Anni Tyynelä

Designing a customer data model and defining customer master data in a Finnish SaaS company

School of Technology and Innovations
Master's thesis
Industrial Systems Analytics

Vaasa 2023

UNIVERSITY OF VAASA
School of Technology and Innovations
Author: Anni Tyynelä
Title of the thesis: Designing a customer data model and defining customer master data in a Finnish SaaS company
Degree: Master of Science in Technology
Discipline: Industrial Systems Analytics
Supervisor: Jyri Naarmala
Year: 2023 **Pages:** 82

ABSTRACT :

In this study a logical customer data model is designed and customer master data in the data model is defined for a case company. During the process of defining customer data and customer itself, the business glossary of a customer is defined to have clear definitions of a customer and to unify the vocabulary across the case company. Defining the important vocabulary ensures the base to define the customer data and customer master data. In addition, quality aspects are studied or ensuring high-quality customer data in the future. ‘

This study aims to understand what customer data in the case company is and model it to a logical data model to unify siloed operations, systems, and data. The case company is a Finnish Software as a Service company. It is in the middle of a merging process due to recent company acquisitions. The case company wants to have common customer data and customer master data. The case company does not have master data defined. It is important to identify, which data is critical to the business so that the case company can have the one truth and the development activities can be targeted into the right direction to ensure the most advantage.

The research method of this study is design research. The empirical part of this study is done by workshops, and there are two rounds of workshops. First round analyses the current situation based on the processes and different functions in the company, that are working with customer data. The outcome of the first workshop round is customer terminology and its definitions and the customer data model. The second round concentrates on iterative development of the terminology and customer data model, and further identifying development needs, restrictions, and possibilities of having a common customer data model and master data. After the workshops, the terminology and data model are developed with internal experts. Lastly, there is a review event, where the participants get to see and comment the designed customer data model and the identified customer master data.

After this study, the case company has a clear definition of what is a customer, and how it should be modeled in a logical data model in the future to have one common customer data structure to unify the case company. Also, the case company has the most important, necessary, common customer data, the customer master data defined. The next step after this study is to plan the implementation of the customer data designs of this study, taking into account the quality principles, that were defined in this study to support the sustainability of the designs.

KEYWORDS: master data, logical data model, customer data, data quality

VAASAN YLIOPISTO
School of Technology and Innovations
Kirjoittaja: Anni Tyynelä
Otsikko: Asiakastiedon tietomallin suunnittelu ja ydintiedon tunnistaminen suomalaisessa ohjelmistopalveluyrityksessä
Tutkintoaste: Diplomi-insinööri
Opinto-ohjelma: Industrial Systems Analytics
Ohjaaja: Jyri Naarmala
Vuosi: 2023 **Sivuja:** 82

TIIVISTELMÄ :

Tässä tutkimuksessa suunnitellaan kohdeyritykselle asiakastiedon looginen tietomalli ja määritellään asiakkaan ydintieto. Prosessin aikana määritellään, mikä on asiakas ja mitä on asiakastieto, ja sitä myötä tehdään sanasto asiakkaaseen liittyvistä termeistä. Sanaston on tarkoitus selkeyttää ja yhdistää asiakkaaseen liittyvää sanastoa kohdeyrityksessä. Tärkeän sanaston määrittely mahdollistaa asiakastiedon sekä asiakastiedon ydintiedon määrittelyn. Lisäksi tässä tutkimuksessa tutkitaan, mitä täytyy ottaa huomioon, jotta tulevaisuudessa asiakastiedot ovat korkealaatuisia.

Tämä tutkimus pyrkii ymmärtämään, mitä asiakastieto on kohdeyritykselle, ja mallintaa sen loogiseksi tietomalliksi, joka yhdistäisi siiloutuneita operaatioita, systeemejä ja dataa. Kohdeyritys on suomalainen ohjelmistopalveluita tarjoava yritys. Kohdeyritys on tehnyt lähimenneisyydessä yritysostoja, ja on nyt keskellä yhdistysmisprosessia. Kohdeyrityksellä ei ole yhteistä määriteltyä ydintietoa. On tärkeää tunnistaa, mikä tieto on kriittistä yritykselle, jotta yrityksellä olisi yksi yhteinen totuus asiakastiedoista, ja kehityshankkeet voitaisiin kohdistaa oikein, jotta voidaan taata suurin hyöty.

Tämän tutkimuksen tutkimusmetodi on suunnittelututkimus. Tutkimuksen empiirinen osa suoritetaan työpajojen avulla, ja ne järjestetään kahdessa kierroksessa. Ensimmäinen kierros analysoi nykytilannetta prosessien ja eri yrityksen toimintojen kautta, jotka ovat asiakastietojen kanssa tekemisissä. Ensimmäisen työpajakierroksen tuloksena muodostetaan terminologia asiakkaasta ja asiakkaan tietomalli. Toinen työpajakierros keskittyy ensimmäisen työpajakierroksen tulosten iteratiivisen kehittämiseen sekä tunnistamaan mahdollisuuksia, rajoitteita ja haasteita, jotka liittyvät siihen, että yrityksellä olisi yhteinen asiakastiedon tietomalli ja ydintieto. Työpajojen jälkeen terminologiaa ja tietomallia kehitetään yrityksen sisäisten asiantuntijoiden kanssa. Tutkimuksen viimeisessä vaiheessa järjestetään tilaisuus, jossa esitellään ja käydään läpi suunniteltu tietomalli ja määritelty ydintieto, ja osallistujilla on mahdollisuus kommentoida tuloksia, ja kommenttien perusteella tehdään viimeisiä pieniä tarkennuksia.

Tämän tutkimuksen jälkeen yrityksellä on selkeä määritelmä siitä, mikä on asiakas, ja kuinka asiakastiedot tulisi mallintaa loogiseksi tietomalliksi, jotta kohdeyrityksellä voisi olla yhtenäinen asiakastiedon rakenne, joka yhtenäistäisi kohdeyritystä. Sen lisäksi yrityksellä on määriteltynä asiakastiedon tärkein, välttämättömin, yhteinen ydintieto. Tämän tutkimuksen jälkeen seuraava askel on suunnitella luodun mallin implementointi ottaen huomioon laatuun liittyvät periaatteet, jotka määriteltiin tässä tutkimuksessa tukemaan suunnitelman kestävyyttä.

AVAINSANAT: ydintieto, looginen tietomalli, asiakastieto, tiedon laatu

Contents

Figures

Tables

1 Introduction

Data is an important asset and resource for any organization. It can be used to create valuable information, and by information can be created knowledge. Information and knowledge can be used in decision-making to react, prevent, or improve. Data is everywhere and the amount of it increases every moment. Every organization has data, but the management and utilization of the data is not as easy. Even harder is to manage and utilize data efficiently. In case of utilizing the full potential of data, one main requirement is to know your data: what data is available and where it comes from, what is the most important data, and how the data points are related to one another?

This study is a case study for a case company, that wants answers to those questions from the viewpoint of customer data. Customers are an important part of any business. That is why it is important to know the customers. To do that, the company has to know who the customer is. This all requires high-quality, current, relevant, and accessible data. The case company has a complex organizational structure due to multiple company acquisitions that have been made recently, and it is in the middle of acquisition and merging processes. Therefore, the case company has multiple operations, systems, and data sources, when it comes to customer data. They want to unify the most important customer data into a common database to unify the data structure internally. That would enable an easier environment for better data management and data quality management.

The target of this thesis is to define a customer in the case company and model the most important common customer data into a logical data model, define customer master data in it, and find out how high data quality could be ensured in the designed model. The case company is a Finnish software as a service (SaaS) company, which does not have common master data due to the company acquisitions and therefore a multi-company environment. This study is part of a large project, which aims to unify the case company into one coherent company in the long term. This study aims to understand

the concept of a customer in a multi-company environment and design a common data model which would enable the modeling of all kinds of customership.

In order to the data sharing and analyzing can work internally in the company, the concept and data must be coherent across the organization. The data quality should be high for effective and efficient data usage. The employees should agree on the terminology so that they can communicate without misunderstandings. With shareable customer master data, the organization can have one common truth of the most important customer information.

1.1 Case company and the research problem

This study is done for a company of a Finnish software as a service (SaaS) company. The company is in reality a cluster of a larger company, that has three clusters in total. The clusters work almost independently and handle their data and systems, so in this study, the cluster in question is referred to as the case company. The case company of this study is producing financial management software as a service. To ensure the confidentiality of the company, it is referred to as a case company in this study, and also no confidential names of company parts, systems, or data are revealed.

The case company has recently acquired four smaller companies, that became new product lines. Therefore, the case company is in the middle of a merging process. The long-term goal is to merge the operations, systems, and data into one united organization. This is not an easy or quick project, and it has a lot of complexity. The purpose is to do efficient and reasonable changes step by step having the long-term goal in mind. An important part of the merging is to take into account the differences between the product lines' businesses and operations, which may not easily convert or unify to one to another.

This study is done for the Data and Analytics team of the case company, which is part of the Internal Operations function. The function aims to ensure the company's daily operations run and develop efficiently. The function has many professionals that have expertise across the business, like marketing, customer service, processes, IT, and enterprise architecture. The function supports this study with their expertise and will be part of the workshops, but also part of the analysis and design phases of the study. This is an important part of the study because experts like this are the developers of the operations and systems, and they are familiar with the case company's operations comprehensively.

The research problem is that the company does not have a common customer data model. They have many different models, which are not coherent. That creates many unnecessary expenses, uses more resources than having common ways and tools for working, and complicates the workflows when the aim is to unify the company operations internally. The challenge is to understand the current situation to comprehensively recognize the problems, restrictions, and possibilities, and then design a future customer data model that could unify the data and would create a customer master data to represent the most important core data throughout the organization.

This study aims to clarify the situation from the viewpoint of customer data. The core aim is to recognize the customer master data in the case company. To do that, customer data points and flow are studied and defined into a data model, first analyzing the current situation, and then designing a new logical data model. The logical data model represents the customer data that is collected and used. In the designing process the challenges and opportunities, that are recognized during the study, are considered to avoid mistakes and harness the possibilities. The purpose of defining customer master data is to have one common truth of customer data, which could be relied on in any case. Part of this study is also to examine the practices that would ensure the high-quality of customer data in the new model because data is efficiently usable only when it is high-quality.

The purpose of the study is to identify customer master data and design a unifying logical data model of customer data. The study information is gathered through workshops, where related employees are involved. The employees are experts in their operations, and the wide expertise of many teams is utilized to create a comprehensive and efficient data model.

1.2 Research questions

This study has three research questions. They aim to design a logical data model of customer data and define the master data in it, ensuring the maintenance of good data quality. This study aims to find answers to these questions to unify the case company's operations, processes, and data:

1. What is customer data and how it should be modeled as a logical data model to support unified processes and operations in a multi-company environment?
2. What is the customer master data in the new data model?
3. How to ensure data quality in the new data model?

The first question aims to define comprehensive and coherent definitions of customer and customer data throughout the organization and design a logical data model of customer data. The data flow of the current situation is modeled, and it is analyzed. The current model is a complex entity with many overlapping systems, data, and processes, which are wanted to unify in the future data model. The outcome of the first research question is a logical data model, which answers to the collected challenges and gathers the good qualities to enable holistic and coherent design.

The second question aims to define the customer master data in the new data model. The case company does not have clear customer master data. The customer master data would ensure the one truth of customer data, which could always be referred to, and it could be shared through the organization. Defining the customer master data would also

ensure better data quality, and it would unify the data and data's definitions throughout operations.

The third question aims to ensure that the quality of customer data could be maintained throughout its lifecycle. The aim is to study ways of data quality assurance and implement them in the new data model. Good data quality affects many operations in the company, like reporting, functioning of systems, and the quality of customer relationship management.

The overall aim of this study is to find and design a functional model of the customer data and define master data in it. The work aims to unify data and therefore processes and operations of the case company.

1.3 Research scope and limitations of the study

This research concerns only customer data that is commonly important in the case company. In this context that means information about the customer, for example company name, business ID, and address information. The study examines the lifecycle of customer data throughout related business processes. Other data areas have been scoped out of this thesis, but this study is designed so that it would be scalable. The framework of this study can be developed and used also to the future work of master data and data models' investigations in the company. Customer data was chosen to be the first master data domain that is examined because it is one of the most important data domains in any company.

This research considers this research problem from the point of view of data. The research does not consider the technical details of systems or software, and in the later parts, it does not state which systems the organization should use. In addition, data privacy and security are not part of this thesis.

This study concentrates on master data, data quality, and data modeling and design from the data wheel of DAMA-DMBOK (Henderson et al., 2017) (see Figure 1). The main theories are referenced from the book of DAMA-DMBOK by Henderson et al. (2017). Business intelligence is an important area where master data will affect and will be useful, but it is not part of this study. Data governance, the umbrella term that includes all the different sectors of data management, is also referred to in this study because the concentrated sectors of data management are closely dependent on data governance.

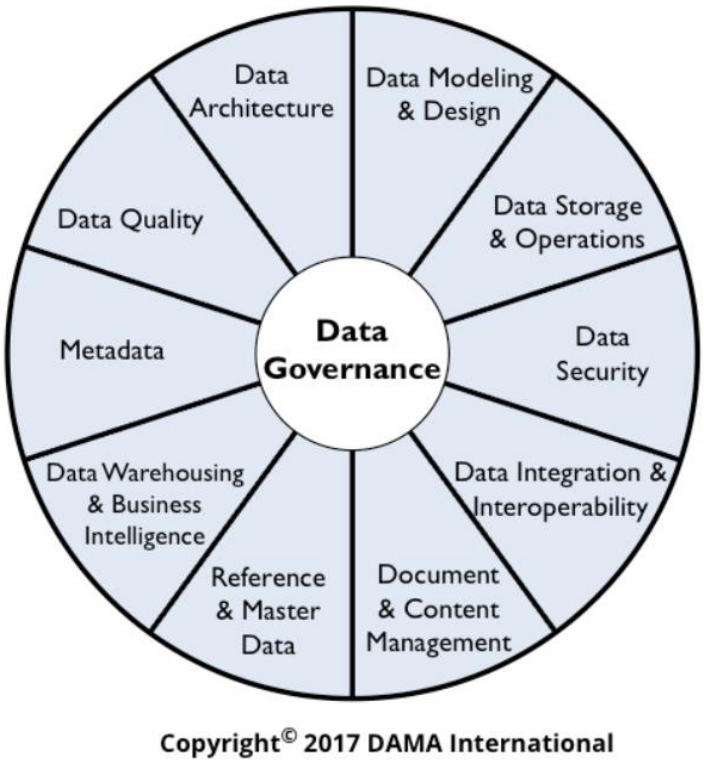


Figure 1. DAMA-DMBOK Wheel of data (Henderson et al., 2017).

When considering the viewpoint of the evolved wheel of data (see Figure 2) the scope of this study is the upper corner of the triangle, the plan & design. Use & enhance and enable & maintain are out of the scope of this study, but they are following actions of this project after this study. From foundational activities, data quality management is

part of the study, but data protection, security and risk management are scoped out. This study aims to build a functional plan and design that can afterward be deployed.

