Universiteit Leiden

ICT in Business and the Public Sector

A data governance framework from the perspective of privacy and security in an Artificial Intelligence-integrated CRM system

Name: Johnny Hu
Student-no: s2713063

Date: 13/04/2022

1st supervisor: Prof. Dr. Werner Heijstek
2nd supervisor: Prof. Dr. Ir. Joost Visser

Master Thesis

Leiden Institute of advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

Data Governance Framework

Privacy and Security in an AI-integrated CRM system

Johnny Hu
S2713063
MSc. Thesis

Universiteit Leiden

prex partners

Supervisors
Prof. Dr. Werner Heijstek
Prof. Dr. Ir. Joost Visser

13/04/2022
Word Count : 30955
Page Count : 93

Abstract

Background

An AI-integrated CRM system is becoming more common in industries seeking to maximise customer satisfaction and retention. Many benefits and risks come with the growth of adopting AI and Big Data systems. Research has shown that such systems can be abused and misused, exposing organisations to legal and ethical issues. People are becoming more aware of privacy and security when handling, storing and using their data. General data governance frameworks exist, but not for privacy and security in an AI-integrated CRM system. Thus, the need for such a framework is growing, especially as AI integrated systems keep developing in complexity.

Objective

This research aims to develop a data governance framework from the perspective of privacy and security in an AI-integrated CRM system. The developed data governance framework is focused on a general application for each organisation using, developing or implementing an AI-integrated CRM system. The effectiveness of the data governance framework will be tested on (RQ1) how such framework can be used in the privacy and security aspect of an AI-integrated CRM system and (RQ2) how such framework impacts the privacy and security aspect of an AI-integrated CRM system.

Method

A design science approach is used to conduct this research, as similar topics in data governance have followed the same methodology. The literature review will form the basis of the framework, where a prototype has been made. Refinements are made to the framework based on an adapted version of the data governance maturity assessment tool, comparing the current situation at PrexPartners and the proposed framework. A final assessment is made over the final version of the data governance framework that considers the refinements.

Results

The results from the survey towards all the dimensions from the maturity assessment tool resulted in an average score of 1,99 for the current situation in the organisation and an improved average score of 4,79 with the proposed data governance framework. Specifically for privacy and security, the current situation resulted in an average score of 3,27 and an improved average score of 4,86 with the proposed framework. Additional feedback from the respondents indicated that the proposed framework was thorough and easy to use.

Conclusion

This research concludes that (RQ1) the data governance framework can be used in an organisation to improve and maintain the privacy and security of an AI-integrated CRM system. The provided scenarios (RQ2), through refinement from the gathered results in privacy and security, have a direct impact when people are accessing the customer's data or when the deletion of the data of a customer is requested. Although, limitations apply as results can differ on how an AI-integrated CRM system is developed and implemented in an organisation.

Table of Contents

Listed Figures

Listed Tables

Abbreviations

- ABAC – Attribute-Based Access Control
- ACP – Access Control Policies
- AI – Artificial Intelligence
- API – Application Programming Interface
- ANN – Artificial Neural Network
- BAC – Basic Access Control
- BD – Big Data
- BDA – Big Data Analytics
- BDAS – Big Data Algorithmic System
- BI – Business Intelligence
- BOLD – Big, Open and Linked Data
- BRAC – Base Realignment And Closure
- CCM – Cloud Control Matrix
- CDO – Chief Data Officer
- CIO – Chief Information Officer
- CMM – Capability Maturity Model
- CPO – Chief Privacy officer
- CRM – Customer Relationship Management
- CSA – Cloud Security Alliance
- CSO – Chief Security Officer
- CSP – Cloud Service Provider
- DAC – Discretionary Access Control
- DPA – Data Protection Authorities
- E-commerce – Electronics commerce
- ENISA – European Union Agency For Cybersecurity
- ERP – Enterprise Resource Planning
- ETL – Extraction Transformation Load
- FISMA – Federal Information Security Modernization Act
- FOI – Freedom of Information
- GDPR – General Data Protection Regulation
- I-commerce – Internet-enabled commerce
- IaaS – Infrastructure as a Service
- ISACA – Information Systems Audit and Control Association
- ISAE – International Standard for Assurance Engagements
- ISO – International Organisation for Standardisation
- M-commerce – Mobile commerce
- MAC – Media Access Control
- ML – Machine Learning
- NDA – Non-Disclosure Agreement
- OAuth – Open Authentication
- PaaS – Platform as a Service
- ROI – Return on Investment

- RFM – Recency, Frequency and Monetary
- SAML – Security Assertion Markup Language
- SAS – Statistical Analysis System
- SCM – Supply Chain Management
- SDI – Scientific Data Infrastructure
- SKU – Stock Keeping Unit
- SLA – Service Level Agreement
- SOC – Service Organisation control
- SSH – Secure Shell
- SSI – Self Sovereign Identity
- SSO – Single Sign On
- SQL – Single Query Language
- U-commerce – Ubiquitous commerce
- UI – User Interface

1. Introduction

This chapter takes into account a brief background that is connected with the motivation of the chosen topic. Two research questions are derived to obtain concise answers to the conducted study. The study's objective is stated, followed by an overview of the thesis structure.

1.1. Background & motivation

A CRM system is regarded as one of the most effective tools various industries use to identify their best customers and maximise their satisfaction value (Chatterjee et al., 2021). A CRM system is a systematic tool that improves organisational performance and customer satisfaction (Nguyen, 2012; Coltman et al., 2011). A CRM system requires a large amount of data to be accurately analysed concerning its customers (Keramati et al., 2010). This data consists of a significant amount of personal information and activities, also known as *Big Data*. Personal data is frequently sensitive and can pose a severe threat to security and privacy (Mansour, 2016). The use of manual techniques to analyse this data has become difficult, necessitating technological advances such as the Artificial Intelligence (AI) application in CRM, dubbed AI-integrated CRM (Molinillo & Japutra, 2017). As a result of the increasing use of AI-integrated CRM systems in organisations, businesses can analyse data with less manual effort and interact with customers on a larger scale (Libai et al., 2020). AI enables organisations to adapt to personalised services at a low cost in the long run. Human-like interactions enabled by AI-powered systems will replace manual efforts and reshape customer service (Kaplan & Haenlein, 2019; Hoyer et al., 2020). The significance of using AI in a CRM system can be seen in ***Figure 1***: Impact of AI on a CRM from Tahir & Khan (2021)**.**

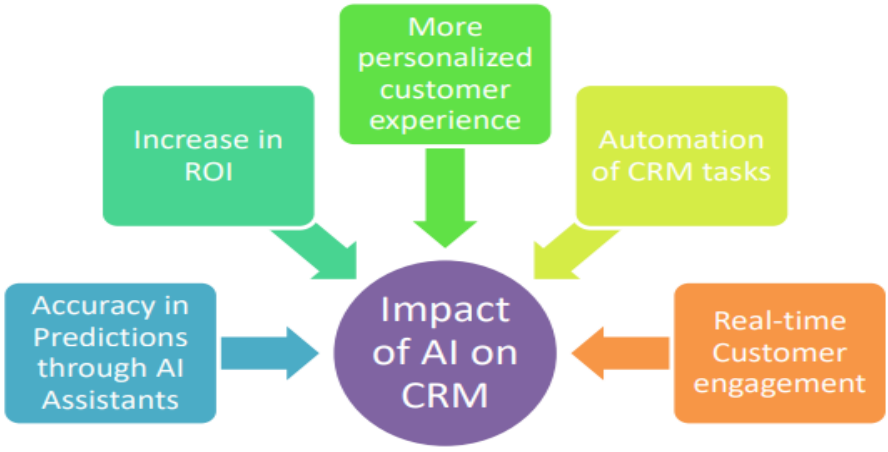


Figure 1: Impact of AI on a CRM from Tahir & Khan (2021)

Despite the potential of AI-integrated CRM systems, it remains unexplored in academia (Youn & Jin, 2021; Chaterjee et al., 2020). Only a few studies have covered the adoption of AI-integrated CRM systems, with a primary focus on the privacy and security aspects of data governance (Chatterjee et al., 2020). As the demand increases for AI systems to make decisions that have consequences for individuals and communities, particularly industries, failures must be eliminated to meet regulatory and ethical requirements (Janssen et al., 2020). The potential for information misuse and abuse has already been discovered concerning the specific form of big data about AI (Mansour, 2016). As a result, a high-risk situation has arisen, involving the

legal and ethical aspects of data usage and highlighting a broken system (Rana et al., 2021). Data governance can be considered the foundation of a reliable and usable AI (Janssen et al., 2020). Without a data governance framework governing the security and privacy of AI-integrated CRM systems, businesses will make less optimal business decisions and have legal and ethical implications (Yang et al., 2019). As a result, there is a perceived risk of operational inefficiency and competitive disadvantages, with the worst-case scenario of an organisation's demise (Rana et al., 2021).

To properly improve and maintain data quality, organisations should implement a data governance framework that includes the processes, policies, practises, and structures required to orchestrate their people, processes, and technologies and optimise data collection, storage, use, and dissemination (Soares, 2010, 2012; Marchildon et al., 2018; Ladley, 2019). A data governance framework is concerned with creating policy to optimise, secure, and leverage information as an organisation asset by aligning the goals of multiple functions (Soares, 2012). This loss of governance and control may severely impact the organisational objectives and thus its ability to meet its mission and goals (Marchildon et al., 2018). The loss of control and governance can also result in the inability to meet security requirements, a lack of data confidentiality, integrity, and availability, and a deterioration in performance and quality of service, not to mention the introduction of compliance challenges (Al-Ruithe et al., 2016). Hence, an approach toward mitigation, planning, monitoring and control needs to be developed through a data governance framework for an organisation that handles valuable data (Khatri & Brown, 2010; Janssen et al., 2020).

