to extract data reports.	System is too demanding and time consuming, flexibility needed.	System structures result low data quality and usage
Q: To what extent do you experience the data to be correct, error-free and accurate?		

K: From 1 to 5, where one is poor and five is accurate, I would say three. **Q: Is the data interpreted or processed in any way before you receive it? How raw do you consider it to be?** **K:** No, it's not. The data received is structured by me, so it's not interpreted or processed before I receive it. I structure and process it, so it makes sense for my manager, but also for the team I am a part of. **Q: To what extent do you accept the data as being true and credible? In terms of its sources; how much do you trust the data?** **K:** It's always subjective. So, there is no true in the data. But I think it provides a good picture of what is happening out there. **Q: To what extent are the data beneficial and provide advantages from its use?** **K:** I think the data provides a big value, from the fact, that we need that data and knowledge on what municipalities think and how they work in order to plan and organize our solutions and tasks and figure out when to contact them and provide our service for them. **Q: To what extent is the age of the data appropriate for the task at hand?** **K:** Well, this should be eliminated in the systems where we use it, so we continuously update the information in it. For instance, when a case gets closed, we should update the fact in the system, so we always have the most timeliness information available in the databases. The update is customer related, so the system gets updated depending on the interaction with the customers. **Q: To what extent is the data in an appropriate language and definitions clear?** **K:** It's easy for me, because the data received comes directly from municipalities, so there are no further struggles on understanding or interpreting the data. Since I am the first one received the data my interpretation of the data becomes the true. I am the middleman	Data is structured and interpreted by employees not the system ' Data provides good value and a general understanding of what is happening Data used to create strategy Regularly system update is needed. Results in freshness of data and information Data is raw and pure as it comes directly from municipalities and business partners	System structures the data and not employees Main purpose of data is to create a good strategy and gain a general understanding of the market

between the raw data and the data entered in the system. My interpretation of the data can be different from the true, and when I indicate the data in the system it would be biased as it's characterized by my understanding of the data. **Q: How easy is it for patients to access their own data? Do they know what kind of data is stored in our systems?** **K:** Based on the GDPR rules, they will always have access to their own data and they will always be able to demand to receive all the data that we have stored of them. But it's nothing that I directly work with. I don't have authorization to provide them their own data. This has to go through our Business Administrator. **Q: Do you think that the data has more potential for use and how can Roche DC use the data in another way more focused patient care?** **K:** In relation to the data I am in charge of, there are not much potential in terms of patient care, because we don't have patient data. But of course, in terms of the patient data there is a huge potential, but the GDPR is holding us back, since there are some clear and strict rules about how to use it.	GDPR makes it more demanding. Data is not as usable as it used to be	GDPR implementation has created more vague data structures
Open	**Axial**	**Selective**
Informant: Rune Kværnstøm – CRM & Business IT Manager **Q: Would you start by briefly describing your position and role in Roche DC?** R: My title is CRM & Business IT Manager and it refers to be the middleman between the IT and business part when it comes to the business applications part in the company. I mostly cover areas of marketing and sales. This is very basically my title and role in the company. **Q: What kind of systems do you work with?** R: Today we primary work with Rex, which is our CRM system. I work a little bit with	CRM & Business IT Manager = IT & Business integration processes	

SAP. I work some with Tableau, which is used as a reporting tool. Then I also work with other different marketing applications, such as Dialogs, which is an intern Roche consumer service tool, where you can store repositories of documents. **Q: So how do you use Rex? What do you store and process in this system?** R: The application is primary used to organize our customers. When it said, then the primary purpose of the SAP system is to store the link of all our customers and stakeholders, so we can have it as a backup. Then REX has many other information stored, such as all the contacts, the stakeholders, we talk with. They are not placed in SAP. What meetings we have with the stakeholders, which opportunities we have with these, and which sales prospective we have. These are all the information stores and processes in REX. **Q: When you say "them" who do you refer to? Is that patients?** R: No, we don't have any patient information stored in any systems. We only have information about our stakeholders, such as, municipalities, companies and hospitals and doctors. **Q: What is the data used for?** R: There are three levels the data can be used for; 1. Sales people can use it for their own tasks, to administrate their daily work and find patterns. 2. Management have an interest in knowing what the salespeople do and try to analyze the sales patterns, so we can see if we can target our intentions better. 3. Global looks at the data to see how each country operates. This does not happen often but happens sometimes. **Q: How has GDPR affected the way you collect and use data?** R: All the contact information we have had on people, companies and hospitals, we have	REX, SAP, Dialog & Salesforce used in DIA Data used for several purposes, but all related to stakeholder relationship Municipalities, companies, hospitals and doctors are relevant. = not patient data Data used by - Sales - Management - Global	Data is mostly used only for investigating stakeholders and the market Data is used by sales people, local management people and global management people to derive patterns

been forced to contact them and tell them that we have data on them. And if they wanted us to delete it, it has to be deleted. This is just a process that comes with the GDPR rules, but this process is the one requiring most time and resources. This happens periodic, so once every quarter we take a look at it and everyone who want their information deleted we delete.	GDPR's impact on data storage and management = time demanding	GDPR has created fuzzy data boundaries
Q: Do you face any challenges with the data in your databases? R: There are some challenges related to the different kinds of data. Our master data which is not person data, we can keep for a long time, however, other data, such as person data is more difficult to store for longer time. When it comes to person data the biggest challenge is that people tend to change information or environments without notifying us. This leads us to have fake data on them.	Company data is easier to store than patient data, since patient data expires easier. Change results in fake data	Outside change has impact on the databases.
Q: Is the data used for a lot? R: It depends a lot what kind of data we talk about. I think the data quality differs a lot and is not that consistent in the company. Our customer data is very structured and organized and is used for what it should be used for. When it comes to contact person data, it is still reasonable, but more possibility for improvements. When we talk about opportunity data, it is more doubtable. So, it depends which objects we look at. But the master data is mostly correct.	Data usage levels: - Customer data - Contact person data - Opportunity data	The best and most usable data comes from customers, second most organized data is contact data, while opportunity data could be improved.
Q: What do management to become better at using the data? R: This is what we are trying to initiate right now. The main thing is to see what you have and what you can assume from it. We have different data in different systems, and we try to see what kind of data we have, what kind of issues we have and what kind of solutions we need and then we try to connect it in a way, so we can provide these solutions. For instance, all the people who have a user in Dialog, we can see what kind of data and documents they have and how they use it.	Different data in different system. What kind of problems can this data solve?	

Basically, we try to use the data to find issues, patterns and solutions to already existing problems.		