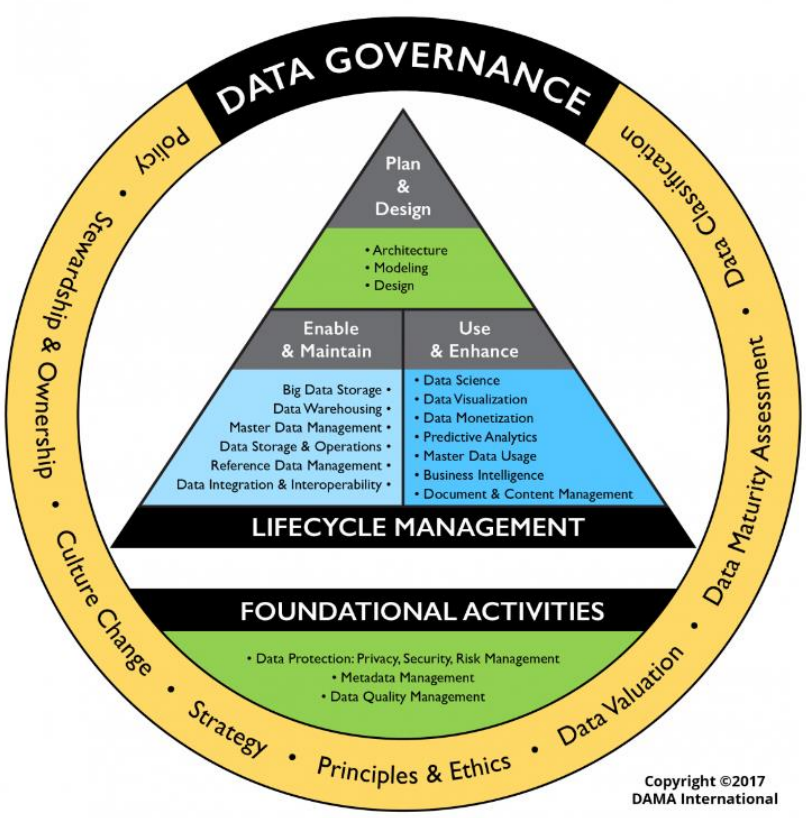


Figure 2. DAMA-DMBOK Evolved wheel of data (Henderson et al., 2017)**.**

1.4 Research structure

This research consists of an introduction, literature review, methods, results, discussion, and conclusion. The introduction outlines the research topic, research questions, the reason and importance of this work, and the scope of the research. The literature review explains the concepts that are used in this research, the main concepts being data modeling, master data, and data quality.

Methods will open the procedures that are used to perform this research. The empirical work of this research is done by workshops. The empirical part consists of three phases. The first two phases include workshops and an analysis of the outcomes of them, and the last phase includes a review. The first round of workshops is about analysis and the second round is about development. The review event presents the findings and allows the participants to evaluate the outcomes to examine the needs for final adjustments. Employees across the case company will be involved in the project, so the problem is considered coherently to form a data model and master data that comprehensively works for the case company. the findings of the research are covered in the result chapter. Discussion of the results includes thoughts and considerations of the study, and lastly, the chapter is summarized for clarity.

Conclusions conclude the research. Managerial implications are considered in the conclusions, and they present what kind of steps should the case company take based on this study. The future research section considers things that this study does not cover, but which could or will be part of the continuing work in this area in the case company.

2 Literature review

This chapter presents the theories of a data model and data modeling, master data, and data quality. The empirical study is based on this literature review. This study aims to design a new common logical data model of customer data for the case company and define customer master data in the model. In addition, the study aims to find how can the high-quality of customer data be ensured. The empirical part of the study uses the principles and concepts that are presented in this literature review.

2.1 Data modeling

According to Finkelstein (2006), data modeling is “a process that is used to identify, communicate, and record details about data and the relationships that exist between data, with its own terminology and conventions”. The aims of data modeling are understanding the organizational data and documenting the common truth from different perspectives of data (Henderson et al., 2017). (Witt, 2021) notes that data modeling can have different benefits, for example it can help a new systems to fit the business needs of an organization, or a data model can represent a common understanding of the data structures in an organization. This study aims to model a common customer data model, and one important part of the process is to understand the concept of a customer and identify the terminology around the customer data. Only by understanding the concept in-depth, the designed model can be functional.

A data model is a visual representation of data structure, and it helps to simplify complex systems into understandable forms (Blaha, 2010). Data models can be compared to process models because like process model tells what is happening and who is responsible for the happenings, data model tells what data exists and how it is structured (Simsion & Witt, 2004). Visual representation of compaction, like a data model or process model,

is easier to understand than a complex written report or a large visualization of different entities. Data modeling is also about understanding the business, its operations, and its terminology. That enables the model to work for the organization comprehensively.

According to DAMA-DMBOK by Henderson et al. (2017, p. 125) business drivers of data modeling are providing common definitions and vocabulary around data, documenting the organization's data and system structure, providing a communication tool for data and system-related projects, and creating the base for future development, like system changes or integrations. In addition, having a data model is important because accurate data models can be used in many development activities or risks can be detected from them (Henderson et al., 2017, p. 159).

(Simonin et al., 2012) present in their article that there are two reasons why data modeling is important from the viewpoint of enterprise architecture. The first reason is the business viewpoint: modeling the data is an illustration of the business heart. The data model should be aligned with the business processes. The other viewpoint is technology's viewpoint. The data model should satisfy the technical solutions of used information technology. A study from Samaranayake (2008) presents that data improvements can improve business process operations, which implicates that the viewpoints from Simonin et al. (2012) are not two separate viewpoints, but they can complement each other.

Simsion and Witt (2004) note that data modeling is an obligatory part of building a database, and Hunka and Matula (2016) state that the design of a data model plays an important role in database quality. In this research data modeling plays an important part of background research, so the master data can be defined and deployed in a way that enables the case company to work coherently from the point of view of customer master data.

Data modeling includes formalization, scope definition, and knowledge documentation (Henderson et al., 2017, p. 125). Formalization is defining data structures and relationships, scope definition is clearing the boundaries of data context, and knowledge documentation is gathering all related knowledge into the documentation of certain context for common use (Henderson et al., 2017, p. 125-126). In this study, the focus is on customer related data. The project is started by getting to know the complex big picture, followed with finding the most important data, and gathering more information of the more exact focus area.

There are three kinds of data models from the viewpoint of how technically informative the model is. The different data model levels are conceptual, logical, and physical data model (Simsion & Witt, 2004). The data models are like different steps throughout the journey of defining a database. Simsion & Witt (2004) present the relations of the different models (see Figure 3). There can be seen how the data model evolves from business requirements to a physical data model step by step.

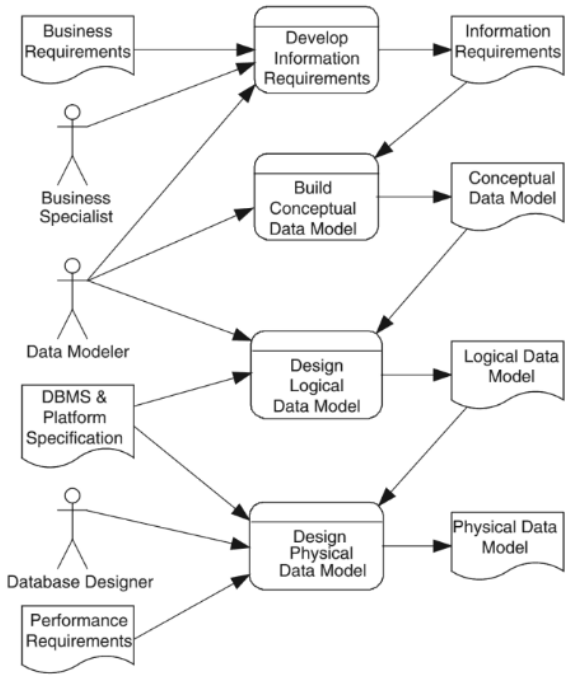


Figure 3. Representation of data models and their connections (Simsion & Witt, 2004).

The conceptual data model aims to understand the key terminology and concepts, and the logical data model is the answer to the business solutions (Henderson et al., 2017, p. 145). According to Simsion and Witt (2004), the conceptual data model is not a specific description of technical details, but more like a representation of the data handled from the business perspective. The conceptual data model is designed considering the business requirements, as can be seen in Figure 3. The conceptual data model is often modeled first, so the concept that the data model aims to represent can be agreed upon and tested before the model is developed into a more detailed logical data model.

The logical data model is a more detailed step of a conceptual data model, and it further specifies tables and columns (Simsion & Witt, 2004). The logical data model is a visual representation of the data and its relationships in an organization, which is tightly connected to the business needs but still independent from technologies and database implementation (Srikant, 2006). Srikant states that a logical data model must fit the business needs of the organization because it can lead to a better commitment to seeing data as an enterprise asset and ensures efficient and effective data storage. This study aims to design a logical data model of customer related data. The logical data model is a more technical representation than the conceptual data model. It includes information of what data exists in the model, but it does not state for example what are the forms of different data values, and it is independent from technologies.

The physical data model is the most technical representation of the different states of data models, and it describes every detail that is needed for the technical solution (Simsion & Witt, 2004). The physical data model concentrates on the technical details (Henderson et al., 2017, p. 148), which are not in the scope of this study. The physical data model is part of the following work and future research. The physical data model needs more technical knowledge and information, and usually, a database designer or similar expert is involved in the designing and planning, as can be seen in Figure 3.

2.1.1 Building a data model

According to Henderson et al. (2017, p. 152) in the DAMA-DMBOK, a plan for data modeling includes identifying requirements, creating standards, and determining the place of data storage. The deliverables of data modeling are the data model, definitions of the components of the data model, list of challenges and questions, and in many cases the lineage of data to understand the big picture of the model and the locations of data (Henderson et al., 2017, p. 152-153). This study aims to design a coherent customer data model but does not study where the database should locate. The study also considers ensuring high-quality data and aims to find ways to do that but does not create standards or policies.

Data models can either be built with forward engineering or reverse engineering (Henderson et al., 2017). Henderson et al. (2017, p. 153) present that forward engineering starts from the conceptual data model and proceeds to the logical data model and lastly, the physical model, but reverse engineering is used to document an existing data model, and it is started from the physical data model, proceeding to logical and conceptual data models. This study aims to model the logical data model of the case company's customer data for the future by forward engineering because there is no existing customer data model for all of the legal entities in the case company.

Henderson et al. (2017, p. 154) present the steps of designing and building a data model. They present that building a conceptual data model begins with selecting a scheme that is wanted to use. The six most common data modeling schemes are relational, dimensional, object-oriented, fact-based, time-based, and NoSQL (Henderson et al., 2017, p. 136). This study designs a relational data model. The relational data model includes the data and the relations between them, and it aims to have exact data and definitions and no redundancies (Henderson et al., 2017, p. 137).

After the framework is selected, the conceptual data modeling is done aiming to gather raw information from related employees. After modeling, the terminology and its definitions are defined and incorporated to avoid overlapping and misconceptions. The last step of conceptual data modeling is ensuring the quality of the model and having it confirmed with relevant facets.

Modeling a logical data model starts with requirement analysis of business process information requirements (Henderson et al., 2017, p. 154). Good analysis and design should include both views and their relations: data and processes. Building the data model begins with identifying entities, then adding attributes, then domains, and lastly keys (Henderson et al., 2017, p. 155-156). This study concentrates on customer domain and its entities and attributes, but the determination of keys is left out form the scope, and it is part of the next phases of the project.

In designing a database should be remembered the following five viewpoints: performance, reusability, integrity, security, and maintainability (Henderson et al., 2017, p. 159). The design's performance should be accessible and usable, and the data should be usable by many applications and for many purposes, like business intelligence and customer relationship management. That is how can be ensured that the model is durable and sustainable even if there are changes. Also, the data should always be necessary and accurate, the data should be available to all authorized users when needed, but only for the authorized users, and the design should me easily maintained.

The quality of a data model can be measured by content, level of detail, composition, consistency, and reaction to change (Henderson et al., 2017, p. 456). Furthermore, the data model should include relevant data and clear definitions, it should be detailed enough and precise with its content, the composition should be natural and proficient, it should be coherent, and lastly robust and flexible (Henderson et al., 2017, p. 456). The data model of this study is designed to be fit for use, coherent, agile, and justifiable to meet these requirements and to be high-quality data model.

2.2 Master data

Master data is necessary and critical information for the business, which is shared and used throughout the organization (Väre, 2019). Master data management aims to represent unequivocal data (Berson & Dubov, 2007). There is a lot of data in any organization, often even more than it is acknowledged. Master data is the most important, necessary and shared data. Berson and Dubov (2007, p. 11) state that master data management aims to have “a single version of truth”. Master data is therefore the data that can be relied on in any case, for example, analysis, reporting, or problem-solving.

Some relevant concepts must be covered when talking about master data so that the concept can be covered comprehensively and enough in-depth. These concepts are domain, entity, attribute, and meta data. First, a domain is a “complete set of possible values that an attribute can be assigned,” and which can be assigned data values based on either data type, data format, list, range, or some rule (Henderson et al., 2017, p. 135). In this study the domain in question is customer.

An entity is something that an organization collects information of, and it can be explained by answering questions who, what, when, where, why, or how, or with a combination of those questions (Henderson et al., 2017, p. 127). Allen and Cervo (2015, p. 17) present in their book that entity is defined as an object or item that is unique in a domain, for example, a business with a contract in a customer domain.

Attributes can identify, describe, or measure an entity (Henderson et al., 2017, p. 133). Attributes of a customer could be for example a name or an address. Metadata is commonly explained as “data about data” (Henderson et al., 2017, p. 417). In other words, metadata describes the data an organization has. Metadata can be for instance a visualization, some information, or statistics of the data in an organization. For example, data models are a form of metadata (Henderson et al., 2017).

2.2.1 Customer master data

Customer master data management focuses on the customer data domain and properties related to that (Cervo & Allen, 2011). Every company has customers, and multiple different business operations in the company are working with customer data. Every business operation can have its own database, and all the business functions value different data of the customer (Loshin, 2010, p. 4). When different business operations store data in their own databases, that can cause problems, like duplicated or contradictory data. Because different business operations maintain different data sets of the same business concept, like a customer, it is important that they agree on the business concept, so there can be a mutual understanding of the objective (Loshin, 2010, p. 5).

According to Allen & Cervo (2015, p. 34), customer and product domains are the most important data domains that should be concentrated on when starting a master data project. The case company of this study is a software company that offers software as a service for customers, so this applies also to this study. Customer data is an essential part of the business and identifying the customer master data enables improvements in the case company. Berson & Dubov (2007) state that having a customer master data enables organizations to have one main data to trust and follow, reduce the number of systems that handle the customer master data, analyze the data more accurately and create a comprehensive and complete view of the customer, and use the view to offer customers unique customer service. The customer master data was chosen to be studied first because it was seen to be the most important master data domain in the case company.

Customer relationship management can be anticipated as a process, strategy, philosophy, capability, or technological tool (Catalan-Matamoros, 2012). From the viewpoint of an organization’s internal operations, customer relationship management helps organizations in strategical development, value creation, multichannel integration, information management, and performance assessment (Baran & Galka, 2016). Considering more

externally directed operations, customer relationship management enables organizations to target profitable customers better, integrate offerings from multiple channels, and improve the efficiency and effectiveness of sales (Catalan-Matamoros, 2012).