Implementing a traditional CRM system is prone to risks and failures, which only emphasises the need for further research into AI-integrated CRM systems in general, which is better suited to the current task at hand in the current business environment (Bibiano et al., 2014). Starting a business is typically fraught with financial constraints, as privacy and security are usually near the bottom of the priority list (Mogaji et al., 2020). Although Big Data is not the only requirement for an AI-integrated CRM system to function, trusted information sharing frameworks, processes, and algorithms must be considered in order to have a secure and trustworthy AI (Janssen et al., 2020). While topics such as AI governance and ethics have been discussed philosophically, societally, theoretically, and legally, there is a lack of work pertaining to businesses and corporations (Schneider et al., 2020). As a result, this field must continue to advance from the standpoint of privacy and security within data governance to establish a fundamental foundation in regular business practices and theoretical implications (Janssen et al., 2020).

Technologies such as AI systems and Big Data have changed how businesses operate and are a disruptor in each industry, becoming more complex throughout their development (Verma et al., 2021). With the growth of AI systems, such as the AI-integrated CRM system, data governance remains the critical element in the further development and trust in the society (Yang et al., 2019). To prevent situations like the Facebook-Cambridge Analytica data scandal, where data was illegally consented to third parties and sets a precedent in data safety and privacy (Mikalef et al., 2019).

1.2. Research Questions

Based on *1.1Background & motivation*, the following research questions to be answered are:

RQ1. How can a data governance framework be used for the security and privacy aspect of an AI-integrated CRM system?

RQ2. How does a data governance framework impact the security and privacy aspect of an AI-integrated CRM system?

1.3. Research Objective

This research aims to enable a discussion and further development of data governance frameworks in security and privacy in the context of an AI-driven data-handling system. Specifically in business industries that use or are considering to implement an AI-integrated CRM system. There is currently little research available that specifically relates to how the security and privacy aspects, as mentioned, are taken into account. Hence, the aim is to contribute by creating a data governance framework that allows for increased resistance to security risks and privacy concerns in AI-integrated CRM systems. When such a system is implemented in an organisation, the data governance framework can contribute in:

- The legal and ethical aspects of using data through an AI-integrated CRM system.
- Standardisation procedures in data governance for an AI-integrated CRM system.
- The mitigation and prevention of data risks when using an AI-integrated CRM system.
- An organisation approach to data governance in an AI-integrated CRM system.
- The trustworthiness of Big Data used in an AI-integrated CRM system.
- A system-level data governance approach towards an AI-integrated CRM system.
- Data governance principles that apply to an AI-integrated CRM system.
- Specific use-case scenarios from the perspective of privacy and security in an AI-integrated CRM system.

The development strategy of the data governance framework is intended to be applied to all AI-integrated CRM systems in general and to any industry that uses or adopts such system.

1.4. Thesis Overview

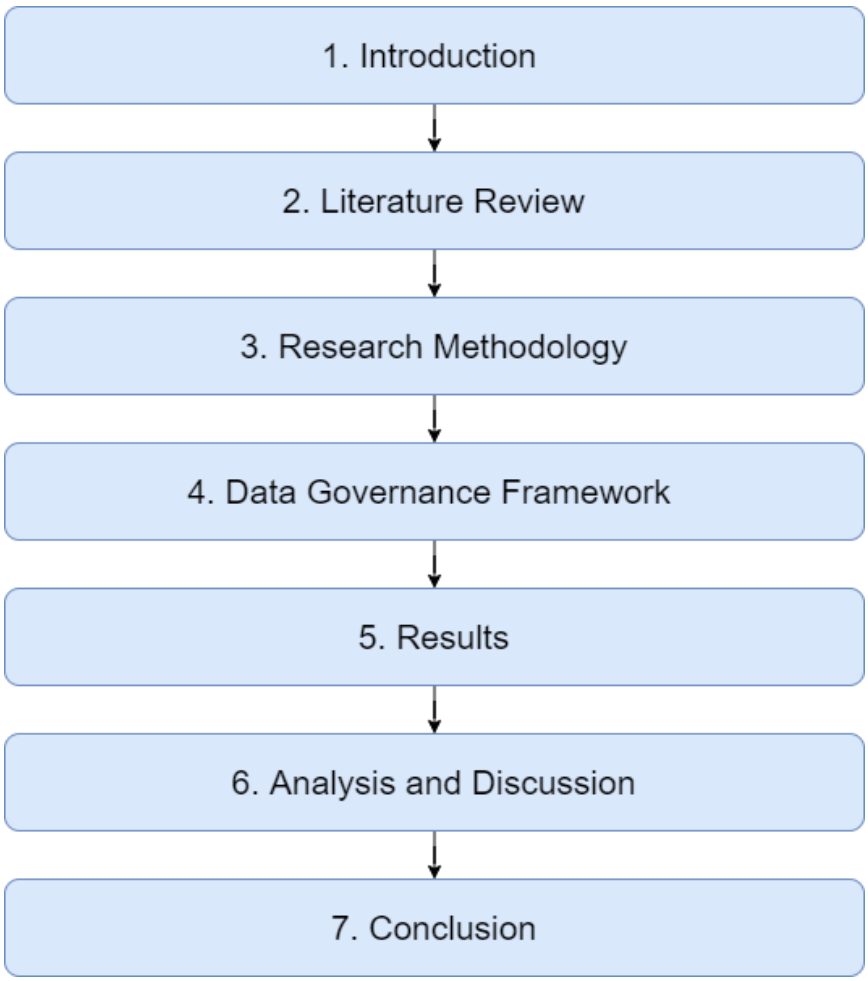


Figure 2: Thesis Overview

The thesis is organised as follows, beginning with the current chapter, which introduces the topic. The second chapter provides a literature review of the foundation of a traditional CRM and an AI-integrated CRM, followed by data governance practice approaches and frameworks. Chapter 3 describes how the research will be carried out through a design science approach. Chapter 4 discusses the development process at PrexPartners and the data governance framework. Chapter 5 describes the results of the surveys conducted, while Chapter 6 focuses on analysing the survey results and improving the data governance framework based on the gathered information. In Chapter 7, a conclusion will be drawn based on the results gathered and the improvements made due to the given feedback from the survey. Suggestions for future research, as well as potential limitations, will be discussed.

2. Literature review

This chapter provides a high-level overview of the topics related to an AI-integrated CRM system. First, consider a traditional CRM system, including its functions and components. AI integration within a CRM system will be thoroughly discussed, as will the algorithms used, their infrastructure, etc. Big Data (BD) is an essential factor that will be referred to collectively as "data." Finally, data governance practices, approaches, standards, and policies relevant to an AI-driven system that handles BD in today's business and academic landscape will be discussed.

2.1. Regular CRM

A CRM is an information system that tracks interactions between the organisation that uses the tool and its customers, allowing system users to see all relevant information, such as their records, past sales, problem calls, etc. (Nguyen et al., 2012). A CRM system is most commonly used in industries where organisations can maximise customer satisfaction and retention (Kennedy, 2006; Chatterjee et al., 2021). There are several definitions of what a CRM is, but the significance lies in its applications. The following are some of the CRM system applications (Garrido-Moreno et al., 2010):

1. CRM is to create value for its customers through knowledge of the gathered information, where personalised products and/or services can be offered through their needs and preferences.
2. CRM can include using a broad range of technologies, such as AI algorithms.
3. CRM is concerned with the maintenance of long-term business relationships. With a secondary emphasis on attracting new customers.
4. A business that is using a CRM needs to be redesigned in such way to orient its customers, hence strategy requires a suitable leadership and the work environment culture.

A success model could be visualised using the CRM applications mentioned above (Garrido-Moreno et al., 2010).

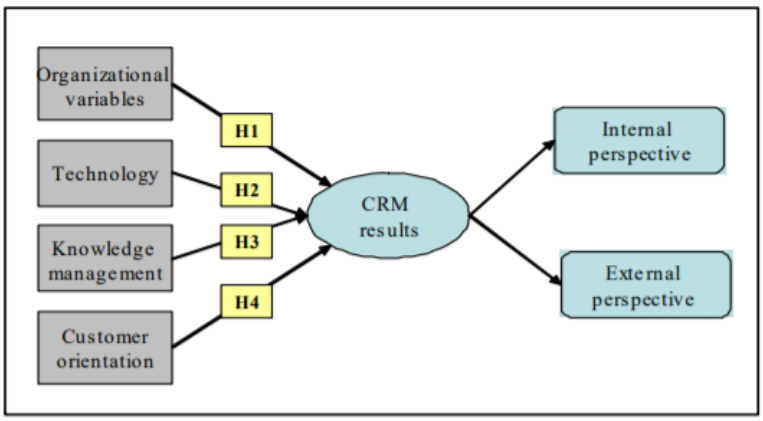


Figure 3: CRM implementation success model from Garrido-Moreno et al. (2010)

Because data is the most critical component of a functioning CRM, analysing and managing such knowledge becomes one of the most important aspects of a business (Li & Nguyen, 2016).

Because of the increasing amount of data, processing these massive data sets has become complex and frequently necessitates the assistance of technologically capable tools (Milinillo & Japutra, 2017). As a result, the terms "Big Data*"* and "Artificial Intelligence*"* have captured the public's attention and are shaping the economic, social, and political environments (Elish & Boyd, 2017). Understanding these terms can help with implementation and benefit a company's core strategy, as it is already redefining areas such as finance, logistics, e-commerce, etc. (Ng, 2017). It is predicted that AI will play a significant role in innovative marketing, business models, customer and sales service, etc. in the field of customer relations and AI (Davenport et al., 2019), including the ability to improve marketers' efficiency and effectiveness through automation, customer insights, prediction, etc. (Kardon, 2019). To begin and discuss an AI-integrated CRM in-depth with a focus on data security and privacy, will need to briefly outline the typical development plan, composed of the relevant parts and components of a regular CRM (Bose, 2002), which serves as the foundation for integrating AI.