From the viewpoint of data, customer relationship management aims to provide accurate and complete data of every customer without duplicates, redundancy, or overlapping data (Henderson et al., 2017, p. 366). (Buttle, 2009) states that databases of customer data are the base of executing customer relationship management strategy. Customer relationship management systems usually manage the master data of customers (Henderson et al., 2017, p. 366), but there are also many other solutions. This study does not consider the location of the customer master data. System-wise customer relationship management should collect customer data from every touch point, store the data, and analyze the data to information (Baran & Galka, 2016). To be able to do that, it is important to have reliable master data to provide working systems and data analysis. Challenges in customer relationship management and customer master data management are often complex roles and relationships of different facets, unique identification, differences in data sources, multiple communication channels, the importance of the data, and customer expectations (Henderson et al., 2017, p. 366-367).

Saarijärvi et al. (2013) found in their study of *Customer relationship management: the evolving role of customer data*, that using customer data efficiently and considering its role important can improve customer relationship management. They found out that with efficient customer data management, organizations can offer insight and knowledge also to their customers, not just use the data and knowledge internally. They present that with data-driven customer relationship management, organizations can find new ways to organize their customer data and therefore offer customers new kinds of services for example by identifying different segments. In addition, data management also includes consideration of privacy and risks, so customers could trust that their data is safe, which can improve customer relationships (Saarijärvi et al., 2013).

2.2.2 Master data management

The goal of master data management is to provide “the single source of truth for the business,” and it includes master data management services, like data governance and stewardship, which further include data quality, hierarchy, and data lifecycle management (Cervo & Allen, 2011, p. 9). According to Berson & Dubov (2007, p. 6-7), proper master data management is an important concept to adapt in any organization because of regulatory compliance of providing and using accurate data, privacy and data protection, and safety and security. Based on the study by Otto and Reichert (2010), master data management is both technical and organizational topic, which considers data quality as an integral part of the management operations.

Gregory (2011) states in an article that if an organization has an IT structure, it has data, and the data needs governance. Master data management is a more technical point of view of data management, and data governance is more about the organizational management and ownership of data. Data governance is related to roles and responsibilities (Hikmawati et al., 2021), or “exercise of authority and control” (Henderson et al., 2017, p. 67). Gregory (2011) also notes that data governance is both organization’s and individual’s responsibility.

Business drivers of master data management are having a common shareable master data through an organization, maintaining data quality, managing costs of data integrations by having a clear knowledge about the data structures, and reducing risks by simplifying data architecture and minimizing the complexity (Henderson et al., 2017, p.349). Loshin (2010, p. 21) states that implementation of a successful master data management will result in better competitive advantage when more effective integration between technology and business leads to better organizational collaboration and productivity.

Henderson et al. (2017, p. 350) present principles, that master data management should follow: master data must be shareable across the organization, master data should be owned by the organization and there should be data stewardship to ensure the master

data quality with continuous monitoring and governance. (Kokemüller & Weisbecker, 2009) note that master data management is a fundament of having high-quality master data, because the enterprise data relies on to the underlying master data. Also, changes in master data should be managed and controlled well, and master data should be authoritative so that master data values would not be replicated in any place, and they would just be referred to (Henderson et al., 2017, p. 350).

Determining the master data in the organization includes identifying what data is repeatedly used, what data is used to describe the most important things, how data is structured, where data is created, stored, and accessed, how data changes during its lifecycle, and what criteria define the quality and reliability of the data (Henderson et al., 2017, p. 359). Master data management needs a proper data model to describe the big figure in the organization from the viewpoint of data. Clear and consistent definitions are an important aspect of master data, and a data model aims to find them across the organization, and therefore a data model supports and enables master data (Henderson et al., 2017, p. 361). In this study, master data is modeled into a logical data model. A logical data model guides the implementation of master data (Henderson et al., 2017, p. 360).

Data modeling is an important part of the master data management process because it aims to present clarity and consistency for data and definitions (Henderson et al., 2017, p. 361). Entity resolution aims to find out if one object can have multiple definitions and to clarify them so that every data point has only one definition, and for one definition there would be one object (Henderson et al., 2017, 362-363). DAMA-DMBOK by Henderson et al. (2017, p. 361) considers data model management and entity resolution as key processing steps of master data management. This study concentrates on data modeling, but also on entity resolution in the case of designing a functional data model.

When speaking about sharing master data, there are alternative architectures, that are usually used. According to Henderson et al. (2017, p. 369-370), master data can be stored in a registry, transaction hub, or a consolidated version of them. A registry is like a list of

the master data points and their locations. A transaction hub is a system that contains the master data, and different systems interact with the hub to access the master data. A consolidated version of these two solutions can work so that there is a registry of master data, but it is managed in the systems it locates. Depending on where the master data is stored, the sharing is also different. Every organization should decide on a solution, that fits the best to their architecture. This study does not include the planning of the location of the data model, but it is an important part of the following steps after this project, so it is important to consider on the background.

Data stewardship is the tactical management of data, and it concentrates on data content, context, quality, and development of data management (Allen & Cervo, 2015, p. 12). Data stewardship usually belongs to a business that works with the organization's IT, and data stewards identify the business needs and requirements, and plans implementations fit for the specific use (Allen & Cervo, 2015, p. 12). Data steward's responsibilities are integrity, accuracy, and privacy (Loshin, 2001, p. 42). Ownership of data is a complex case because many facets collect, use, and produce data. According to (Loshin, 2001), an organization has internal and external data producers, and operational, tactical, and strategic data consumers. Otto and Schmidt (2010) also find in their study that definitions of ownership for business objectives are important part of defining master data architecture.

(Haneem et al., 2017) present in their study that there has been a lack of in-depth studies of master data management risks, even if acknowledging the risks is a necessary part of decision-making in the implementation process. Their study identifies many strategy, process, people, and technology related risks. Therefore, can be stated that master data management implementation is not an easy task, and many aspects must be considered in case of avoiding risks.

Das and Mishra (2011) found out in their study that the main challenges of master data management are duplicates in customer information data, integrations of data and applications, coherent data, issues in data governance, and metadata management. This study aims to tackle these challenges by developing a common coherent data model, which supports high data quality. Das and Mishra (2011) also state that master data is both for the IT and business sides of an organization, and needs viewpoints and dedication from both. Similar findings has been done by Silvola et al. (2011), and they present that implementation of one common master data requires commitment from all around the organization, both from business and IT side, even though the master data management project is often lead by IT. Silvola et al. (2011) found out in their study that problems of implementing one company-wide master data are more organizational and process based, rather than technical problems.

2.3 Data quality

High-quality data should be clear, accurate, and complete (Henderson et al., 2017, p. 450). According to Loshin (2001, p. 10), low data quality can decrease operational efficiency and constrain data-driven decision-making, and eventually lead to customer dissatisfaction. Data, which is high-quality, is reliable, and can be trusted (Henderson et al., 2017, p. 448). Can be stated that data quality is not only a technical concern but an important part of business, hence high-quality data can notably affect the efficiency of the business, positively or negatively.

Data quality as a concept is not only that the data has characteristics of high-quality data, but it also includes the processes to measure and improve the quality of the data (Henderson et al., 2017, p. 453). A successful data quality program would increase the value of data and improve the possibilities to use it, reduce costs and risks that are caused by low-quality data, increase the organization's efficiency and productivity, and help to pro-

tect the reputation of the organization, that could be harmed by low-quality data (Henderson et al., 2017, p. 452). Poor data quality can lead to mistakes, extra work, misinterpretations, or other extra costs of unnecessary resource usage.

High-quality data is not the same for different organizations even if the characteristics are the same or similar in every case, because data quality is also measured by it being fit for purpose (Henderson et al., 2017, p. 454). Therefore, the purposes and needs of the data must be identified with stakeholders, so the quality can be managed properly. The task is not an easy one, because nowadays organizations have huge amounts of data. It is beneficial for any organization to relay know their data. In addition, not every data value is as important as another. Data quality should concentrate more on critical data, like master data as Henderson et al. note (2017, p. 454).

Bahgi et al. (2013, p. 1) state "High-quality data is one of the most important prerequisites for making strategic business decisions and executing business processes," in their study *Controlling Customer Master Data Quality: Findings from a Case Study*. Therefore the quality aspects must be considered when designing a data model of master data.

2.3.1 Barriers to data quality

Haug and Arlbjørn (2011) write in their research paper that data is an important part of decision-making because it is included in almost all operations in an organization, but many organizations do not concentrate enough on data quality. According to their research about barriers to master data quality, five overall barriers are lack of delegating master data maintenance responsibilities, lack of rewards for ensuring high-quality master data, lack of routines or data control, lack of employee competence, and lack of usability of the master data management software. As an interpretation, according to their research, master data quality barriers are more likely to be a lack of management and knowledge, than technology based.

Haug et al. (2013) found in their study that when comparing smaller and larger companies, they tend to have different master data quality barriers. Both smaller and larger companies had issues with technology and IT systems, but for larger companies the organizational issues were more severe. From the study results can be interpreted that when an organization grows, like the case company has done, it is important to concentrate on the data governance as its totality, not only on the technical side.

According to the research of Ibrahim et al. (2021) that studied the factors that affect data quality and master data quality, they found that 19 factors affect the quality of master data. The most discussed factor was data governance, information systems and data quality policy and standards. The next most discussed factors were data quality assessment, integration, continuous improvement, teamwork, data quality vision and strategy, understanding of the systems and data quality, data architecture management, and personnel competency. The least impactful factors were top management support, business driver, legislation, information security management, training, change management, customer focus, and data supplier management. Almost a third of the factors are managerial in that study. The next biggest sections are organizational and stakeholder sections, and lastly external and technological. The managerial section is the most relevant when considering the number of occurrences in the studied papers. Next relevant are organizational factors, then technological, and lastly stakeholder and external sections.

The quality of master data is highly affected by managerial factors because it has the most influence on master data quality, can be interpreted that in the study of Ibrahim et al. (2021). The next influencing factor is an organizational section, which has for example the most influencing factor in it, data governance. Data governance includes having an organizational structure for managing master data quality, which further means that the organization has clear definitions of roles and responsibilities for managing master data (Ibrahim et al., 2021). Master data management roles can be divided in different ways: in strategic, managerial, and operational levels, or by ownership of concept owner, support function, and domain level (Ibrahim et al., 2021).

2.3.2 Data quality issue causes

Henderson et al. (2017, p.468) list five root causes of data quality problems, which are lack of leadership, data entry processes, data processing functions, manual issue fixing, and system design. Lack of leadership as a cause of data quality problems indicates that high-quality data needs organizational commitment in the forms of both governance and management. Data entry processes are a critical part of the data lifecycle, and errors can be caused by systems or humans. Data processing functions mean technical actions in systems, that can cause problems, like changes or misunderstandings in data sources, business rules, or data structures. Fixing data quality issues manually can cause data quality issues, because the possibility of human error can be high, due to multiple aspects of possible effects to consider.

According to Henderson et al. (2017, p.467-468) system design-related issues that can cause data quality problems are for instance data model inaccuracies, like setting different data format than is needed in some data fields, or poor referential integrity, which causes data flow errors between hierarchical data tables. In addition, coding issues in data processing or data mismatches, like using multiple date formats are system-related issues that can cause data quality problems. Moreover, other issues of system design are not having unique constraints to have coherent data, having unnecessary duplicates of data values, or reusing data fields for multiple purposes. Also, immature master data management can cause data quality issues if system designs are not managed properly. Can be interpreted that having a functional and defined data model and master data, which are part of this study, can already improve the quality of data.

(Lee et al., 2006) list ten root reasons for data quality problems, which are having multiple sources and therefore same values of data, using personal judgment in data protection, having limited computing resources, trading between security and accessibility, coded or hard accessibility data, complex data representations, large data volume, too restrictive input roles that can cause information loss, changes in data needs, and complex distributed systems. Can be stated that enhancing high data quality is not a simple

task, because it is affected by systems and people, it has to be considered in creating and using data, and it has to be taken care of by policies and standards, but also by repairing actions.

2.3.3 Data quality enhancement

High-quality data is necessary for a well-functioning company. It enables good reporting, relying on the data, good customer relationship management, and many more things. According to Henderson et al. in DAMA-DMBOK (2017, p. 450), the effects of high-quality data are discussed and understood more than the effects of low-quality data. Hikmawati et al. (2021) present in their *study Improving Data Quality and Data Governance Using Master Data Management: A Review,* that master data management processes can improve data quality. Master data management also encourages improving data governance and moreover, data governance impacts data quality ensuring the maintenance of high-quality master data.

Enhancing data quality is not a project, it is a program (Henderson et al., 2017, p. 450). That means that data quality can not be improved sustainably as one project, but the plan and the work must be continuous. The focus of a data quality program is making a data governance plan that fits the organization, defining the metrics and standards for controlling the data quality, defining processes of monitoring the data quality, and identifying the development opportunities to change systems and operations to generate better quality data (Henderson et al., 2017, p. 452). Data quality programs should be guided by criticality, lifecycle management, prevention, root cause remediation, governance, standard-driven, objective measurement and transparency, embedded business processes, systematically enforced, and connected to service levels (Henderson et al., 2017, p. 452-453), which are further explained in Table 1.

Table 1. Data quality programs should be guided by these principles (Henderson et al., 2017, p. 452-453)**.**

Principle	**Description**
Criticality	The focus should be on the most critical data.
Lifecycle management	The whole data lifecycle from creation to processing and disposal should be covered.
Prevention	Data quality management is not only correcting errors, but it should also concentrate on prevention.
Root cause remediation	Data quality management should always look for the root cause and solution for that in case of errors, not only correcting errors when they occur.
Governance	Data governance and data management should support each other.
Standard-driven	There should be rules and requirements for high-quality data so that the quality can be measured.
Objective measurement and transparency	The measurements should be done objectively, and the finding should be openly communicated with stakeholders.
Embedded business processes	Business process owners are responsible for the quality of the data they process, and they should enforce that the data quality standards are met.
Systematically enforced	System owners are responsible for that systematically the data quality standards are met.
Connected to service levels	Reporting and issue management of data quality should be incorporated with Service Level Agreement.

Data quality improvement actions require preparation to be effective and precise. In Figure 4 there are seven steps that Henderson et al. (2017, p. 4573-477) present. To be able to achieve high-quality data it is important to first define what it means to the organization and its operations (Henderson et al., 2017, p. 473). After understanding what high-quality data for the data values in question is, a data quality strategy can be defined. Data quality strategy includes methods to understand and prioritize business needs and

plan and implement actions, like standardizing integrating controls, to achieve better quality data (Henderson et al., 2017, p. 474).

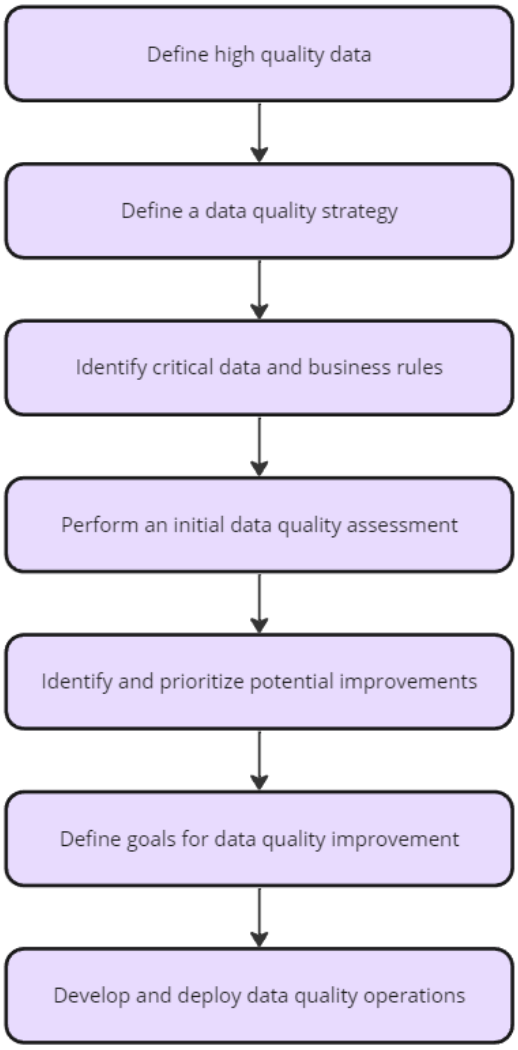


Figure 4. Data quality improvement process (Henderson et al., 2017, p. 473-477).

Some data should be prioritized in data quality improvements, for example, master data (Henderson et al., 2017). Master data is commonly the most important core data in an organization, and having it identified can notably affect data quality improvement projects' impact by prioritizing the important issues.

Data quality assessment is a more technical investigation of the data quality and the systems around the data. Data quality assessment aims to identify the technical causes of the data quality issues (Henderson et al., 2017, p. 475). Data profiling is a form of data quality assessment (Henderson et al., 2017, p. 470), and it aims to get to know the data. Data profiling enables identifying the appearance of data and is the groundwork for standardizing and requirement setting, issue initializing, and root cause analyses of deviations.

Data processing is actions that aim to improve data quality (Henderson et al., 2017, p. 470). Furthermore, data processing is data cleansing, data enhancement, data parsing and formatting, and data transformation and standardization. Data cleansing is correcting faulty data values, implementing controls to manage data entry, or improving business processes, that create data quality (Henderson et al., 2017, p. 471). Data enhancement is adding attributes to increase data quality and usability quality (Henderson et al., 2017, p. 471). Data parsing and formatting means comparing data to the rules that guide it, and it aims to find patterns of invalid data so that the root cause would be detected and corrected by transforming and standardizing new rules quality (Henderson et al., 2017, p. 472).

After the root causes are technically identified, the improvement needs must be identified and prioritized. The goals of quality improvements are set to guide the projects. Finally, after all preparations and research can the data quality improvement operations be developed and deployed. As was stated before in the literature review, data quality management should not be a project, but rather a program. That is why after this kind of data quality improvement process, it is important to remember that data quality management is iterative and continues after the deployment.

ISO 8000 is an international data quality standard, and ISO 22745 is a standard for defining and exchanging master data (Henderson et al., 2017, p. 451). The two main goals of ISO 8000 are to remove as much ambiguity as possible and to have master data in a

portable format so that the exchange between systems is possible (Talburt & Zhou, 2015). ISO 22745 is defining more specific instructions for implementing ISO 8000 (Talburt & Zhou, 2015). ISO 8000 is a set of standards, ISO 8000-110, -120, -130, and -140 (Talburt & Zhou, 2015). These standards go deeper than this study, but they should be considered when the project is proceeding to the implementation planning phase.

3 Methods

In this section, the research method and the process of the study are presented. The academic approach of the research method and the theory behind it are introduced first. After that, the more practical execution of this study is explained.

Research problems can be divided into two main types: nomothetical and normative (Helo et al., 2019, p. 14). Nomoethical research problem aims to find out what is the situation now, and normative seeks how things could be done in the future (Helo et al., 2019). The research problem of this study is that the case company wants to find a logical data model for customer master data, which fits the multi-company environment and unifies the data while ensuring the good quality of the customer master data. This study's research problem is therefore normative because the focus point of the study is on the future solution. The problem also includes elements of a nomothetical research problem, because to be able to design a solution for the future, the current situation must be examined.

This is a case study of an actual problem in a case company, which means that the study is empirical. The research will be qualitative, and the data and information will be gathered in workshops. In the later parts of this study, the artifacts designed based on the information from workshops are developed with internal experts of the case company.

3.1 Design science

The research method of this study is design science. The design science research process produces design knowledge, which consists of problem, solution, and evaluation (vom Brocke et al., 2020). Simsion & Witt (2004) describe data modeling as a design process, as it needs, for instance, analysis, past experience, and creativity. This study aims to design a logical data model with wanted qualities and features to meet the objectives of the future.

Design science is a sub-method of operational research, but it is more design-oriented and often information system-related (Peffers et al., 2007). Compared to other methods, the design science method is more design-oriented than the problem-focused method, and it combines analytical and empirical methods for determining and creating system designs (Peffers et al., 2007). The problem and the solution of the design science process can be independent components, but the evaluation reflects both of them together (vom Brocke et al., 2020).

Figure 5 represents the conceptual framework of design science research by vom Brocke et al. (2020). The environment is the facts and elements, that together define the research problem (vom Brocke et al., 2020). The knowledge base is the tools and guidelines to solve the research problem and to conduct the new design (vom Brocke et al., 2020).

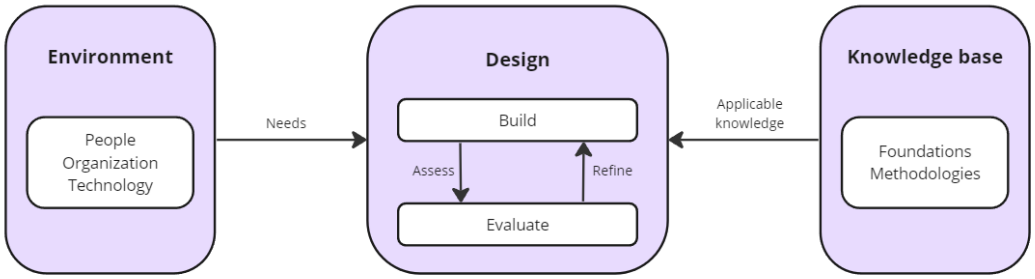


Figure 5. The framework of design science research method (adapted from vom Brocke et al., 2020)**.**

In this study, the environment is in the case company with a complex structure of the organization and complex architecture of systems and data. There are a lot of people, almost everyone in the company, that this customer master data affects. All of that together define the research problem. The knowledge base becomes from previous research, theories, and methodologies, which bring the theoretical knowledge that help to design the needed customer data model to solve the research problem. When the environment and knowledge are brought together, a solution can be designed. To have

an iterative development process, the design is evaluated and further developed during the design science process.

According to Peffers et al. (2007), the steps of design science research are problem identification, objective defining, design and development, demonstration, evaluation, and communication. The process model of design science research by Peffers et al. (2007) can be seen in Figure 6. The elements that are presented in the process model are also part of this study, and the progress of the study is adapted by these process steps. The viewpoint of this study is to design the customer master data model, and the implementation is not part of the thesis. Therefore, the demonstration phase is done at a conceptual level.

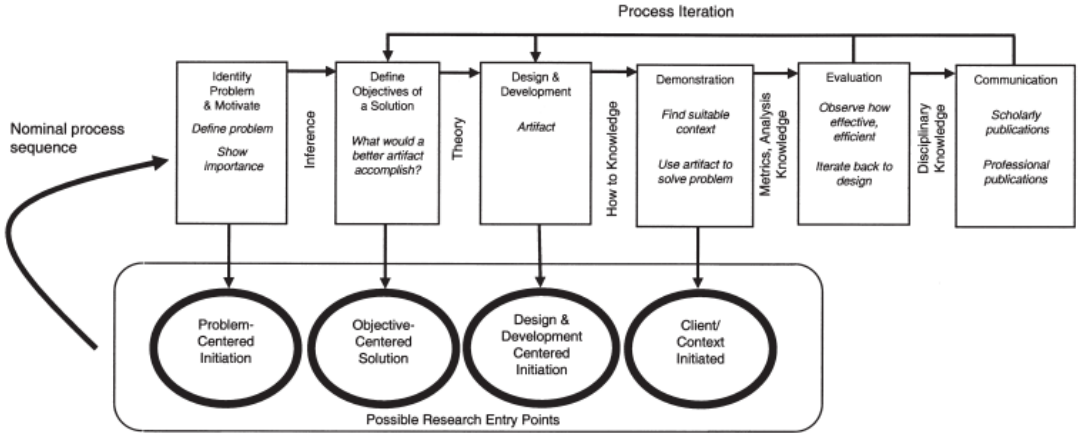


Figure 6. Design science research process model (Peffers et al., 2007)**.**

Like from Figure 8 can be seen, the process is iterative and does not certainly go through the process steps chronologically. There are also different possible research entry points presented in the figure. This study concentrates most on generating the “Design & Development Centered Initiation”, and the outcome is an artifact, the customer master data model. More specific explanations of the main process steps of design science research are presented in Table 2.

Table 2. Steps of design science research (Peffers et al., 2007)**.**

Problem identification and motivation	Define the specific research problem and justify a solution. Capture the complexity of the problem
Define the objectives for a solution	Objectives can be quantitative or qualitative but should address the role of the artefact in the solution
Design and development	Create the designed artefact in which the research contribution is embedded in the design
Demonstration	Demonstrate the use of the artefact
Evaluation	Observe and measure how well the artefact supports the solution of the problem. This may result in further development
Communication	Communicate the problem solution, its utility and novelty, and the rigour of the design to relevant professionals

In this study, the problem is identified, and the outcome of the study is structured and scoped. Then in workshops, that are described more detailed later, the objectives of the solution are defined with relevant employees by analyzing the current situation and identifying the development needs and opportunities. The logical data model of the future is designed and developed based on the outcomes of workshops. After the design process, the outcome is demonstrated to employees, and they can evaluate it in case of any further development needs. The implementation process does not include in this study.

3.2 Workshops

In this study, workshops are used as a data-collecting method. Workshop as a research method is related to action-oriented methods, which are group research, action research, action learning, and participatory research (Storvang et al., 2018, p. 3). Therefore could be stated, that workshop is a versatile concept. Workshops were chosen to be the main method of data collection because there were many people and many functions and

teams, who should be part of the process. According to Storvang et al. (2018 p. 9), a workshop has three types of participants: researcher, facilitator, and participants. As Storvang et al. (2018) note, sometimes the researcher is also the workshop facilitator, like in this study.

The four phases in workshops presented in figure 7, diagnosis, planning, facilitation, and analysis, are all based on the purpose of the research. The diagnosis phase aims to determine the research questions reflecting the purpose, identify the relevant people to participate in the workshops, and examine the existing practices (Storvang et al., 2018). The planning phase includes planning how the workshops are executed; where, when, and how they are done, what exercises are done there, and who will participate (Storvang et al., 2018). The next phase is facilitating the workshops, and there the facilitator should concentrate to manage, guide, and support the workshop to get the expertise of the participants to unite in the point of view of the research purpose (Storvang et al., 2018). The last phase is the analysis, when the gathered data is treated in some chosen way to get answers to the research questions (Storvang et al., 2018).

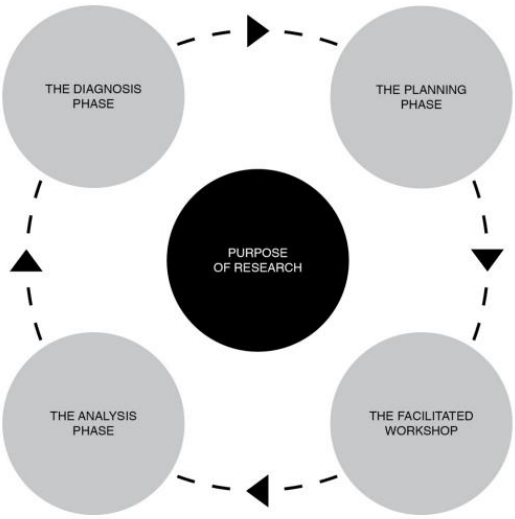


Figure 7. Workshop frameworks cycle that all round around the purpose (Storvang et al., 2018)**.**

3.3 Study plan

This study's empirical part is executed in three phases, which can be seen in Figure 8. Before the empirical part, the literature review is formed, and that is used as the base for the empirical part. The literature review forms the Knowledge base of the design, as can be seen in Figure 5. The first two phases include workshops and the third phase includes a review event. All the phases also include analysis and development. The first phase is called the Analysis phase, and it aims to analyze the current situation and environment. This phase aims to cover the second phase from the design science research process model (see Figure 6), which is defining the objectives of a solution and examines the environment from the design science framework (see Figure 5). The outcome of the first phase is an objective-centered solution (see Figure 6).

The second phase aims to cover the third phase from the design science research process model (see Figure 6), which is designing and development. the second phase aims to develop the design according to the knowledge base and environmental information. The outcome of the second phase is a design and development-centered initiation (see Figure 6). The third phase covers the fourth and fifth steps of the design science research process model (see Figure 6).

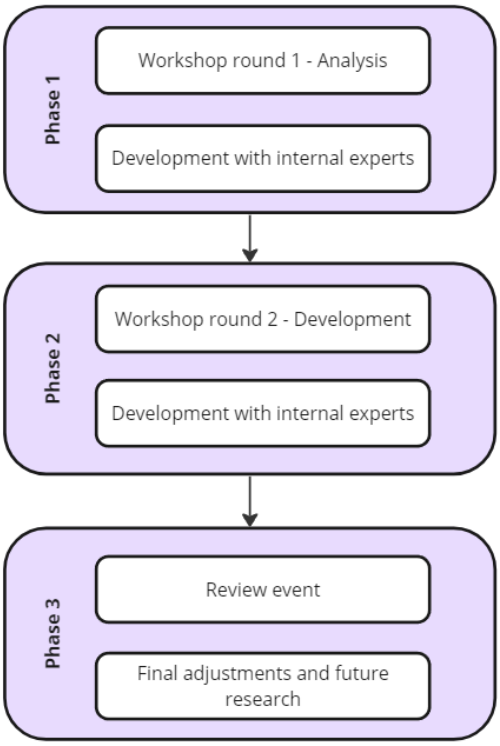


Figure 8. Plan of the execution of the empirical part of this study.

The workshops are held entirely remotely. Work-life has changed so that the majority of employees work remotely most of the time. The pandemic changed the way of working globally and especially in the technology field, where most of the work is done via computer or laptop. Therefore, online tools have become more and more popular, and they have developed during and after the pandemic. That is also a reason why it is reasonable to organize the workshops remotely.

For each workshop, there is a 1,5-hour meeting booked. Every workshop is booked approximately two weeks before because of the participants' busy schedules. Meetings are recorded with the consent of participants. That way it is possible to go back to the discussions when needed. The platform that is used to organize the workshops is Microsoft Teams. The case company uses Microsoft Office tools, so it was a clear choice of a platform to hold the workshops.

The base tool of the workshops is Miro board. Miro is an online white board tool, which helps easily involve co-workers remotely, visually build together what is wanted, and store it online (Miro, 2022). Miro board enables everyone to easily contribute to the workshops at the same time easily remotely. Miro board is a modern version of the physical whiteboard used in many in-person workshops, and it includes many similar objects, like sticky notes. The virtual sticky notes are Miro's well-known feature. Virtual sticky notes are familiar to many employees from the times before remote working. Workshops, which were held in person, used actual sticky notes to do the workshop exercises, so the virtual ones are an easily understandable concept.