2.1.1. Collecting Data

Because of the increase in information storage, collection, and inspection due to the transition to a digital civilisation, organisations focus on developing and maintaining databases containing large amounts of customer information (Zahay et al., 2012). However, research indicates that working professionals will not use such databases unless there is a guarantee that the data is of high quality and can be relied on to be correct (Payton & Zahay, 2003). Data integrity can only be achieved when an organisation employs a customer-focused strategic approach (Ewing, 2009). As a result, CRM has already begun to focus on maintaining and ensuring data consistency (Zahay et al., 2012). Maintaining high-quality data (Marsh, 2005) requires auditing, cleaning, and implementing compliance measures. Data collected for CRM use can be described as learning activities involving customer information, such as a segmented view based on customer data (Zahay et al., 2012). Which is derived from dynamic sources, such as primary data collections, and secondary sources such as web search results are derived from static sources (Liu & Wang, 2010). CRM systems can identify different types of data from customers, and the relationships can be shown in a pyramid form (Zayah et al., 2012):

Figure 4: Quality data type pyramid from Zayah et al. (2012).

- Because of its personal relationship with each customer, the top indicates the activities, intricate relations, and preferences of customers that are difficult to replicate.

- Customer touchpoints focus on tailoring marketing towards the customer through communication. Concepts such as preferences in marketing materials and communication channels are the highlighted data.
- Psycho-Demographic data consists of information, such as motives, beliefs, values, attitudes and lifestyle. This information can be gathered through internal forms, such as surveys or external sources, such as acquired prospect information.
- The transactional tile represents historical purchases, frequency and other data related to financial transactions, abbreviated in RFM (recency, Frequency and Monetary). These values form the core of management to both customer and shareholder relationships.
- The base represents the minimum amount of effort required to collect. The base provides a foundation for learning activities and collecting and using other data types. Representing fundamental data such as contact and prospect information, any data-driven approach would be impossible to develop without this data.

Organisations collecting data are becoming more vulnerable to insecure practices as they collect more information than before (Malhotra & Malhotra, 2010). Failures in unsafe data practices cannot be tolerated because of the negative consequences, such as the legal implications (Janssen et al., 2020).

2.1.2. Data Mining

Data mining is the process of gathering and recognising information in a database that can be used to gain insight about the customer using statistics, mathematics, AI, and machine learning techniques (Bahari & Sudheep, 2015). Data mining aims to collect the information required to provide adequate customer service in a CRM (Cooley et al., 1997). Prior to the implementation of AI, various techniques such as clustering, pattern discovery, classification, and customer value evaluation, among others, were used manually (Yong et al., 2003). Data mining has grown in popularity, gaining much traction in various CRM applications, and is now the most commonly used method (Bahari & Sudheep, 2015). The acquisition and retention of potential customers are critical in developing any CRM system. Maximising customer value via a CRM-data mining framework can be supported by various data mining models depending on the task at hand (EWT et al., 2005). Because most models take into account the prediction of customer behaviour, data mining is primarily used for forecasting and decision-making (EWT et al., 2009). Predictions can be made using predefined classes based on data classification from the database (Mi et al., 2002). The Naïve Bayes classifier is a simple probabilistic classifier that employs the Bayes theorem (Tom, 2010). The Multilayer Perception Neural Network is based on artificial intelligence classification via machine learning (Hany, 2014), which will be covered in greater detail later. These two classification models can be visually represented in the following framework (Bahara & Sudheep, 2015):

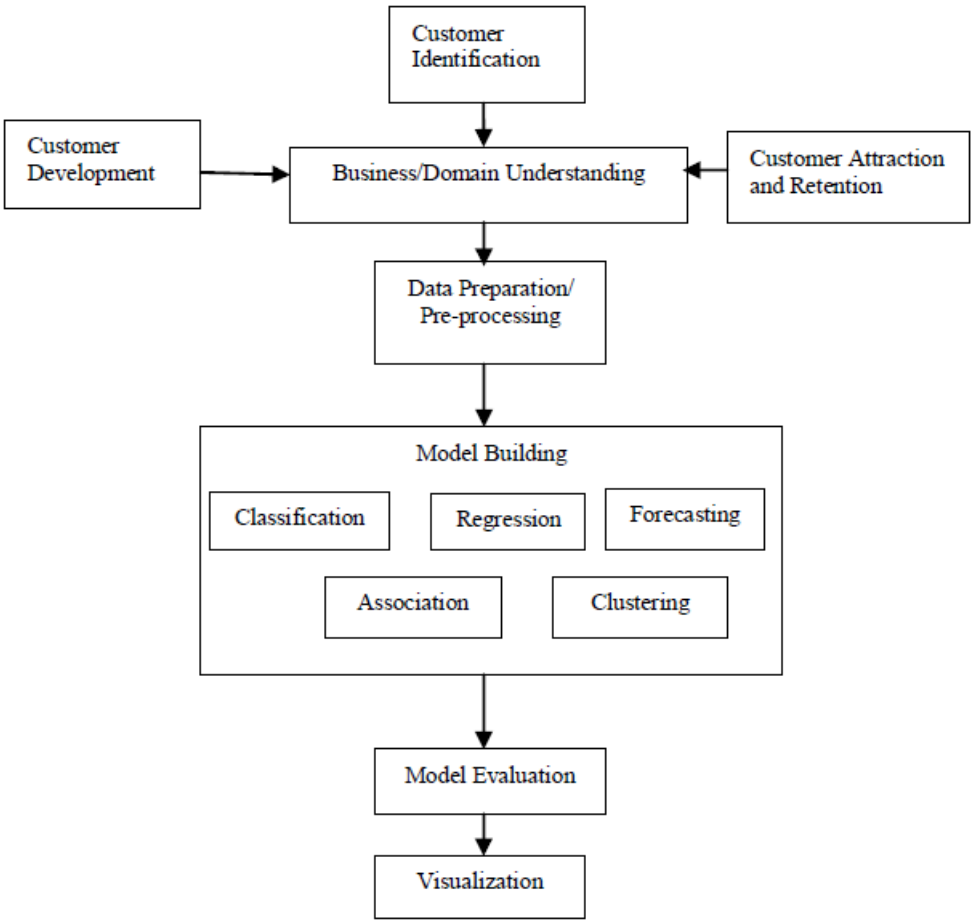


Figure 5: CRM-data mining framework from Bahara & Sudheep (2015)

Starting with the business goals and problem domain requirements that need to be understood in the first phase of data mining, observing the management of customer relationships and their activities helps develop, identify, and retain customers in its domain. Data preparation/pre-processing prepares the data through several processes, such as cleaning, selection, transformation, etc. These processes enable the data mining process to evaluate the model before visualising the data, enhancing the performance of predictions and decision-making by measuring its effectiveness (Bahara & Sudheep, 2015).

2.1.3. User interface

Aside from gathering and categorising valuable data, the UI (User Interface) design is one of the most important aspects of a CRM (Yong et al., 2003; Zahay et al., 2012). Understanding who the users of a CRM system are and what they want to achieve is a fundamental principle that must be followed (Cooley, 2002). Hence, the following design elements must be considered in order to create a compelling UI for a CRM system (Yong et al., 2003):

- Results of data analysis
- Decision making support in real-time
- Data mining integration
- Information process management

Because the primary goal of a CRM system varies per organisation and can include service personalisation, information can be displayed in various ways (Bose, 2002). Visualisation and critical decision-making information are not only displayed through stored data. However, they are also mentioned by modifying data mining techniques and the environment of data warehouses according to organisations (Yong et al., 2003). The importance of the UI design, in this case, lies with its users, who must be granted a certain level of access to complete the task, as the potential for data abuse increases when broad access is granted for general use (Vance et al., 2012).

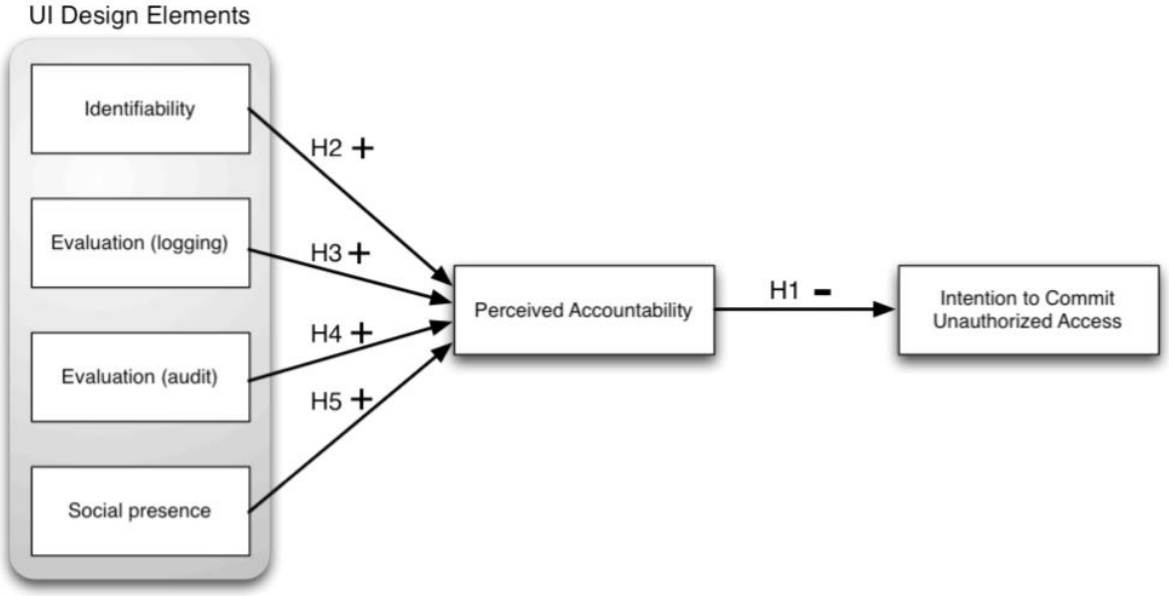


Figure 6: Unauthorized access UI design model through user accountability from Vance et al. (2012)

Figure 6: Unauthorized access UI design model through user accountability from Vance et al. (2012) comes into play by taking four UI design elements into account. The model will increase user perceptions of accountability and mitigate the effects of these elements on unauthorised data access abuse (Vance et al., 2012). Given all of the sensitive data that an (AI-integrated) CRM system handles, the long-standing solution has been to limit user access. However, this does not account for the abuse of the privileges granted to complete the job (Saltzer, 1974). When it comes to granting data access, the complex part is when AI becomes involved because data should be readily available in case any decision-making or real-time information is required (Tawalbeh & Saldamli 2021). In the context of an AI-integrated CRM, this includes data storage and data analysis and manipulation to be formed as valuable data (Ardagna et al., 2017). Thus, data governance guidelines are an essential component for regulating AI behaviour, user access, maintaining data security and privacy when data is used (Oussous et al., 2018).