There are 2 rounds of workshops and a review event. The workshop rounds are named Analysis and Development. Round Analysis aims to identify the used terminology and to find out the current data points. The second workshop round, Development, reviews the aggregated terminology and the data model, which is modeled based on the first round's gathered information. The aim of the second round is also to gather the challenges and opportunities of the current and future data models when the aim is to unify the customer master data between the product lines. The review aims to present the new customer master data model and the aim is to gather feedback of the new model for iterative changes and final adjustments.

The Miro board for this study is designed for this project. The scalability is considered, but the base is targeted directly for this project. Both workshop rounds' bases include a "To begin with" section. That section aims to open the most important terminology, the aim of the project, benefits for the participants, outcomes of the project, and the steps that will follow this project. The section aims to explain the purpose and meaningfulness for the participants so that they know what kind of project they are part of, and why it is important. The "To begin with" section also includes the timeframe of the first workshop.

The vocabulary and subject in this project are technical, and because of that, the most important terminology is explained at the beginning of the workshop, and the language

that is used to organize the workshop and explain the project and exercises, is as easy as possible without any unnecessary technical vocabulary, so that everyone, regardless of their background, was able to understand.

Both workshop rounds also include two steps of individual working with the virtual sticky notes. The exercises are explained and instructed first to make sure everyone knows what is meant to be done. Then there is a certain time allocated for individual work, which is followed by a discussion of things that were written into the board. Both workshops also end with a wrap up what was done during the workshop, and what are the next steps of the project.

The workshop base in Miro board was designed so that it would be as intuitive as possible to do the exercises in workshops. The different parts of the workshops were divided into different areas. The workshop's agenda is introduced first, and the exercises are explained before each exercise. The participants are informed that they can ask anything in any phase of the workshop, and before any exercise it is checked if anyone has anything to ask or if there is anything that needs clarification. The workshop bases in Miro are built from different shapes and lines, and they are locked so that the participants are not able to accidentally move anything that is not supposed to be moving. The sticky notes, that are used to write the thoughts and answers to the base are the only objects that are moving during the workshops.

The second round's workshop Miro board platform is built after the first workshop round. Storvang et al. (2018, p. 21) wrote that "the analysis will change the researcher's pre-understanding to a post-understanding of the problem", and that is why the second workshop rounds' Miro board base is developed after the analysis of the first workshop round. That enabled adjusting the process to the results of the first workshop.

The first workshop round, analysis, was decided to divide into three sections based on the processes: the first workshop analyzed marketing processes, the second opportunity,

contract, order, and invoicing processes, and the third customer onboarding, churn, delivery, service, and production processes. In these workshops participated different functions: marketing, sales, and finance, and lastly customer service and product functions.

The second round, development, was divided into two sections, where the first workshop analyzed marketing, opportunity, contract, order, customer onboarding, and churn processes, and in the second workshop delivery, service, invoicing, and production processes. The marketing and sales function will participate in the first workshop with part of the customer service and revenue function. The revenue function is responsible for the revenue in the whole case company operations, so they will also participate in the second workshop.

The employees that are invited to the workshops are from different sides of the case company, and the aim is to include one from each relevant function and firm. Thinking of the participants and groups that would be included in the project was challenging because of the complexity of the case company's organizational structure. Participants were chosen with the help of the Data & Analysis Team Lead. The first thought was to involve end-user employees, who would know phase by phase which information they are using. After iterating and considering the plan, a decision was made to involve only hierarchically one step higher participants, like team leaders or function leaders, and to inform them about the tasks and ask them to prepare so that they can present their teams' point of view in the workshops.

Before sending the invitations, the participants are contacted via email and asked if they are the right people to participate in this project. This aimed to make sure that the invited employees were the ones, who were familiar with the customer data that was used in their operations. The email included basic information about the project and the scope, and the subjects of the workshop were explained. The contacted people were asked to answer if they think that someone else would be a better option, or if someone should be also invited to the workshop.

The project has a process-based approach. In Figure 9 can be seen the workshops that are divided into three workshops in the first round and two workshops in the second round based on different processes. Processes are also connected with the functions. The workshops rounds are divided into smaller workshops, because of the large number of participants. Larger groups can lead to more discussion, and workshops wanted to enable as much discussion as possible, but at the same time keep the length reasonable.

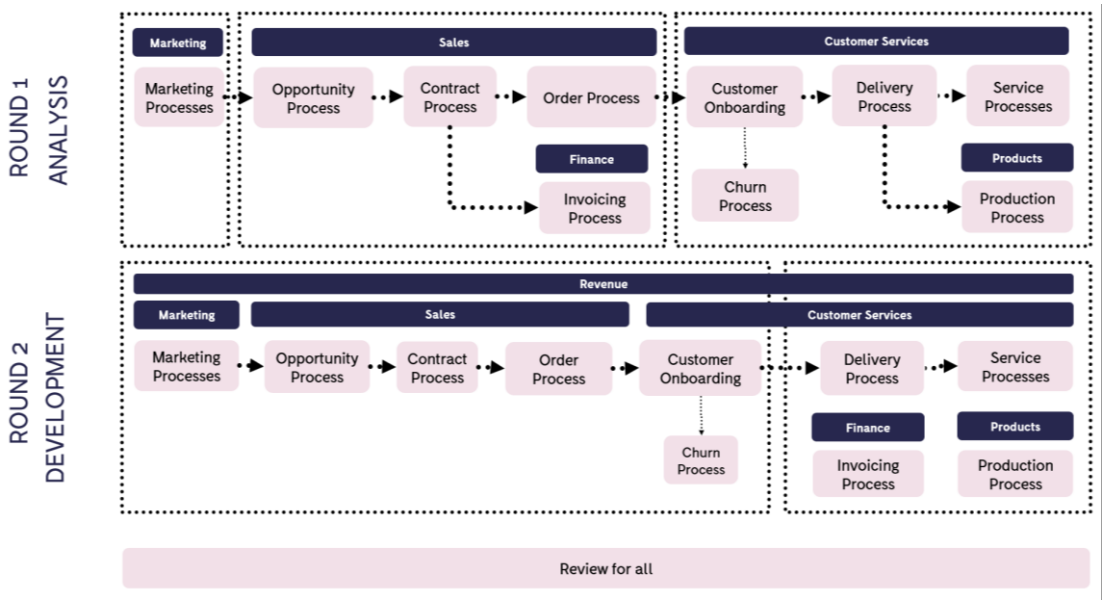


Figure 9. The participant functions and processes analysed in the workshop rounds**.**

As was previously stated, the case company in this study is one of the three clusters. Another cluster had started a similar project focusing on customer master data earlier. There were two discussion sessions with the project owners from the other cluster before and after the workshops. The project was executed there differently, but still, it was interesting and insightful to share knowledge at the beginning and the end.

4 Results

This chapter presents the result of the empirical part of this study. This chapter presents how the study finds answers to the research questions. First the results of the first workshop round's analysis and results are presented, then the second round's analysis and results, and then lastly the review event's analysis and results. The outcomes of studying the data quality in this case are presented in an own chapter. At the end of this chapter is a discussion of the empirical part of this study and lastly a summary.

4.1 Analysis

The workshop started with an introduction to the project's aim, targets, benefits, concrete outcomes, and next steps. The proper introduction was important so that participants could understand why they were in the project and what was the project about. Also, the main concepts, master data and data model, were explained so that the participants were familiar with the concepts that were talked about. The visualization of the introduction is presented in Figure 10.

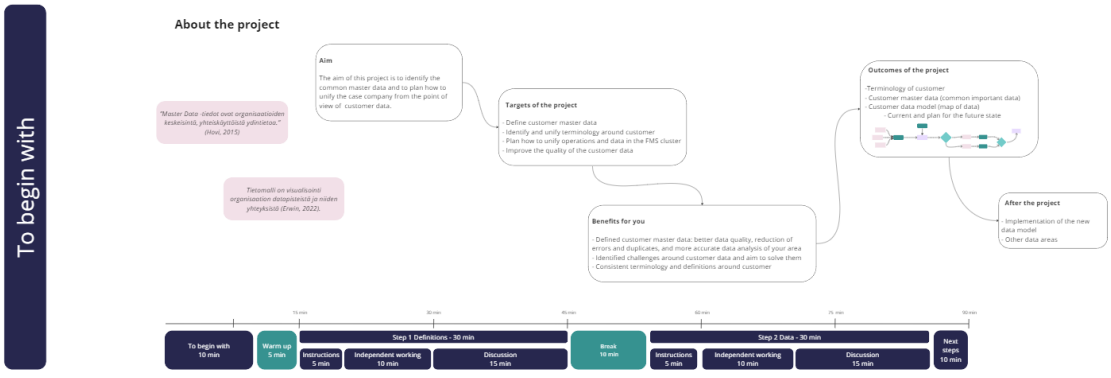


Figure 10. The introduction of the project and workshop timeline in the project's Miro board.

The introduction also included the timeline of the workshop. The workshop was divided into four main steps: introduction, definitions, data, and next steps. Definitions step

aimed that the participants listed terms and definitions around customer data, and the data step aimed that the participants listed customer data points that were used in their area of responsibility.

From Figure 11 can be seen the workshop's warm-up material, which is meant to teach the participants how to use the Miro board. The warm-up exercise aims to answer the question "Who are you and what are you doing in the case company?" by writing their name and role on a sticky note. Sticky notes are used in the later parts of the workshop, so this warm-up exercise is an important part of the project to teach everyone to use these virtual sticky notes. In addition, this exercise gave the participants an understanding of who the other participants were, if they did not know each other's roles beforehand. Also, this part of the workshop was a step of documenting who were part of the workshops.

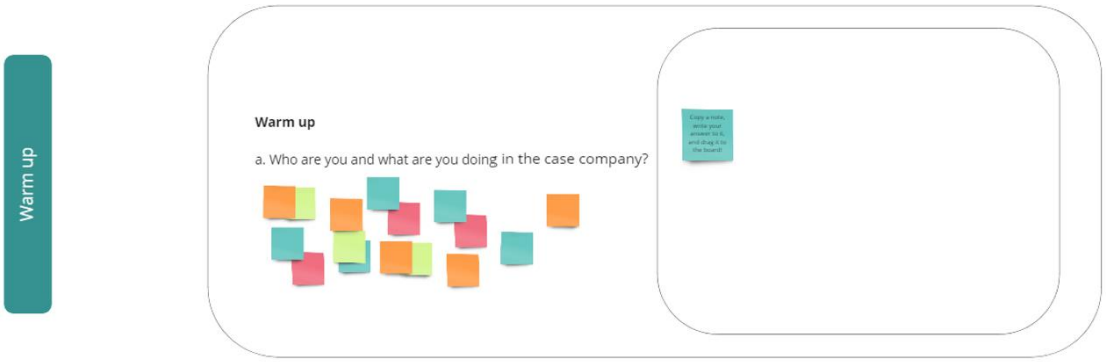


Figure 11. Warm-up section of the workshop.

Figure 12 represents the first workshop round's base. The workshop has two exercises: Step 1, the terminology, and below that the Step 2, data point analysis. The purpose of both exercises is to write sticky notes on the marked areas. Each product line has its own colour sticky note because the terminology and processes vary between product lines. There are three different areas in the workshop phase because of the three workshops that were organized in the first workshop round. During the workshops, the areas that were not in question were covered so that the participants knew which area is the one

the exercise was meant to be done. For example, the first workshop was from a marketing point of view, and during that, the other two areas of sales, customer service, finance, and product processes were covered so that they could not be seen or modified.

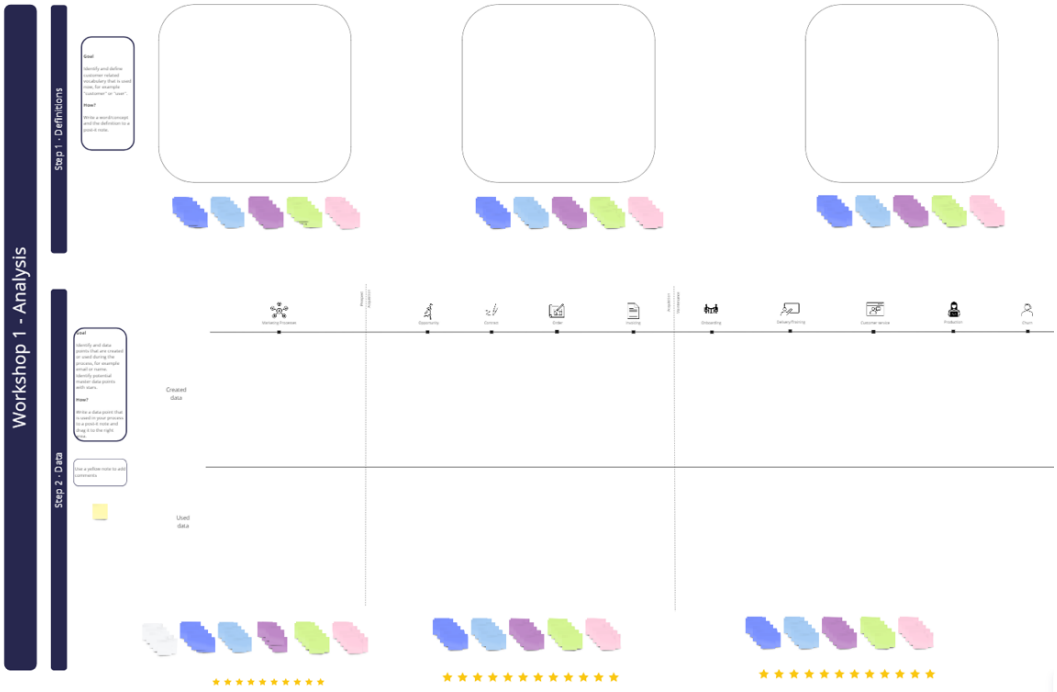


Figure 12. Miro board of the workshop round 1.

The aim of the first exercise was to list as many terms and their definitions of the customer as possible. The exercise aims to map the different terminologies in the case company from various points of view comprehensively. In the second part of the analysis workshop, the aim is to place sticky notes in the “created data” or “used data” area. The starts below the sticky notes are meant to mark the most important case company-wide shareable data, that could be potential master data. The participants could place a star on the sticky note if the participant thought that the data is important and could be master data. In the second exercise, there was also a light yellow sticky note on the left, and it was meant to enable commenting on any data point if there were any comments.

Figure 13. The overview of the outcome of the analysis workshop round's three workshops.

Altogether, 390 sticky notes were written in the analysis workshop round, which of 167 for the terminology exercise and 223 for the data point exercise. The final state of the workshop base with the sticky notes can be seen in Figure 13. The number of sticky notes was high, which indicates that there were a lot of data to analyze. Some invited employees could not make it to the workshops or had to leave in the middle of them, so after the workshop round, there were few individual discussions. The individual discussions included the same exercises as the original workshops, but often they were gone through more discussion based.

The analysis of the first workshop round was done in Miro. The sticky notes were copied so that the original answers remained, and the copied ones could be analyzed. Miro was a good platform for analysis because the sticky notes could be easily copied, moved around and for example categorized.

The terminology part of the workshop was analyzed by first dividing the terms into categories. The categories were customer, user, marketing, partner, accounting office, accountant, and customer service. All these categories had three or more relevant sticky

notes related to them. Customer and user were the biggest entities having both over 40 different sticky notes. There are several different angles and types of a customer and user, and that is one reason why there were so many different terms around them.

The categories were analyzed further, and the same terms' explanations were compared and combined if reasonable. The terms were divided into tables by their theme: the lifecycle of the customer, marketing terms, the customer's relationship to the case company, the user, and terms around customership, and they are presented in Figure 14. Not all terms were included in the final table because this study wanted to concentrate on the most important main terms and most basic terms to get them correct to be the base for future iterative developments.

Figure 14. Terminology based on the first workshop round divided into different categories.

The data points that the workshop participants identified were first divided into entities. Then attributes were assigned to the recognized entities. The entities along with the attributes were organized into a logical data model (see Figure 15). The conceptual data

model (see Figure 16) was also drafted based on the logical data model and the understanding that was gathered during the discussions in workshops and in individual discussions that were held with the experts in the case company.

The data models were tried to simplify so that the analyses would be as understandable as possible to the participants who were not so familiar with the technical side of their operations. Like in terminology, not every identified data point was included in the data models, if they were not recognized to be part of these data models, which concentrated on customer information.

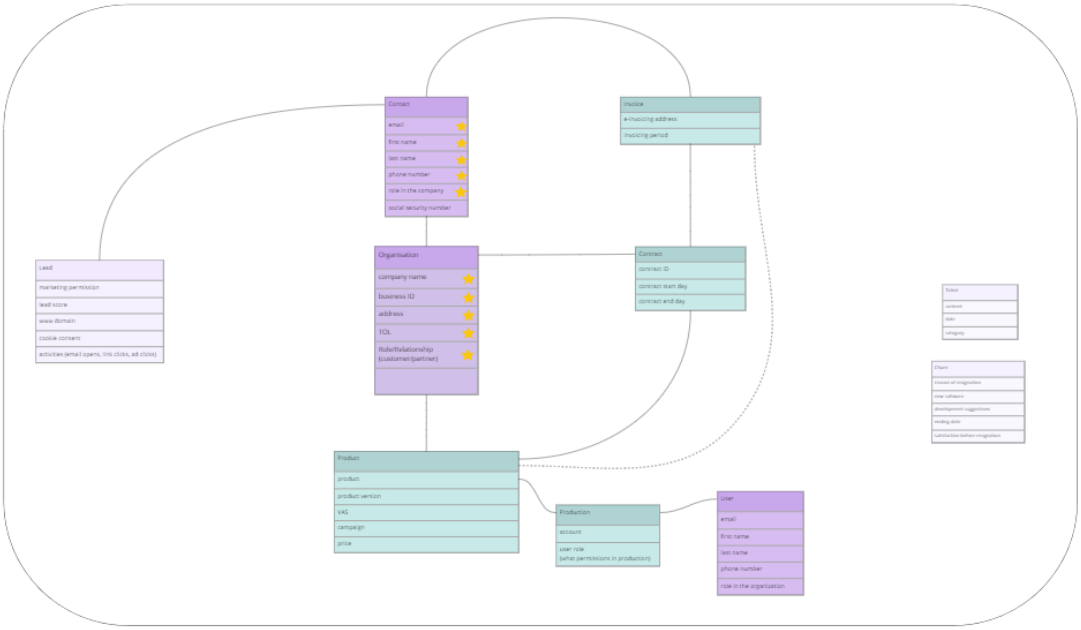


Figure 15. Logical data model from the analysis of the first workshop round.

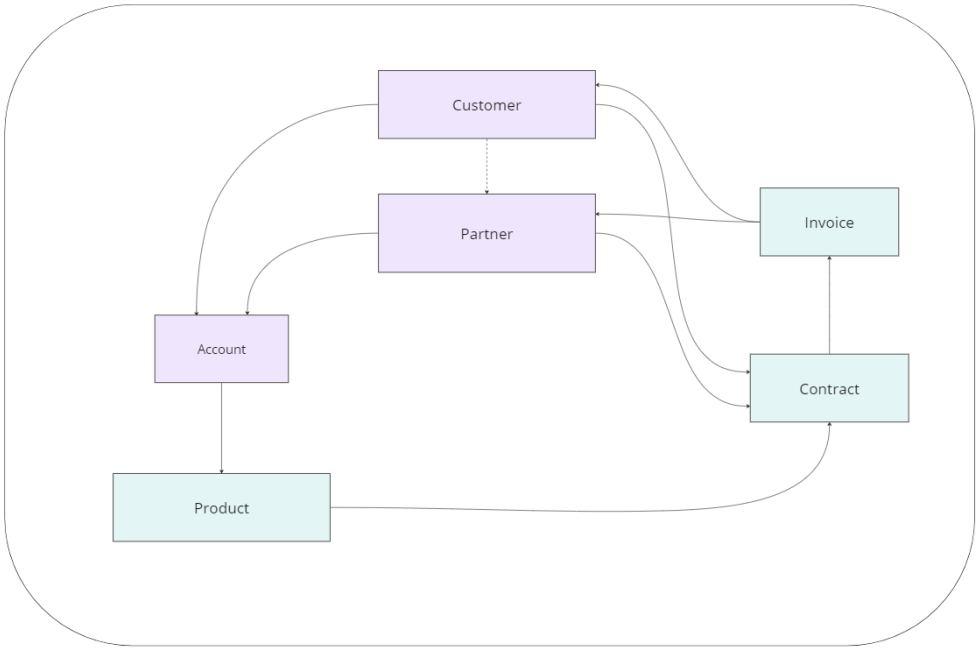


Figure 16. Conceptual data model from the analysis of the first workshop round.

4.2 Development

The second workshop round was about development. First, the outcomes of the analysis of the round one were iteratively reviewed and further developed. After that, the challenges, and possibilities of now and future were analyzed. The workshops were started with an introduction. The introduction included background information of why this project is done, and the agenda of the workshop with a schedule (see Figure 17). Because the last workshop's introduction was held already a few weeks ago, the introduction part was meant to remind participants of the project.

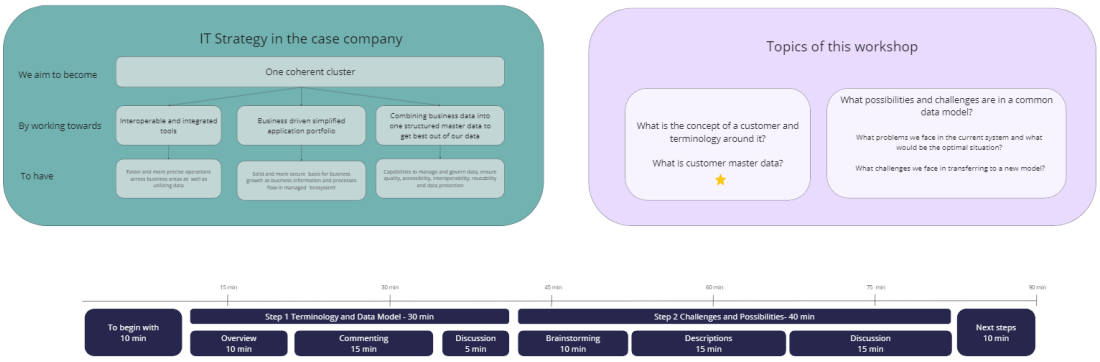


Figure 17. Introduction to the second workshop.

The introduction was held from the viewpoint of the IT strategy of the case company. The case company wants to integrate the tools, applications, and data of the product lines into one coherent case company. The starting situation is that the product lines, that have previously been own independent companies before the company acquisitions, have each their own operations, tools, and data. The integration is planned to take approximately two years, so the project is a long one.

As can be seen from Figure 18, the topics of the second workshop were to review the terminology and concept of the customer and the customer master data, and then think about the challenges and possibilities of current and future customer concept and data model. The workshop was divided into two steps: a review of the terminology and data models and a determination of challenges and possibilities. The review was meant to be an iterative development step of the project. Determination of the challenges and possibilities was meant to deepen understanding of the customer data related issues from the viewpoint of the experts in the case company. Like in the analysis workshop, in this workshop also in the first step included an introduction, individual work, and discussion.

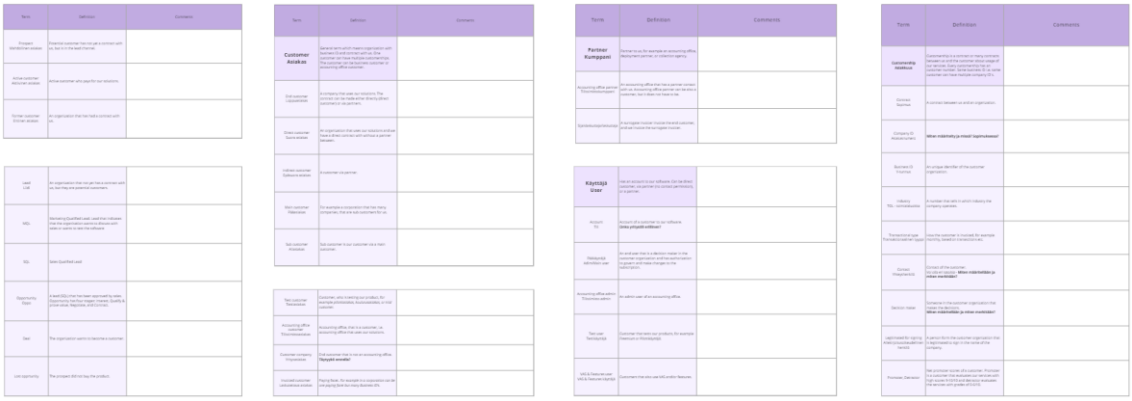
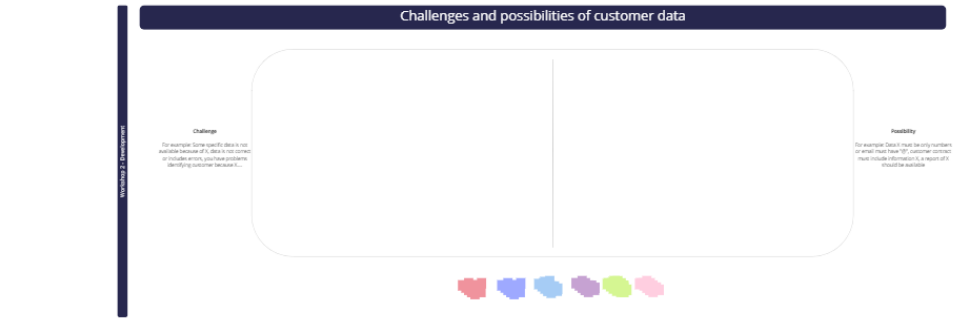


Figure 18. Tables of the customer terminology. There is a column for the term, a column for the definition, and then a column for comments.

The conclusions from workshop round 1 were presented at the beginning of workshop round 2 workshops to have iterative development to them. The data models were presented as conceptual and logical models (see Figures 15 and 16). The potential master data points were marked with yellow stars. The terminology was presented in tables divided into different entities (see Figure 18). The terms were presented in English and Finnish, because the working language of the case company is English, but many of the participants use the terms in Finnish, so it is clearer to present them in both languages. The descriptions of the terms were in English. The participants were given the opportunity to comment on the data models and terminology table with sticky notes.



Challenge											
Possibility											
Business area											
Comments											

Figure 19. The base of second workshop round.

The second step of the development workshop included two parts. First part aimed to brainstorm as many possibilities and challenges as possible about having a common customer master data to the area above. The second part was to pick the most important ones and elaborate them in the table below, which can be seen in Figure 19. In the table, participants could drag the virtual sticky note that they wrote in the brainstorming part to the table and elaborate and define the challenge or opportunity further. In the table, there is also a business unit row, where the participants could specify if the possibility or challenge concerns some specific business function or product line.

Figure 20. The overview of the outcome of the development workshop round's results of the review of data models and terminology**.**

The data models got a relatively small number of comments or sticky notes, as can be seen in Figure 20. That may be because the concept of a data model is rather technical. It also encloses a wide concept that has multiple entities and connections. They are not visible in normal daily work life to the people who work more in the business side rather than in the technical site, so the data models are not easy to comprehend.

The sticky notes in the step one were all analyzed and the terminology was modified according to them. Also, some new important terms were added according to the participants' comments. All analysis was done critically considering so that all the aspects were considered.

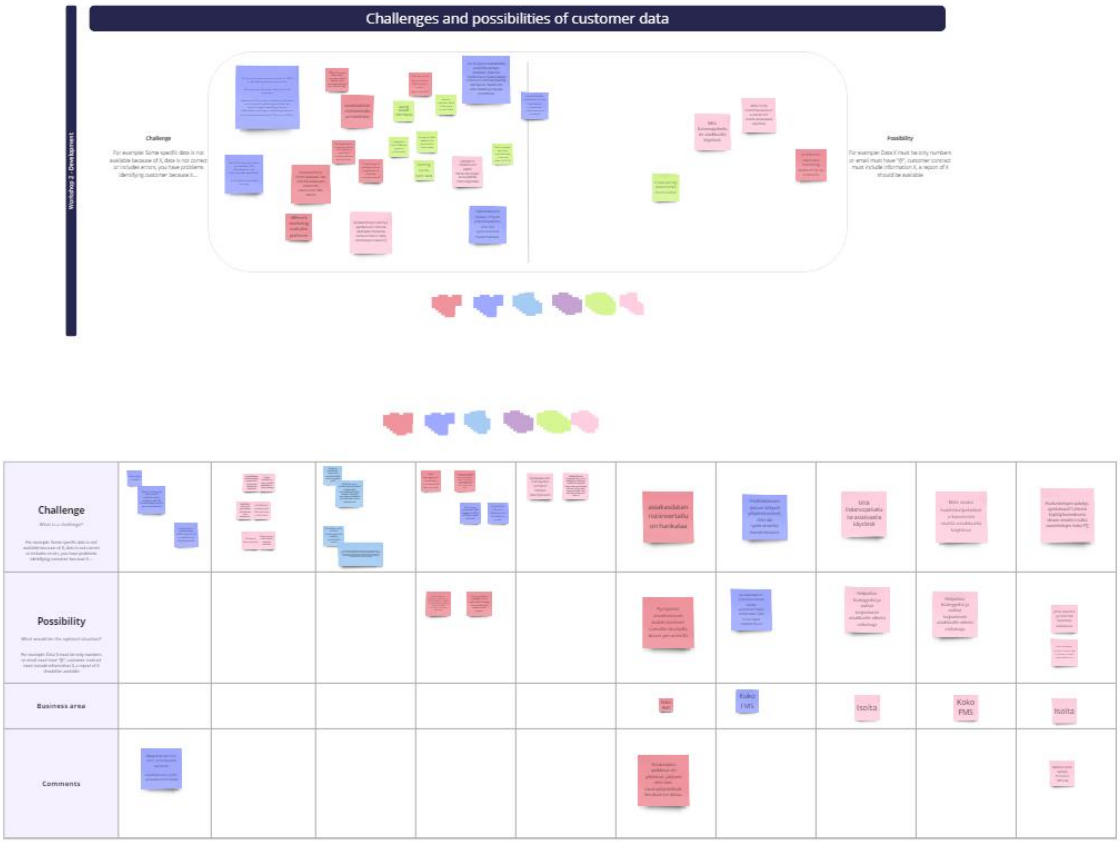


Figure 21. The overview of the outcome of the development workshop round's results of the determination of challenges and opportunities.