2.2. AI-integrated CRM

The need for AI is growing, and it is being used for various tasks involving data analysis to find patterns and make predictions (Chatterjee et al., 2019). Companies fall behind and cannot increase revenue and optimise customer loyalty without an AI-integrated CRM, resulting in a

competitive disadvantage (Grace et al., 2015). AI-integrated systems enable accurate decision-making by autonomously analysing Big Data, improving the organisation's overall business processes (Chatterjee et al., 2019). Customers' data is now ready for use thanks to the integration of AI and can be further implemented to be deployed quickly for analysis (Verma & Verma, 2013). Processed data allows for the quick integration of other platforms, such as cloud computing and reduces complexity for a lower cost (Wen & Chen, 2010). As a result, the combination of AI and CRM is more than technological advancement; it is necessary to help analyse data and improve decision making, allowing businesses to expand their business processes and strive for success (Chatterjee et al., 2019).

The impact of AI on a CRM has been mentioned previously (Tahir & Khan, 2021) and its applications. With an AI-integrated CRM, these additional applications can help optimise the business operations by (Chatterjee, 2019):

1. Reducing time in repetitive tasks still needs to be executed manually. Through automating routine tasks, AI can help handle activities, data modification, decision making, predictions, etc. The AI-integrated CRM will help organisations make decisions and recommendations through the processed data. Such as targeting the right customer to build a long-term relationship.
2. Learning historical patterns and habits by its customers, making customised services available. Targeting the appropriate category by customers and helping prioritise the best leads that will help the procurement department to pursue.
3. Gathering correct insight by analysing Big Data swiftly can help build target profiles. With the focus on the specific customer, interactions can be highly personalised and enhance customer satisfaction. Enabling future retention of customers and a streamlined communication process.
4. A realistic roadmap can be made based on predictions and historical events. Creating a stepwise team approach lead to deal results in effective and better results under any circumstance.
5. Assisting virtually, automating customers' responses, question-answer prediction, etc. Sending appropriate responses through the right communication channel to the customer, saving the employees time and helping the customer within an appropriate time.

Highlighting the most important additional values, an AI-integration CRM will be an indispensable asset to the business organisation (Schrage & Kiron, 2018). Research showed that companies believe in wanting to implement AI due to its competitive advantages, yet only 20% have implemented such capable systems (Ransbotham et al., 2017). Another study revealed that 39% of those firms have a proper setup to execute AI in its implementation (Tahir & Khan, 2021). Prominent organisations like Salesforce, Zoho, SugarCRM, etc. have successfully applied AI in their CRM systems throughout their whole platform (Chatterjee et al., 2019):

- Salesforce: Called "Einstein" is their AI tool. Delivering recommendations through machine learning predictions based on gathered customer data.
- Zoho: "Zia" is introduced as a "conversational AI assistant". Which can be used on any provided data, regardless of its complexity and can hold conversations in data analytics, such as a mobile application.

- SugarCRM: “Hint” searches independently based on a simple command, providing inputs that help gather business or personal information.

2.2.1. Machine learning

AI is a broad term that is used in a variety of industries and is frequently misunderstood in its applications (Holzinger, 2018); therefore, specifying the AI used in CRM systems is primarily through Machine Learning (ML) in data mining (Vafeiadis et al., 2015). Artificial Neural Networks (Kirui et al., 2013; Kraljevic et al., 2010), Naïve Bayes (Bahari & Sudheep, 2015; Kirui et al., 2013), Decision trees learning (Radosavljevik et al., 2010; Kraljevic et al., 2010), logistics regression (Kraljevic et al., 2010), etc. One of the most well-known types of AI is machine learning (ML). Its application is indecision and outcome prediction using historical data, requiring no technical knowledge (Iqbal & Khan, 2021). These algorithms are specifically designed for processing large amounts of data via pattern recognition, resulting in immediate future outcome-based predictions for humans to understand (Mahdavinejad, 2018). Compared to manually going through all collected data, analysing and finding patterns would be too much (Iqbal & Khan, 2021). ML is a supervised learning technique because the goal is for the machine to learn a classification system that will be created based on given inputs, as shown below (Nasteski, 2017):

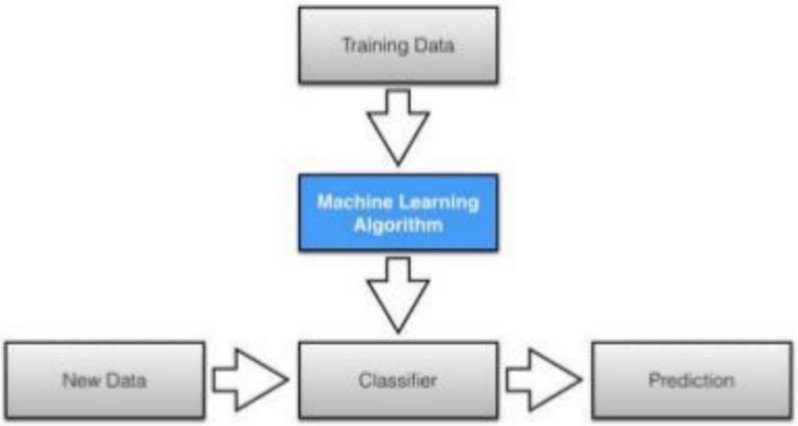


Figure 7: ML supervised learning model from Nasteski (2017)

An Artificial Neural Network (ANN) can either be hardware-based (represented as neurons in physical components) or software-based, using a variety of learning algorithms (Vafeiadis et al., 2015). A popular supervised model with a variation of the Backpropagation algorithm is the multi-layer perceptron and is feed-forwarded (Way et al., 2003). Simply put, the Back Propagation algorithm evaluates the output of ANN against the desired output, where if results are not satisfied, a connection between layers is changed repeatedly until the expected outcome is satisfied with a small enough error margin (Cilimikovic, 2015). ***Figure 5: CRM-data mining framework from Bahara & Sudheep (2015)*** shows a visual representation based on ANN and the Naïve Bayes method. ANN’s implementation into CRM systems is one of the more popular approaches towards data mining, as it outperforms decision trees and logistic regression (Vafeiadis et al., 2015).

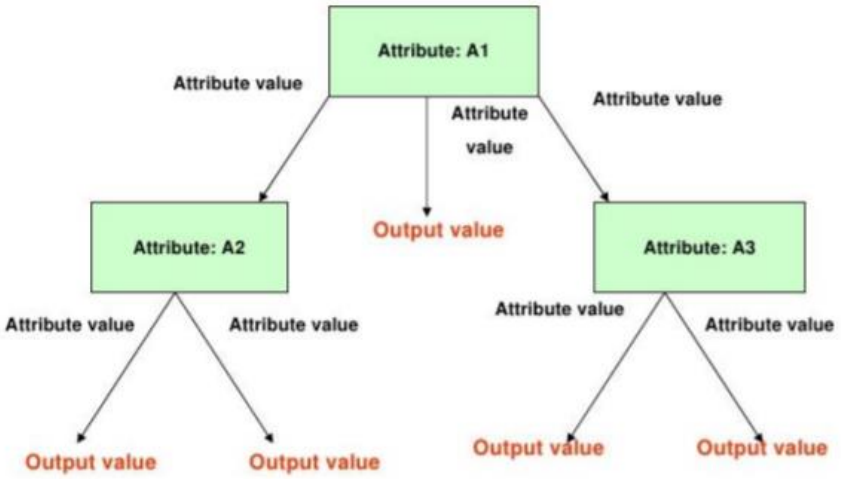


Figure 8: Decision tree example from Vafeiadis et al. (2015)

Tree-shaped structures (Nasteski, 2017) represent sets of capable decisions generating rules that classify datasets into specific categories are decision trees (Lee & Siau, 2001), which can be translated to a structure that is used for dividing up extensive collections of customer data from the database(s) into smaller sets of categories, through a succession of simple decision rules (Linoff & Berry, 2011). The so-called leaves from the tree (squares in the figure) can be said to represent categories for its customers, and branches (square lines) can be said to represent conjunctions of features that lead back to the specified categories (Vafeiadis et al., 2015). Hence, significant in the use of finding patterns and relations in customer behaviour, as the performance can differ depending on the complexity and linearity of the data to be filtered through the decision tree; accuracy can be of high quality depending on how the data is formed (Hadden et al., 2006).

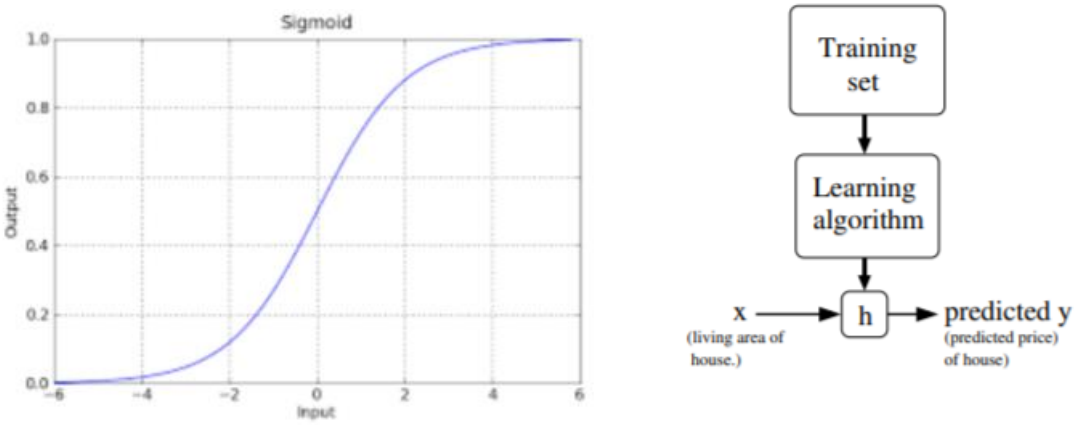


Figure 9: Logistics regression function examples from Hadden et al. (2006)

Logistics regression (Nasteski, 2017) can be described as analysing statistical processes for relationship estimation between variables, including many techniques for variable distinction and modelling depending on the focus (Vafeiadis et al., 2015). Similar to Naïve Bayes, logistics regressions work by extracting some weighted features from the input data, taking logarithms and combining them linearly (Nasteski, 2017). Simply put, the addition is performed by multiplying each feature by weight. This type of regression predicts the probability of an event, which can be applied to historical customer data and be put through a logistics function to

calculate its probability (Vafeiadis et al., 2015). Taking into account several predictive variables to be a numerical or specific category (Kraljevic, 2010).

The described Naïve Bayes framework, through the visualisation of ***Figure 5: CRM-data mining framework from Bahara & Sudheep (2015)***, can be said to be a simple probabilistic classifier based on solid independence assumptions. In simple terms, the classifier assumes that either a particular class feature, such as a piece of contact information, is unrelated to the absence or presence of any other features, such as customer preferences (Kraljevic, 2010). Combining these observed data and determined outcomes can provide a calculated probability through the proposed situation, e.g., a higher percentage for the customer to prefer the colour red or blue through the independent variable data of customer preferences or customer contact information (Nasteski, 2017).

2.2.2. Big Data

The term "Big Data" has been mentioned a few times already, having significantly impacted all aspects of the current industry landscape (Elish & Boyd, 2017). *BD (Big Data)* can be defined by the amount of information involved, its complexity, and how fast it is coming in (Hoffman, 2015). The following three V's in BD can be identified through variety, volume and velocity (Pence, 2014; Yang et al., 2019):

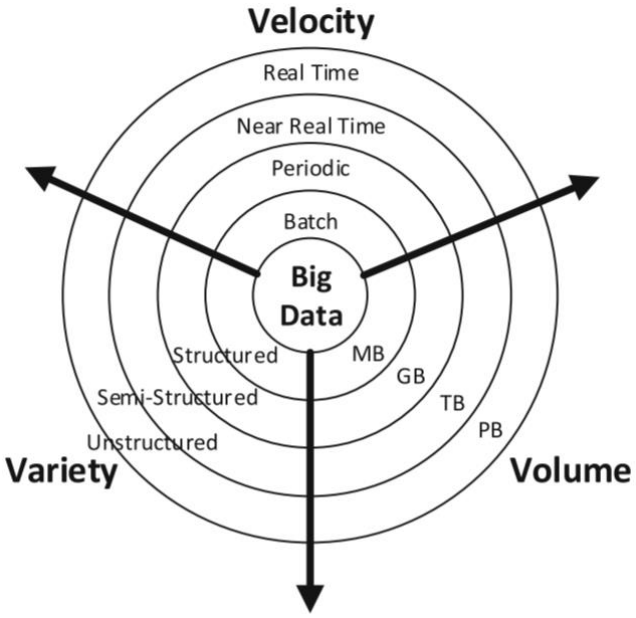


Figure 10: Big Data factors from Yang et al. (2019)

To understand the context where an AI-integrated CRM relates to BD is the breadth and scope of customer data (Libai et al., 2020). Organisations at this time already have databases filled with data fields that are ready to be processed and used (Deighton, 2019). BD can be achieved through multiple partnerships with external connections to customers, but perceived trustworthiness needs to be achieved first as a sense of safety and security (Bart et al., 2005). As organisations are actively seeking more data that can be used, external entities provide and exchange data where that value can be reused in other ways (Libai et al., 2020). Although increasing the volume of data is easy, the challenge lies in its variety (Pence, 2014). Suspecting

that once enough volume has been categorised for validation, the more opportunities there are for an organisation to discover a variety of different and new patterns that can be used (Deighton, 2019). Hence, the key factors explain the expectation that the analysis of BD will allow for more accurate identification of customer characteristics, providing better forecasting and decision-making predictions (Pence, 2014).

BD is a potential enabling factor for business process innovation and a possible new form of value creation (Fosso Wamba et al., 2015). However, such factors remain to be explored further (George et al., 2014). Such innovations can potentially be triggered by the increase of data available through the sheer volume and velocity that data is processed, as a variety follows (Zerbino et al., 2017). BD and its analytics are increasingly transforming customer-facing industries, collecting large amounts of data, such as customer behaviour and preferences, enabling real-time decision-making (Bean & Kiron, 2013). As companies are still trying to cope with the amount of data and its spread amongst the increasing data sources that can be both structured and unstructured, organisations are trying to figure out the potential value of creating customer insights (Zerbino et al., 2017). Aligning innovative data sources into an AI-integrated CRM can be a struggle for some organisations (Phillips-Wren & Hoskisson, 2015). However, many organisations have already overcome this obstacle and lead companies in customer relations, such as Salesforce (Chatterjee et al., 2019). Through the use of BD and AI, from the described machine learning types, Big Data Analytics (BDA) takes place in an AI-integrated CRM system to perform actions, such as predictions and decision-making (Libai et al., 2020).

2.2.3. Sources of Big Data

A traditional CRM would mostly take advantage of data already presented in a particular structure, such as ERP systems, SCM (Supply Chain Management) systems, etc. (Zahay et al., 2012). However, with the growth of BDA and its success (Shahbaz et al.,2020). A change of focus has taken place where data is now also gathered through E-commerce (Electronic commerce), I-commerce (Internet-enabled commerce), M-commerce (Mobile commerce) and currently, U-commerce (Ubiquitous commerce) (Liu, 2015). As there is a lack of literature regarding how BD could affect the development and management of an AI-integrated CRM, the importance of understanding these available data sources can change the critical success factors of how an organisation might be run (Zerbino et al., 2017). Hence, the types of data sources could be visualised and focused on further BDA implementation for more in-depth use (Liu, 2015).

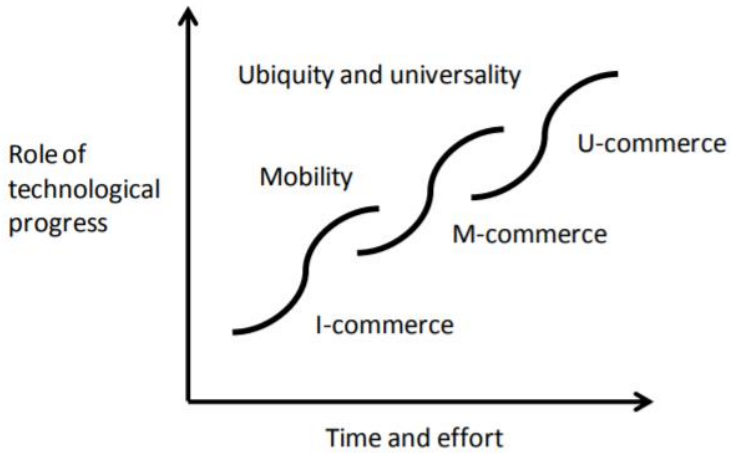


Figure 11: S curves ranking commerce innovations from Liu (2015)

In the context of BDA, E-commerce enables organisations that provide products/services to track each user's behaviour and connect patterns to allow for long-term customer relationships (Akter & Wamba, 2016). However, the shift from E-commerce has taken place to I-commerce, allowing customers to have a commerce experience without the physical restriction of a brick and mortar store (Liu, 2015). Adding even more valuable data that can be coupled to the customer, from browsing behaviour, online activities, connected social networks, etc. (Anshari et al., 2019) and arriving at M-commerce, which is an extension relating to the involvement of all kinds of electronic transactions through the use of mobile devices (Niranjanamurthy & Kavyashree, 2013). Allowing for an unreasonable amount of consented data access, such as face/finger recognition for password login through the website and geolocation, can pinpoint its customers' exact location (Anshari et al., 2019). The growth of ubiquitous computing allows organisations to acquire all detailed information from customers or prospective customers at any time and place, from any capable enabled device with an internet connection (Liu, 2015).

Ubiquitous computing can be thought of as small, wireless, intercommunicating microprocessors embedded into objects with a range of sensors and capabilities to map the surrounding environment (Krum, 2011). Basically, through the use of all mentioned commerce methods and access to internet data that is public domain, uninterrupted communication/transaction can take place between an organisation and its customer (Liu, 2015). Resulting in an indescribable amount of personal data that can relate to an individual or another organisation that stores such information; privacy and security are significant concerns regarding the used IoT (Internet of Things) devices and access to data (Niranjanamurthy & Kavyashree, 2013). Hence, data storage practices and AI regulation need to take place in deploying an AI-integrated CRM. If concerns come to light, trust is broken, and data exchange will stop at that concerning event (Libai et al., 2020). Resulting in a competitive disadvantage to an organisation and damage to its customers (Zerbino et al., 2017). This opens up dangers to the legal aspect and can be the downfall of any organisation (Janssen et al., 2020).

2.2.4. Data storage & Deployment

Having mentioned that BD is foremost large amounts of data sets that cannot be manually analysed through its regular database tools (Anshari et al., 2019), data still needs to be stored somehow for further processing through BDA (Deighton, 2019). As databases cannot solve all

the aspects of BD and the machine learning algorithms used throughout the process for users to be translated into simple visualisations (Katal et al., 2013). Hence, Scientific Data Infrastructures (SDI) provide a basis for building interoperable systems with data sharing (Yuri et al., 2012). With current trends in deploying an AI-integrated CRM, cloud-based infrastructures have been easily implemented with an SDI (Katal et al., 2013). Compared to traditional storage methods, such as a Single Query Language (SQL) database without machine learning implemented (Taft et al., 2020), cloud-based solutions are cost-efficient, accessible, reliable, responsive and cheap to maintain (Anshari et al., 2019). An SQL database can be defined as a relational database sublanguage through the concept of a table with columns and data types represented as rows (Melton, 1996).