The participants identified many opportunities and challenges, which can be seen in Figure 21. After analysis and categorization, five main themes were risen. The first of them was an opportunity that the common customer data model and master data would enable coherent, up-to-date, and accessible customer data. It was pleasant to discover that the participant also recognized the opportunity and importance of having common master data. The second identified theme was an opportunity of having a common understanding of the services the customers of the case company use. This would be a needed thing, because now the case company has separate data for different legal entities in the case company, and being able to compare the different data sets is a complex process and even impossible in some cases.

The third theme was an identified opportunity to have one commonly agreed terminology in the case company. The different products, systems, and processes use different terminology, and in some cases, it is difficult to communicate coherently, which creates risks of misunderstandings. The fourth theme was the ownership and stewardship of customer data. Participants were thinking about who is responsible for the data and who can edit the data. These are important aspects to consider and determine. The last theme was data privacy and information security. There was a concern about privacy and security if so much customer data is in the same place, which is an important concern to act upon in future research. In addition, there are some customer data areas that can not be accessible to everyone, and they must be administered carefully.

The five themes identified as opportunities and challenges are important aspects that must be concerned when considering the quality aspects of master data and when planning the implementation and deployment of the designs that are done in this study. Taking care of that these themes are well considered and maintained, can already be stated that the model is providing good quality for the business.

After analysis of the second workshop round's results, the data model, master data, and terminology were developed with internal experts in the internal operations function in the case company. Especially data team and enterprise architect were part of the development. This was an important part of the design development because the internal experts were able to give more technical and conceptual input to the project.

The business side offered practical input for the project, but more technical and wider knowledge and information were also needed. The data team offered more technical support from data architecture and data engineering viewpoints, which were valuable when considering the final solution and also future work. The enterprise architect could offer information and knowledge from both technical and business points of view. By working as an enterprise-wide architect in the case company, the knowledge helped to unify the information that was gotten from all around the case company. With the data

team and enterprise architect the solutions could be elaborated to a concept, that would work both practically and technically.

In addition, the other cluster that had done a similar project with customer master data was invited to a discussion session. In the session information was shared and they showed what were their final customer master data points, and how their project progressed. the customer master data was compared to the draft that was formed in this study.

The data model then was agreed to concern both the customers and partners. That way the case company can manage their customer data, but at the same time manage the partner data, and have a clear knowledge about the different relationships that different organizations have with the case company. This also enables to have only unique representations of an organization, that can be either a customer or partner. Loshin (2010a) also highlights the importance of having unique representations of real world entities, and suggests using roles to enable differentiation of entity roles. The designed data model can be seen in Figure 22.

The case company can have customers who can also be partners. A partner can use the case company's solutions, which makes them a customer, but at the same time bring their customers to use the case company's solutions with a partner contract. In other words, an organization that the case company has some kind of contract with, can also have different relationships or contracts with different product lines or legal entities inside the case company. Also, there can be multiple contact persons for an organization, and the contacts might be related to different relationships in the organization. That is why the organizations the case company is dealing with will be named in the data model as an Account.

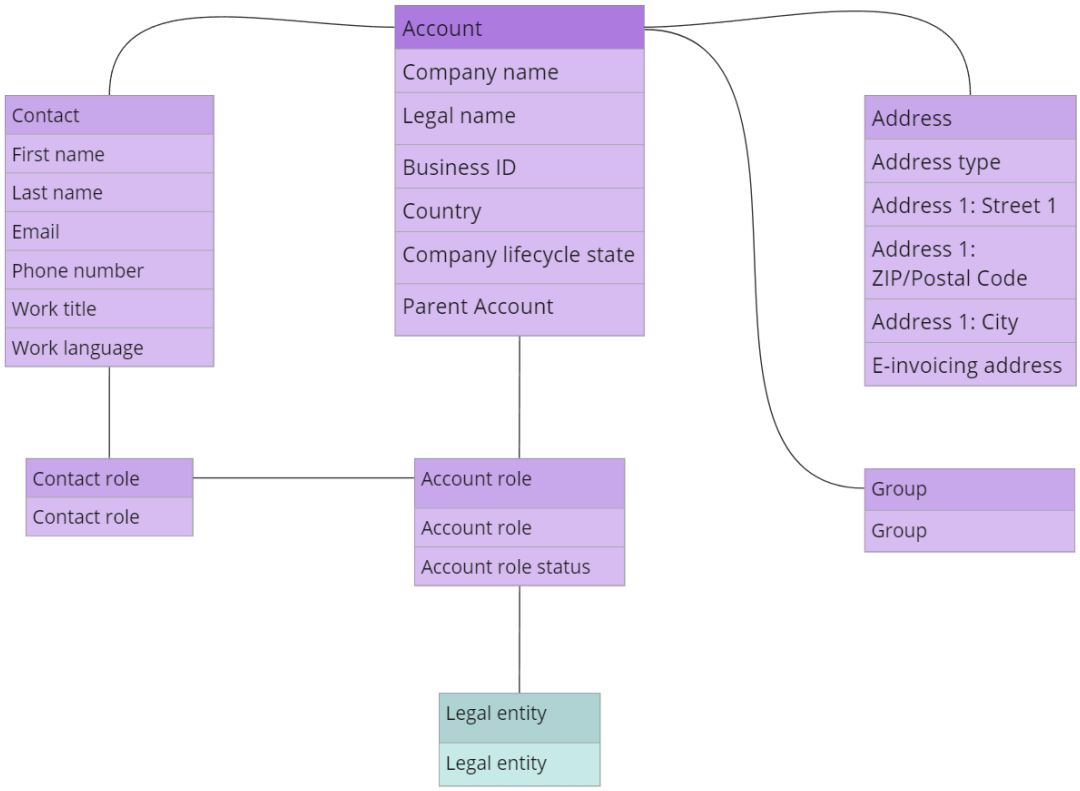


Figure 22. The designed data model.

In the data model, there are entities/tables of account, account role, contact, contact role, address, group, and in addition legal entity. Every table has its attributes, and the tables are related to others in different ways. The lilac color represents the information of the customer, and the turquoise table is information about the case company. Even if the data model is about customer data, the one turquoise table was included in the model, because it was important information, that was related to the role of the organization and the relationship between the case company and the customer or partner organization.

The organization would be referred to as an Account, and the company's relationship to the case company would be stated in an Account role table. This enables representation of an organization, that can be a customer or a partner, or both. Also, because the case company has different product lines, that have previously been individual companies,

there would be a table for the Legal entities in the case company, that table would be connected to the Account role -table, and it would elaborate that with which legal entity the account has a relationship with.

The account is for the organization's main information, company name, legal name, business ID, country, company lifecycle state, and parent account. The company name is the name that the organization gives to the case company, but the legal name is the name that is in the legal documents of the organization. Business ID is the identity code of the company, that has been given by authorities, and the format varies between countries. Country information of the organization can affect for example the service language or currency of invoicing, so that was seen as an important common data point. Company lifecycle state tells if the organization is in an active state, or if not active anymore. The parent account is related to the hierarchy of organizations. Designing and building the hierarchical system is a project for the future. The hierarchy might affect for example service or invoicing processes.

Account role is related to an account and it has attributes of role and account role status. The role tells if the organization is a customer or a partner for the case company, and there might be more than one account role for one account. The account role status tells if the role is active or passive. That indicates if the organization is a prospect, lead, customer, or former customer related to a legal entity of the case company. The legal entity that is related to the organization is in the legal entity table. The legal entities are the different product lines in the case company.

The address table is its own table because an organization might have multiple addresses. The address table has attributes of different address information: street, postal code, city, and e-invoicing address. The entity also has an attribute or address type, which indicates if the address is for example the organization's office address, postal address, or e-invoicing address.

The contact table includes the attributes of an organization's contact information, and there can be more than one contact for an account. The attributes are first name, last name, email, phone number, work title, and work language. The name, email, and phone number information are necessary information to be able to contact the organization. The work title of the contact is information that tells if the customer is for example CEO, accountant, or user, and which affects the situations and matters that the contact is contacted. The work language of the contact is necessary information so that the contact can be served with the right language. Every contact is related to at least one contact role entity, which tells what is the relationship of that customer related to certain account role.

The group entity includes different groups, where the organizations can be classified. For example, industry classification identifies in which industry the organization operates. Turnover class tells about the scale of the organization, which can be interesting information about the company for example from the point of view of financial function or marketing. The group table also enables that different groups can be added in case of a need for that information to be available for all inside the case company.

The final customer master data is presented below as a list:

Customer master data

1. Account name (company name)
2. Legal name
3. Business ID
4. Account role
5. Account role status
6. Parent account
7. Company lifecycle state
8. Country
9. Address(es) (Street, e-invoicing, etc.)

10. Contact first name
11. Contact last name
12. Contact email
13. Contact phone number
14. Contact role
15. Contact's work title

There are both account name and legal name presented because they can be different, and both can be needed. Business ID is the identification number on authoritative sources, and it can be used to identify the organization. The names and business IDs are slowly changing data values and might not ever change. They are also all important data values of a customer or partner, and therefore master data. Account role and the status of it tells what kind of a relationship this organization has with the case company, and if it is active or not. Parent account tells if the organization is hierarchically under some other organization. Company lifecycle status informs if the organization is in operation or if it does not exist anymore, which might affect marketing or sales processes.

Country and address information were seen as important company-wide information. Country information indicates for example the service language or currency of invoicing. Contact information of the organization is important because any customer or partner must be able to be contacted in some way. Contact information can change, but it is considered as slowly changing information, and eligible and necessary as master data values. There are many data values of customer information, but every one of them was considered important for different reasons and operations, but still company widely impacting values.

4.3 Review

Phase three was about informing the stakeholders about the solutions, that were designed based on the workshops. Everyone that participated in the workshops, was invited to the review event. The review event was also held in Microsoft Teams and Miro board was used to work together.

The third phase was a review event and not anymore a workshop, because the aim was only to review the plan and gather comments if there would be a need for some minor adjustments. During the event the outcomes were presented and explained with examples. At the end of the event was a possibility to discuss, question, or comment on the designs, and for that was also created a Miro board. That enabled people to comment with low threshold and also allowed participants to write their comments even when there was a conversation going on about another case.

The concept was introduced by determining customer data in this project and explaining the concepts of a data model and master data. Then the purpose of the project and a recap of the project steps were gone through. Also, the participants were reminded of the purpose of the review meeting, which was to present the designed customer data model and master data with examples, and to comment on and discuss about the designs with the participants, which were stakeholders and participants of the workshops. Lastly, before the presentation of the designs, the next steps were communicated.

The customer data model was presented and explained to the participants. They were informed of the entities and attributes of the data model and the relationships between them. After that, the model was used to visualize the master data, which is the attributes in the model excluding the grouping information and the legal entity. The concepts were elaborated with examples of imaginary organizations. The examples that were shown were about a company, that was simultaneously a partner and a customer. The organization then had the same account, group, and address information, but other information was different in the examples.

After the presentation, the participants were offered a possibility to discuss about the topic and give comments. In addition to the spoken conversation, a Miro board link was shared, and participants could use sticky notes to give comments virtually. During the discussion, also the comments on Miro board were gone through. Many comments took up a specific situation and asked how the model could work in those situations. Also, there were comments and questions about things that were scoped out from this project, like invoicing or contract related. There were no questions or comments that would demand changes to the designed data model or master data, so no adjustments were made. There were also questions about the project's next steps, which are planning and executing the implementation. Those questions concerned about things that were not yet determined and the schedule of the implementation phase, which was not yet determined. One subject of a longer conversation was the quality of data: how the model will consider it and what possibilities for improving customer data quality the new designs offer after the implementation. All questions were answered, and there were many good discussions.

4.4 Data Quality

The quality of the customer data is examined in this study by literature review, which is interconnected to the case company’s situation, and enriched with the things that were recognized during the empirical study. In the literature review of this study, both data quality and more especially master data quality and their special were studied. In this section of the study, the concentration is on the designs that were done in this study and the scope of those.

According to the literature review, one of the most important things in data quality is data quality management. There are multiple factors that influence data quality. Based on the study it could be stated that data quality is affected the most from data governance and data management. Therefore, a data quality program could be beneficial and effective way to ensure high data quality in the case company. A data quality program is

a systematic approach, which aims to ensure good data quality. It is not a project, but it is a continuously ongoing set of principles and policies, which aim to improve existing processes, prevent quality issues, and preserve high-quality data. Data quality and therefore data quality program work as part of data governance, like can be seen in Figure 1. Data quality program supports data governance and data governance gives data quality guidelines and frameworks to follow.

A data quality program, like data quality, is best when it is fit for purpose. For the case company, the data quality program should concentrate on five identified areas. The five chosen areas were identified from the literature review of data quality and considered in the environment of the case company. The five areas are:

1. Data lifecycle management
2. Focusing on critical data (e.g. master data)
3. Preventing issues, finding the root causes of issues
4. Develop metrics to monitor and control data quality, and identify development opportunities or demands
5. Ownership of data and data processes

Data lifecycle management is an important area, which includes the processes of data entry, data managing and processing, and lastly data disposal. Data lifecycle management from the point of view of data quality includes both what happens in business processes and in technical processes in systems. It is important that data structures, design, and processes are working correctly in systems, but it is also highly important that business processes, that create or use data are aware of the high data quality principles and are working according to them.

A data quality program should prioritize critical data, as was also found in the literature review. This study identified the customer master data, and for the case company, it is critical that the master data is high-quality. The high quality of customer master data

affects significantly customer relationship management and its operability. A master data management framework could be part of the data quality program because master data management concentrates on the quality improvement of the most critical data, master data.

An issue-oriented focus area of the data management program should focus both on fixing and preventing the issues. Some issues need concentration and fixing, like redundancy of data values. There should be planned a data quality assessment, which would give insights into the state of the data quality in the case company. As was discovered in the literature review, manual issue fixing can cause even more issues, so it would be beneficial to work also on preventing issues. That includes improving the existing systems by identifying best practices from external sources, like literature, but also reacting to the demands of the business side.

One critical area of the data quality program is developing metrics and policies. Metrics implicate how the data quality can be measured, and policies govern the procedures. Metrics should be defined together with the data team and business so that high-quality data can be defined and furthermore analyzed. For example, while building a physical customer data model based on this study's outcomes, the data quality aspect should be taken into account. Different policies, that should be developed would be for example how to enter data in a business process that includes data entries, or how to act if someone notices an inconsistency in data while processing it.

Ownership and responsibilities of data and the data processes are necessary things to consider because it is not always straightforward question in an organization. Many employees create, use, and process data, but who is truly the owner of the data? As was found out in the literature review, (Henderson et al. (2017, p. 453) state that business process owners should have data ownership, and the system owners should have stew-

ardship of the data and should have the responsibility of ensuring data quality with technology. There should be clear communication about that in the case company, aligned with proper training of what the ownership of data really means in practice.

4.5 Discussion

When reflecting on the organizing of the workshops, the workshops' rounds 1 and 2 would have been organized as two bigger workshops compared to the three and two workshops that were organized. Every workshop was booked for 1,5 hours, but none of them took the whole time. In all the workshops the cause of the shorter needed time was the lack of discussions and conversation. A bigger group of people might have generated more discussions, but of course, it would not have been promised. A bigger group of people is also more challenging to get to participate at the same time, which was one of the original reasons why the workshops were chosen to divide into smaller workshops.