Tools such are Hadoop and MapReduce allow for BD scalability, processing and data storage in a database (Katal et al., 2013). Hadoop Distributed File System can be used as the backbone, storing multiple types of large data sets that can have any structure, and MapReduce allows for the logic of processing the BD (Merla & Liang, 2017). Different types of databases have since come out with the same basic architecture of SQL, such as MySQL, Oracle, and NoSQL, that allow for complex logic and fast retrieval of data (Zaki, 2014). With the amount of data that needs to be accessed globally, geo-distributed solutions (such as a cloud setup) need to be optimised for performance and regulatory reasons (Taft et al., 2020). With the lack of research in data governance that takes into account the security and privacy aspect of the components in an AI-integrated CRM (Chatterjee et al., 2019), abuse of such information is only a matter of time, resulting in catastrophic consequences for both organisations and customer (Janssen et al. 2020). The controversy of applications to BD has already risen regarding privacy and security (Pence, 2014).

2.3. Data Governance

Data governance can broadly be defined as decision rights and responsibilities through a system that outlines who or what can take action with the data in question and under which circumstances, what methods, and when specifically (Gupta et al., 2020). With the rise of Big, Open and Linked Data (BOLD), the increase in BDA and Big Data Algorithmic Systems (BDAS) that is based on the mentioned machine learning takes place in an AI-integrated CRM (Janssen et al., 2015). The rise in decision-making consequences to society, from individuals to organisations, cannot tolerate failures and need stringent regulations and ethical requirements (Janssen et al., 2020). Critical concerns are being raised about how BD needs the appropriate governance frameworks for quality data access that allows machine learning techniques (Tsai et al., 2018). Hence, ensuring that a BD governance framework regulates the storage and processing of data from organisations in a responsible way that is both in order from the legal and ethical side (Yang et al., 2019). In order to fulfil this goal, BD governance should focus on the systems where data is collected, managed and used (Benfeldt et al., 2020).

BD governance is dependent on the collaboration of the system's many affiliated organisations and people (Janssen et al., 2020). Ensuring reliable and secure data-sharing between the involved parties and ensuring that the data complies with the General Data Protection Regulation (GDPR) is required (EU commission, 2017). As a result of data governance, BDAS is thriving and has an overall positive effect on organisations implementing systems such as AI-integrated CRM while lowering data-related costs and risks (Abraham et al., 2019). BDAS

decision-making errors in BD governance frameworks can affect system functionality and legal, financial, and social consequences for organisations and their stakeholders (Kroll, 2018). Mistakes caused by faulty deployment result in systemic bias, significant financial exposure, a political crisis, illegal decisions, and even the loss of lives, among other adverse outcomes (Janssen et al., 2020). BD governance serves several functions in an organisation and aids in the achievement of goals through the data that must be protected, from which the following two aspects can be extracted (Yang et al., 2019):

- Data risk identification, such as personally identifiable information and constitute sensitive data (medical records), which the outcome has been discussed by Janssen (2020). Hence, a BD governance framework can identify sensitive data, preventing such situations.
- Safe access control practices where certain users do not need to view or access sensitive data for daily usage. Hence, a BD governance framework helps control sensitive data by applying the correct data governance methods.

Traditional methods that are applied in a regular CRM do not take into account the BDAS, BDA, BOLD and BD in general (Janssen et al., 2020), as it is becoming too complex for regular tools to handle the increasing BD with the growing IoT devices that gather data (Wang, 2017) for the following reasons (Yang et al., 2019):

- Traditional methods do not consider semi-structured or unstructured data, which is very common in BD. Hence, manually transforming the data, such as in an ERP, is time-consuming and can be very costly.
- Traditional methods do not consider retrieving, accessing, storing, processing, and retaining large volumes. Hence, it is inefficient as the wrong tools are being used for BD.

Janssen et al. (2020) redefine BD governance as "organizations and their personnel defining, applying and monitoring the patterns of rules and authorities for directing the proper functioning of, and ensuring the accountability for the entire life-cycle of data and algorithms within and across organizations".

2.3.1. Privacy & Security

The privacy and security of BD, in general, have gained much momentum in the research space due to the growth of Cloud Computing, Social networks, etc. (Cuzzocrea, 2014). In terms of BD research, the ethical issue of ensuring the privacy and security of databases is challenging (Wu & Guo, 2013). Privacy can be defined to have the privilege to control how personal information is collected and used, such as restricting organisations or groups to identify a person during the use of the internet (Porambage, 2016). Whereas security can be defined as defending the information and its assets through training, processes and technology; preventing disruption, modification, recording, inspection, disclosure, unauthorised access and destruction of the data and infrastructure in question (Jing et al., 2014). Where the primary differences

between security and privacy in BD governance can be shown in the following table (Jain et al., 2016):

S.No	Privacy	Security
1	Privacy is the appropriate use of user's information	Security is the "confidentiality, integrity and availability" of data
2	Privacy is the ability to decide what information of an individual goes where	Security offers the ability to be confident that decisions are respected
3	The issue of privacy is one that often applies to a consumer's right to safeguard their information from any other parties	Security may provide for confidentiality. The overall goal of most security system is to protect an enterprise or agency
4	It is possible to have poor privacy and good security practices	However, it is difficult to have good privacy practices without a good data security program
5	For example, if user make a purchase from XYZ Company and provide them payment and address information in order for them to ship the product, they cannot then sell user's information to a third party without prior consent to user	The company XYZ uses various techniques (Encryption, Firewall) in order to prevent data compromise from technology or vulnerabilities in the network

Table 1: Difference between privacy and security in BD governance from Jain et al. (2016)

Despite the differences, it is well understood that BD's data security and data privacy are closely linked (Maniam & Singh, 2020). However, a framework emphasising BD governance in privacy and security has yet to be clearly defined (Al-Badi et al., 2018). As a result, the need to rethink and redesign established data processing frameworks, such as ML (Ishibuchi et al., 2013), ubiquitous computing (Belsis and Pantziou, 2014), etc., must be refocused on a per-system basis (Cuzzocrea, 2014). As people become more aware of how their data is being used, organisations that use data exchanges for commercial purposes must be regulated to protect personal and sensitive information (Ogbuke et al., 2020). To avoid situations like the Facebook-Cambridge Analytica data scandal, where data was abused at the expense of its users' safety (Mikalef et al., 2019). In addition, due to unsafe practices, databases containing sensitive data have been leaked on the internet several times (Maniam & Singh, 2020). Arriving at the ethical issues in business practises regarding the use and exploitation of sensitive data, which is primarily present in an AI-integrated CRM (Ogbuke et al., 2020).

Due to the vulnerability of sensitive data, it is critical to consider data security, as the security and privacy of data about its customers can otherwise be accessed without consent (Mikalef et al., 2019). As a result, customer concern about their lack of privacy due to unregulated practices by organisations is growing, particularly with the use of live-tracking information such as their mobile devices (Ogbuke et al., 2020). Studies have shown that data protection was among the top issues where only a few people were willing to share the information collected for its original purpose (Kshetri, 2014). The organisation is responsible for ensuring that user and product data are not redistributed without the owner's permission (Ogbuke et al., 2020). Cyber security must be considered with an estimated 20 billion connected devices with IoT capability (Gartner & Brocca, 2015). The situation worsens when using cloud computing platforms on which an AI-integrated CRM relies (Musa & Dabo, 2016). Luckily, several approaches enable secure and privacy-enabled platforms, where risks are reduced or eliminated (Maniam & Singh, 2020).

2.3.2. Approach

For BD governance to occur, the organisation's structure needs to be set up to support a data-driven BDAS or, in this case, an AI-integrated CRM (Janssen et al., 2020). Without data sharing, silos will occur, and the information mismatch due to the lack of the systems that an organisation uses do not work together(Kroll, 2018). As well as the other mentioned points in security, such as the vulnerable aspect in unauthorised access. The choice of a BD governance approach is critical but not always explicit or clear due to a lack of standard regulations (Koltay, 2016). Hence, the following approach toward a BD governance can be used, complementing each other from the organisational aspect to planning and control and risk-based data (Janssen et al., 2020), as shown in ***Figure 12***: Approach towards data governance from Janssen et al. (2020**.**

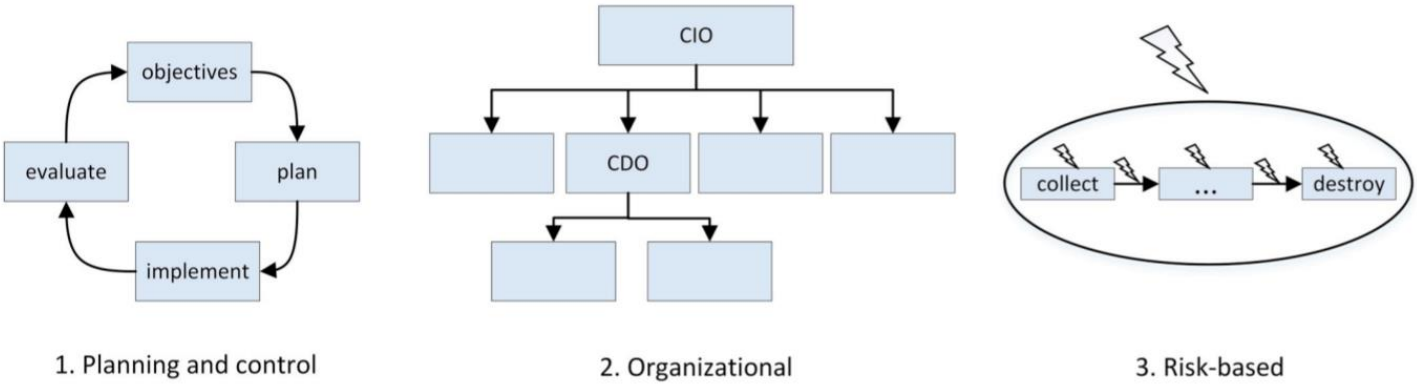


Figure 12: Approach towards data governance from Janssen et al. (2020)

The planning and control approach is based on an annual cycle, in which objectives are set, projects are defined, implemented, and evaluated, and budgets are allocated. This approach is also common in IT governance frameworks, where BD governance is carried out through procedures that can be checked, modified, and repeated (De Haes et al., 2013). Departments within the organisation compete in terms of performance, which is evaluated and aligned between business processes and goals for the implemented technology goals (Janssen et al., 2020). Planning is the infrastructure of a project in which specific areas of data quality can be improved and potential risks identified (De Heas et al., 2013). However, while constant oversight can help adjust project resources and timelines, this approach is frequently criticised for its complexity to change (Janssen & Van der Voort, 2016).

The organisational structure's approach emphasises the importance of top-down accountability, responsibility, and reporting for BD governance, defining authority (Mullon & Ngoepe, 2019). As a result, decision-making structures in AI, privacy and security ethics, etc., can be established, such as Chief Information Officer (CIO), Chief Data Officer (CDO), Chief Privacy Officer (CPO), etc. (Janssen et al., 2020). Responsibilities for data stewardship can be defined within such a command structure (Rosenbaum, 2010). Because of the increased concern about AI in GDPR, the risk-based approach allows for identifying risks in a BDAS and a means to prevent these risks through governance (Ladley, 2019). Often referred to as the foundation of data governance, it can be applied to a wide range of AI-specific risks, including algorithmic error, data discrimination, irregularity, and so on (Janssen & Kuk, 2016). These issues are typically caused by ML algorithms from an AI-integrated CRM failing to function correctly due to the large amount of structured and unstructured data that can be outdated, stolen,

missing, biased, or inaccurate (Beretta et al., 2018). After discussing these approaches, it should be noted that different governance mechanisms should be implemented with caution, as too much governance can result in excessive overhead due to the specific nature of each situation (Janssen et al., 2020).

Given that each industry can be driven by BDA, from marketing to information technology, there are only a few studies on regulatory issues for BD governance (Al-Badi et al., 2018). Studies are foremost focused on its analytics, cloud, architecture, social media, etc. (Akoka et al., 2018). Hence, several existing frameworks on BD governance can be shown for general purposes (Al-Badi et al., 2018) that are relevant for building a data governance framework in an AI-integrated CRM.

Frameworks	Year	Components/ characteristics
BGF1-[36]	2018	Objective, strategy (personal information protection strategy, data quality, and the data disclosure/accountability strategy), components (organization, standards and guidelines, policies and process), IT infrastructure (audit and control, Big Data infrastructure).
BGF2-[37]	2017	Data consumers, self-provisioning data portal, optimize and compute, data infrastructure or tired storage
BGF3-[38]	2017	Data analytic, data querying, distributed data processing, distributed data storing, data acquisition.
BGF4-[39]	2017	Governance objectives, the top-level design, governance objects, governance methods, the internal and external environments and contributing factors.
BGF5-[40]	2016	Quality and consistency, policies and standards, security and privacy, compliance, retention and archiving.
BGF6-[26]	2016	Organization, metadata, privacy, data quality, business process integration, master data integration, information lifecycle management.
BGF7-[18]	2015	Big Data governance framework (content in accessible).
BGF8-[31]	2015	Discover, define, apply, measure and monitor.
BGF9-[29]	2013	(a) a maturity assessment to determine readiness for data governance, (b) a business case to justify implementing data governance, and (c) a roadmap to guide the data governance implementation.
BGF10-[41]	2013	Establish difference between traditional data and Big Data governance, establish basic rules for where new data governance can be applied, establish processes for graduating the products of data science to governance, and establish a set of tools to make governing Big Data feasible.
BGF11-[17]	2012	Strategy, organizations, policies processes and standards, measurement and monitoring, technology, communication.
BGF12-[42]	2012	Big Data types (i.e. web and social media, machine-to-machine (M2M), big transaction data, biometrics, and human-generated), Information governance disciplines (i.e. organization, metadata, privacy, data quality, business process integration, master data integration, and information lifecycle management), industries and functions (i.e. marketing, customer service, information security, or information technology).

Table 2: BD governance frameworks from Al-Badi et al. (2018)

However, there should be noted that a general BD governance framework does not suffice per BDAS, as each application is specific, and improper governance implementation can result in improper system functioning, such as in an AI-integrated CRM, leading to the aforementioned negative consequences for organisations (Kim & Cho, 2017; Yang et al., 2019). Traditional data governance attributes, such as business processes, organisational structure, information life cycle, etc., could still be identified and used in BD (Al-Badi et al., 2018). In addition to the critical aspect of the required data protection policies, privacy, optimisation, and BDA implementation (Dai et al., 2016). ISO (International Organization for Standardisation) 8000, which includes the global standard for data quality and master data for enterprises (Al-Badi et al., 2018), and ISO 38505, which includes data governance guidelines, such as company structure setup (ISO, 2017).

Returning to the topics discussed in the literature review, the importance of organisational structure and responsibilities in data stewardship is highlighted (Morabito, 2015). Any data, structured or unstructured, needs to provide quality data through cleaning and processing (Ularu et al., 2012), which can be further processed through data mining (BDA) and analysis

where data is being stored in a database and can be accessed anytime from anywhere (Al-badi, 2018). Hence, the following governance approach discusses all the mentioned aspects and is perfect as a building stone for the data governance framework for an AI-integrated CRM system. Due to the extensive research performed by Prof.dr.ir. Marijn Janssen in the field of data governance, BD, AI, etc. As well as being a field expert and chair of ICT and data governance at Delft University of Technology. The approach pertaining to the process of a BDAS, in general, is described where BD and ML are interconnected (Janssens et al., 2020):

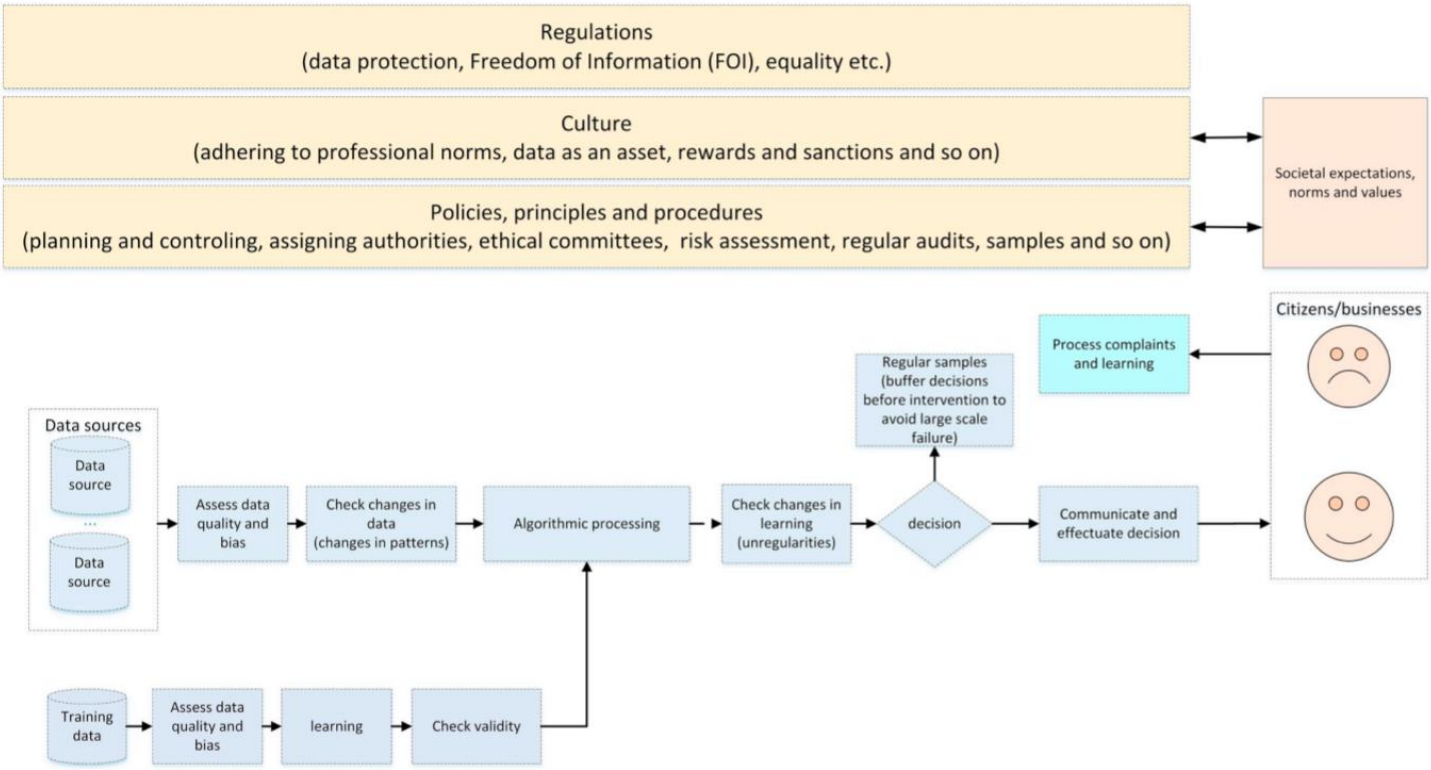


Figure 13: BD governance for BDAS from Janssens et al. (2020)

Describing a system-level governance framework that encompasses BDAS when handling data, showing the data sources at the left, the algorithm processing in the middle and the expected output on the right. This framework takes in the input data, which enables automatic decision making and is hidden from a regular user, which can be dubbed to be playing the "hidden bureaucrat" (Wihlborg et al., 2016), which is essential because of the accountability aspect, where people are responsible for mediating the response of the BDAS (Kroll, 2018). The top of ***Figure 13: BD governance for BDAS*** from Janssens et al. (2020) displays how the system is guided through regulations first, from the described aspects, such as protection of data, principles and procedures (Janssen et al., 2020). Where policies relate to the user's behaviour that pertains to the access of data, and the algorithm and principles define the logic of the governance in data (Kim & Cho, 2017). The bottom part of ***Figure 13***: BD governance for BDAS from Janssens et al. (2020) displays the learning processes of the ML algorithm in order to make decisions, as such process needs to be monitored due to historical inconsistencies of data or the generalisation that each algorithm reacts the same to the given input data (Janssen et al., 2020). Although going through the framework, it becomes clear that there are a lot of fail-safes in place, additional governance mechanisms may still be required (Kroll, 2018).