One problem with the workshops was that some people that were invited to the workshops did not show up, which complicated the analysis process. There had to be some individual discussions, and it lengthened the analyzing process. Fortunately, there was scheduled enough time between the workshop rounds, so the project did not have delays. Some people also had vacations or overlapping meetings, which caused them to be unable to join or participate only in some sections of the workshops. This is of course understandable and probable to happen in a company environment.

There could have been more discussion in the workshops. If the workshops were face-to-face, that could have generated more discussions more easily. On the other hand, in today's hybrid working life, there would have been at least some participants online, so that could have made managing the workshop more difficult and even unequal for some participants. Otherwise, a better and more activating activity at the beginning of the workshops could have lowered the bar to open the mic and start a conversation, ask more questions, or raise some concerns. On the other hand, that could have taken more

time from the actual workshop subject. Also, many of the participants are busy with their other daily work, and the motivation could have lowered if they had to participate in some activity, that did not seem so important.

As a reflection, a few things could have been done differently during the second workshop round. The second workshop could have had a concrete list of potential master data already. There was potential master data marked in the data model, but because the data model is a technical and complex presentation, that might have affected the lack of comments. That is how it could have gotten more feedback. Also, the data model could have been gone through more detail and with examples, so that the participants could have had a clearer picture of how it could work in real life. Now the concept might not have been clear enough for the bigger audience.

One challenge in this project was that it was very important to involve employees from the business side of the case company in this project, but the concepts were rather technical. This problem was tried to solve by explaining with non-technical language the concept and targets of the project. Also, participants were asked multiple times during the workshops if they have anything that remained unclear, so everything could be re-explained in case of any uncertainty. Still, the concept of data models and master data is not an easy concept to comprehend. That caused some drawbacks to the process because the iterative process of developing the data model did not get so many improvement ideas.

Overall, workshops were a good way to gather knowledge and information from the company. The workshops gathered people from all the functions and product lines efficiently. If the data and information were collected in interviews, that could have made the data-gathering phase multiple times longer. A workshop is working together around some common subject and they were an effective way to collect data. Also, the usage of Miro application was an effective and working way to collect, document, and analyze data.

It was interesting to have conversations with people across the case company within the same subject. The conversations were helpful addition in addition to the materials that were gotten from the workshops.

The aim was that this study could be used as a base for future research in the case company, for example when considering other areas of data, like product data. Therefore this study structure was designed to be scalable for similar master data projects. Starting a new project based on this research requires an examination of relevant processes and teams, but the basic principles and workshop frame could be used in many other cases of defining data models and master data. This study also creates a base for implementing the created model.

4.6 Summary

This study had three research questions. The research questions aimed to define a customer and find out a common customer data model for a case company, identify customer master data in the new data model, and find out how high data quality could be ensured. The study was empirical, and the data collection was done in virtual workshops with relevant stakeholders from the case company. Stakeholders that were involved in the workshops were chosen with the data team lead, and they were chosen based on if they were working with customer data in their daily work.

There were three phases in the empirical study that involved stakeholders. The first two phases were workshops with a concentration on the analysis of the existing and the development of the new. The two workshop rounds consisted of altogether five workshops, three for the first round and two for the second. The workshops were divided into smaller workshops because the number of participants was relatively high, and it was wanted to ensure that there would be enough time for discussion of the topics. The last phase of the study which involved business stakeholders, was a review where the development was communicated, reviewed, and commented.

The first workshop was held in three phases, and the focus points were terminology around customer and customer data points. Based on the first workshop first drafts of the customer data model, customer master data, and customer terminology were made.

The second workshop concentrated on the iterative development of the customer data model, customer master data, and terminology and defining possible opportunities and challenges. Opportunities and challenges were meant to identify possible or existing problems and identify the possibilities of future plans. The opportunities and challenges are also tightly related to quality issues, that must be focused on in the later phases. After the second phase, the customer data model and master data were developed with the case company's data team and enterprise architecture.

In addition, a framework for ensuring customer data quality was created. The quality framework was based on theoretical analysis, and it also aimed to cover the issues that were risen in the workshops and conversations.

In the last phase of the project a review event was organized for participants of the workshops, so the outcomes could be communicated and reviewed. Also, the participants were offered the possibility to comment on the new customer data model and customer master data. There were good questions that were answered and insightful conversations. There was reserved a possibility of adjusting the designs after the review, but because there were not any needs for changes risen, no adjustments were made.

5 Conclusions

Figure 7 was presented a design science research method in visualization adapted by vom Brocke et al. (2020), where can be seen that the needs of the environment, in this case the case company, and knowledge, in this case the literature review, are combined to the design. The design is iteratively built by building and evaluating. In this study, the literature review was done on topics of data modeling, master data, and data quality. The literature review was used as a base to understand the theories and to do the empirical research in the case company. Combining the theories and the outcomes of the empirical study could be answered the three research questions, which were:

1. What is customer data and how it should be modeled as a logical data model to support unified processes and operations in a multi-company environment?
2. What is the customer master data in the new data model?
3. How to ensure data quality in the new data model?

To answer the first question, the customer terminology and created and used data points were identified with the business, and iteratively the customer data model and terminology were defined and designed together with the internal experts of the case company. The customer master data, that the second question comprehends, was part of that process and was defined and designed with the customer data model. The last question about the data quality in the new data model was based on both literature and the issues that were collected during the workshops and conversations.

5.1 Managerial implications and future research

The next step after this thesis is to plan how to continue with the implementation of the findings of this study, and how to extend the project to other areas of data. Allen & Cervo

(2015, p. 34) state that customer and product data domains are the most important domains when considering the master data in a company. This study concentrated on customer master data, and the next domain to concentrate on would be product data.

This study was done keeping in mind that it could be scalable. It was certainly beneficial, because this project was instructive, and many things will be considered as part of a project of the next data area, but many things can also be developed and done more efficiently. This study was used as a practice from which a lot of things can be learned. For example, this study aimed to get a comprehensive understanding of the topic by involving many employees across the organization at the same time, but when starting a new project around product data, might smaller group conversations work better. That can be done after the quality aspects that this study pointed out are deployed, and the data ownerships are determined.

The implementation of the master data requires building a physical data model of it. A physical data model is a more technical approach, that includes more detailed information about the data points. Building the physical data model should be done by having in mind the quality aspects, that were identified in this study. Designing a physical data model also requires a decision, where the data model should locate. In addition, data privacy and information security should be considered carefully, because they were not part of this study, but they were identified as an important aspect.

Overall, this study was an interesting project, which gave a lot of new information for the case company. This study created a bas understanding of common customer data and enables the continuation of the project. This study also works as a framework for future projects of similar context.

References

Allen, M., & Cervo, D. (2015). *Multi-Domain Master Data Management: Advanced MDM and Data Governance in Practice*. Morgan Kaufmann.

Bahgi, E., Otto, B., & Oesterle, H. (2013). Controlling Customer Master Data Quality: Findings from a Case Study. *International Conference on Information Resources Management (Conf-IRM)*.

Baran, R. J., & Galka, R. J. (2016). *Customer Relationship Management: The Foundation of Contemporary Marketing Strategy* (2nd ed.). Routledge.

Berson, A., & Dubov, L. (2007). *Master Data Management and Customer Data Integration for a Global Enterprise*. McGraw Hill Professional.

Blaha, M. (2010). *Patterns of data Modeling* (1st ed.). Taylor & Francis Group.

Buttle, F. (2009). *Customer Relationship Management: Concepts and technologies* (2nd ed.). Elsevier.

Catalan-Matamoros, D. (Ed.). (2012). *Advances in Customer Relationship Management*. InTech. https://doi.org/10.5772/1795

Cervo, D., & Allen, M. (2011). *Master Data Management in Practice: Achieving True Customer MDM*.

Das, T. kumar, & Mishra, M. R. (2011). A Study on Challenges and Opportunities in Master Data Management. *International Journal of Database Management Systems (IJDMS)*, *3*(2).

Finkelstein, C. (2006). *Enterprise Architecture for Integration - Rapid Delivery Methods and Technologies*.

Gregory, A. (2011). Data governance — Protecting and unleashing the value of your customer data assets. *Journal of Direct, Data and Digital Marketing Practice*, *12*(3), 230–248. https://doi.org/10.1057/dddmp.2010.41

Haneem, F., Kama, N., & Ali, R. (2017, October). Risk Factors in Master Data Management Implementation. *PostGraduate Annual Research on Informatics Seminar 2016 At: Universiti Teknologi Malaysia, Kuala Lumpur*.

Haug, A., & Arlbjørn, J. (2011). Barriers to master data quality. *Journal of Enterprise Information Management*, *24*(3), 288–303. https://doi.org/10.1108/17410391111122862

Haug, A., Stentoft Arlbjørn, J., Zachariassen, F., & Schlichter, J. (2013). Master data quality barriers: an empirical investigation. *Industrial Management & Data Systems*, *113*(2), 234–249. https://doi.org/10.1108/02635571311303550

Helo, P., Tuomi, V., Kantola, J., & Sivula, A. (2019). *Quick guide for Industrial Management thesis works*.

Henderson, D., Earley, S., & Sebastian-Coleman, L. (Eds.). (2017). *DAMA-DMBOK* (2nd ed.). Technics Publications.

Hikmawati, S., Santosa, P. I., & Hidayah, I. (2021). Improving Data Quality and Data Governance Using Master Data Management: A Review. *IJITEE (International Journal of Information Technology and Electrical Engineering)*, *5*(3). https://doi.org/10.22146/ijitee.66307

Hunka, F., & Matula, J. (2016). *Conceptual and logical level of database modeling*. 120014. https://doi.org/10.1063/1.4951897

Ibrahim, A., Mohamed, I., Safie, N., & Satar, M. (2021). Factors Influencing Master Data Quality: A Systematic Review. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 12, Issue 2). www.ijacsa.thesai.org

Kokemüller, J., & Weisbecker, A. (2009). *Master Data Management: Products and Research*. 8–18.

Lee, Y. W., Funk, J. D., Pipino, L. L., & Wang, R. Y. (2006). *Journey to Data Quality*.

Loshin, D. (2001). *Enterprise Knowledge Management: The Data Quality Approach*.

Loshin, D. (2010a). *Considerations: Mastering Data Modeling for Master Data Domains*.

Loshin, D. (2010b). *Master Data Management*. Elsevier Science.

Miro. (2022, December 19). *How to help your team get started with Miro*. https://miro.com/blog/starting-remote-team-collaboration/

Otto, B., & Reichert, A. (2010). Organizing master data management. *Proceedings of the 2010 ACM Symposium on Applied Computing*, 106–110. https://doi.org/10.1145/1774088.1774111

Otto, B., & Schmidt, A. (2010, August 13). Enterprise master data architecture: Design decisions and options. *Enterprise Master Data Architecture: Design Decisions and Options*.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*, 45–77.

Saarijärvi, H., Karjaluoto, H., & Kuusela, H. (2013). Customer relationship management: the evolving role of customer data. *Marketing Intelligence & Planning*, *31*(6), 584–600. https://doi.org/10.1108/MIP-05-2012-0055

Samaranayake, P. (2008). Enhanced data models for master and transactional data in ERP systems – unitary structuring approach. *Lecture Notes in Engineering and Computer Science*.

Silvola, R., Jaaskelainen, O., Kropsu-Vehkapera, H., & Haapasalo, H. (2011). Managing one master data – challenges and preconditions. *Industrial Management & Data Systems*, *111*(1), 146–162. https://doi.org/10.1108/02635571111099776

Simonin, J., Bigaret, S., & Gourmelen, J. (2012). A data warehouse logical design method based on the alignment with business processes. *2012 Sixth International Conference on Research Challenges in Information Science (RCIS)*, 1–12. https://doi.org/10.1109/RCIS.2012.6240442

Simsion, G., & Witt, G. (2004). *Data Modeling Essentials* (3rd ed.). Elsevier.

Srikant, S. (2006). Logical Data Modeling: A Key to Successful Enterprise Data Warehouse Implementations. *DM Review*, *16*(9).

Storvang, P., Mortensen, B., & Clarke, A. H. (2018). Using Workshops in Business Research: A Framework to Diagnose, Plan, Facilitate and Analyze Workshops. In *Collaborative Research Design* (pp. 155–174). Springer Singapore. https://doi.org/10.1007/978-981-10-5008-4_7

Talburt, J. R., & Zhou, Y. (2015). ISO Data Quality Standards for Master Data. In *Entity Information Life Cycle for Big Data* (pp. 191–205). Elsevier. https://doi.org/10.1016/b978-0-12-800537-8.00011-9

Väre, T. (2019). *Master data* (1. edition). Alma Talent.

vom Brocke, J., Hevner, A., & Maedche, A. (2020). *Introduction to Design Science Research*. 1–13. https://doi.org/10.1007/978-3-030-46781-4_1

Witt, G. (2021). *Data Modeling for Quality* (1st ed.). Technics Publications.

Appendix 1 – Terminology and definitions

Unknown	For example, a website visitor who is not recognized yet.
Potential customer	A potential customer is an identified person or organization.
Lead	An organization that has shown interest towards our solutions.
MQL – Marketing Qualifies Lead	A lead that has shown interest or wants to test the software.
SQL – Sales Qualified Lead	A lead that marketing has handed to sales.
Prospect	A potential customer has not yet a contract with us but is identified and in the lead channel.
Opportunity	A prospect. Opportunity has four stages: Interest, Qualify & prove value, Negotiate, and Contract.
Deal	The opportunity is won, the contract is made, and the company is now a customer.
Lost Opportunity	The prospect was not interested after all.
Sign up customer	The customer has signed up or registered to be a customer. This path does not necessarily include the lead channel or prospect phase.
Active customer	Active customer who uses our solutions.
Passive customer	Has a contract but does not actively use the solutions.
Non-paying customer	Customer who uses the solutions for free.
Former customer	An organization that has had a contract with us. A former customer can stay as a non-paying customer in some cases. A former customer still might have some data stored with us.

Customer	A general term that means a registered/legal/juridical business organization that has a contract with us. One customer can have multiple customerships.
End customer	An organization that uses our solutions. The contract can be made either directly or via partners.
Direct customer	An organization that uses our solutions and we have a direct contract with, without a partner between.
Indirect customer	An organization that uses our solutions but has a contract via a partner.
Main customer	For example, a corporation that has many companies, that are sub-customers for us, for example, a parent company in a corporation.
Sub customer	Sub-customer is our customer via a main customer.
Test customer	Customer, who is testing our product.
Invoiced customer	Paying facet.
User	Has an account to our software. Can be a direct customer, via partner (no contact permission), or a partner.
Account	Account of a customer to our software.
Admin/Main user	An end user that is a decision-maker in the customer organization and has the authorization to govern and make changes to the subscription.
Test user	A customer that tests our products.
VAS (value added services) & Features user	Customers that also uses value added services and/or features.

Customership	Customership is a contract or many contracts between us and the customer about the usage of services. Same business ID i.e. same customer can have multiple company ID´s.
Contract	A contract about the usage of services between the case company and an organization.
Business ID	A unique identifier of the customer organization.
Industry (TOL)	A standard industrial classification number that tells in which industry the company operates.
Contact	Contact of the customer.
Decision maker	Someone in the customer organization that makes the decisions.
Legitimated for signing	A person from the customer organization that is legitimated to sign in the name of the company.