The main component of data governance is collecting data responsibly to prevent abuse or misuse when sensitive data is collected, such as race, gender, health status, residential address, etc. (Janssen et al., 2020). As a result, data stewards must ensure that information sharing is done responsibly, ensuring the quality and validity of data and the aspect of security in managing risk to preserve data with integrity (Dawes, 2010). Having concluded that data stewards are also responsible for information security (Cuganesan et al., 2017). Trusted frameworks must have authorisation services that can check for identification and authentication before allowing access to data (Janssen et al., 2020). Even though organisations rely on collaborations for data exchanges, compliance standards such as GDPR must still be followed for interoperability (Kroll, 2018). Considering the following elements to regulate data sharing and its security (Janssen et al., 2020):

- Requirements for trusted data sharing.
- Set standard for trusted data sharing to take place.
- Contracts and agreements that allow for trusted data sharing.
- Authorization platform, allowing only authorized users to access data under certain circumstances.
- Recording mechanism to allow for monitoring which parties are certified.
- Auditing mechanism that allows verifying all the points mentioned above.
- Enforcing mechanisms for compliance with the agreements and rules.

Data sharing should be limited to those who need to know and, if necessary, anonymised as a preventive measure (Potiguara et al., 2020). If data is shared without consent, the person or organisation should be made aware to avoid misuse and correctness, such as hacked passwords that have been reused for other services (Kroll, 2018). The data governance framework has been mentioned as the foundation of a trustworthy BDAS. However, AI systems that handle BD are still a complex field, and developing such a framework requires attention to avoid risks (Janssen et al., 2020). As a result, taking into account the following BD governance principles from Janssen et al. (2020) allows for the construction of a solid framework:

Name	Description
1. Evaluate data quality and bias	When data is used by BDAS, its quality, and possible embedded bias should be evaluated.
2. Detect changing patterns	When the outcomes of the algorithms change, their validity should be checked, and the reasons for such changes investigated.
3. Need to know	Minimize the amount of data that is shared by only sharing what is necessary, e.g. answers to questions instead of complete datasets.
4. Bug bounty	Rewards could be used to encourage people to spot errors and issues and report them back.
5. Inform when sharing	When governments share data about a person or an organization, these entities should be informed to ensure transparency and avoid misuse.
6. Data separation	Separate personal from non-personal data, and sensitive from non-sensitive data (Janssen et al., 2017).
7. Citizens control of data	Empower citizens and organizations to be in control and check the accuracy of their data.
8. Collecting data at the source	Collect data at the source to ensure its correctness and to know how such data is collected (Hammer, 1990).
9. Minimize authorization to access data	If a party does not need data, access should not be granted.
10. Distributed storage of data	Distributed systems are less vulnerable and avoid easily combining data without permission.
11. Data stewards	Assign data stewards to formalize accountability for managing information resources while adhering to the principle of the separation of concerns (Dawes, 2010).
12. Separations of concerns	Responsibilities for data should be distributed in such a way that no single person can misuse or abuse data.
13. Usefulness	Data should be recognized as a valuable asset that can be used by BDAS (Dawes, 2010).

Table 3: Data Governance principles

These principles may appear simple to implement, but the challenge is making them a reality (Kroll, 2018). As there is little research into such trustworthy frameworks, it can be said that only a limited number of practices allow for the successful adoption and implementation of

data governance in a BDAS (Janssen et al., 2020). In addition, Self-Sovereign Identity (SSI) provides control and ownership of an organisation's gathered data from customers and can share such data with others (Toth & Anderson-Priddy, 2019). As a result of discussing a general BDAS and BD governance, several components of an AI-integrated CRM may require additional attention (Chatterjee et al., 2019). Both privacy and security must be in order for an AI-integrated CRM to function correctly, as both are intertwined, as previously stated (Maniam & Singh, 2020). Accountability and authorised access are the starting points for any user to interact with an AI-integrated CRM, as ***Figure 6: Unauthorized access UI design model through user accountability from Vance et al. (2012)*** have mentioned techniques that rely on user accountability. Data storage and the cloud environment have also been discussed in *Data storage & Deployment*, as well as numerous security practices that might help the BD governance aspect in an AI-integrated CRM (Sarmah, 2019).

2.3.2.1. Cloud infrastructure

The shift in how an AI-integrated CRM is deployed is being driven by the rise of Infrastructure as a Service (IaaS)/Platform as a Service (PaaS) providers and the benefits they provide, such as low cost (Serrano et al., 2015). Third-party resources are typically provided to organisations/users in the case of IaaS, which is based on cloud computing and takes the form of a lease (Li et al., 2012). Servers, networking, databases, and other services are provided, with customers only paying for what they require on an as-needed basis (Iosup, 2014). PaaS cloud provides a condensed version of IaaS, including a container environment where applications can run, such as own application development (Vaquero, 2011). As a result, it is not surprising that organisations are shifting to cloud-based solutions, which, as previously stated, are more reliable, accessible, affordable, efficient, etc. (Anshari et al., 2019). However, there has been an increase in cyber-attacks on IaaS/PaaS cloud platforms, primarily due to configuration errors in overprivileged users, Access Control Policies (ACP), and no logging enabled for troubleshooting (Torkura et al., 2020). According to the Cloud Security Alliance, these errors are among the most severe security threats in a cloud configuration, with user error and system misconfigurations accounting for 49 percent of database leaks (CSA, 2019). Concluding, when attacking an IaaS/PaaS, the cloud user is usually the most vulnerable entry point (Torkura et al., 2020).

Before constructing an AI-integrated CRM in an IaaS/PaaS cloud, organisations must be aware of the security and privacy risks, such as a lack of data leak detection and prevention, authentication and authorisation, incidents, encryption, and infrastructure hardening (Maduka et al., 2017). Cloud Service Providers' (CSP) Application Programming Interface (API) is typically one of the breaching points when using an IaaS/PaaS, as these incidents occur as a result of a vulnerable application compromising the stored information (ENISA, 2009). As a result, before adopting the CSP's policies and management processes, the cloud user should assess current compliance requirements, control risks, disaster recovery, business planning, and adequate provisions in the Service Level Agreement (SLA) (ISACA 2017). Such things can be assessed using risk frameworks standards and the established security controls in the specific IT domain (Maduka et al., 2017). The Cloud Control Matrix (CCM, 2019), NIST SP 800-146 (Badger et al., 2012), ISO 2700:2013 (Tariq & Santarcangelo, 2016), COBIT 5 for assurance (De Haes et al., 2013), and the previously mentioned ISO 8000 are some of the frameworks

that provide best practises and control in cloud computing security, such as IaaS/PaaS. (Maduka et al., 2017).

There are plenty of renowned CSPs that organisations are using, such as Amazon AWS, Microsoft Azure, etc. (Torkura et al., 2020), that have the mentioned standards (Tariq & Santarcangelo, 2016).

Organization	Security Compliance
Amazon	SOC 1, SOC 2, SSAE 16, ISAE 3402, FISMA, DIA-CAP, FedRAMP, PCI DSS Level 1, ISO 27001, FIPS 140-2, HIPPA, CSA and MPAA
Salesforce	ISO 27001, SysTrust, SAS and 70 Type II
Microsoft	FISMA, PCI DSS, HIPAA, SOX, ISO 27001, SAS 70 TYPE I and II and NIST SP 800-53

Table 5: CSP security Compliance from Tariq & Santarcangelo (2016)

Threats	Controls/Mitigation
1.Data location aware 2.Cloud business continuity 3.Data protection plan and best practice 4.Regulatory Compliance 5.Data Lock-in	Audit checklist
1.Platform Virtualization 2.Network and Internet connectivity 3. Computer Hardware	1.Xen access 2. Encryption of traffic, two-factor authentication 3. High secured locked rooms
Data Leakage protection and usage monitoring 2.End to End Logging and Reporting: 3.Authentication and Authorization 4.Infrastructure Hardening: 5.End to End encryption	1. Information usage. 2. logging and reporting 3. Two level authentication 4.IPsec Encryption 5.VM template hardening
Governance and compliance	Audit checklist

Table 4: Threats and controls from Tariq & Santarcangelo (2016)

Expressly, ISO 27001 can be understood as a security standard to assist organisations in implementing, establishing, and maintaining effective information security management, such as data protection, access management, compliance (e.g. GDPR), malware protection, etc. (Maduka et al., 2017). As a result, selecting the suitable CSP and assessing an organisation's practices and standards that use a cloud platform is critical (Serrano et al., 2015). Due to the abundance of standards and frameworks available, there is a dearth of comprehensive guidelines or controls that address every risk in the IaaS and PaaS cloud environments (Chen & Yoon, 2010). As a result, identifying a general approach for best practices in cloud computing integrity, confidentiality, and availability could be established (Chen et al., 2010). An assurance model that relies on four security compliances with every aspect of security and assurance control is used (Maduka et al., 2017).

2.3.2.2. Authorised Access

An AI-integrated CRM system and its implementation are only as secure as its weakest link (Chatterjee et al., 2020). Having mentioned the *User interface***,** a specific approach through user accountability was mentioned, which would partially solve the problem of data misuse and abuse from overprivileged users (Vance et al., 2012). Because of the restricted control, user access challenges are limited when using a PaaS cloud (CSA, 2019). Cloud users in this situation cannot control or manage the infrastructure, such as data storage, operating system and network servers (Karthiban & Smys, 2018). However, with an IaaS cloud, an average cloud user acts as an administrator. The user refers not only to an organisation's end-user but also to Developers, Network Architects, and Data Stewards (Cuganesan et al., 2017). As a result, changes to an IaaS can fail the proper operation of a BDAS, resulting in insecure data (Janssen et al., 2020). Many security risks can compromise data that contains sensitive information in both situations, from IaaS cloud to PaaS cloud, whether hosted by third parties or an organisation's own database (Indu et al., 2018). Because of the rapid development of cloud computing and BD, specific security measures and protocols may be challenging to keep up with (Xiong et al., 2019).

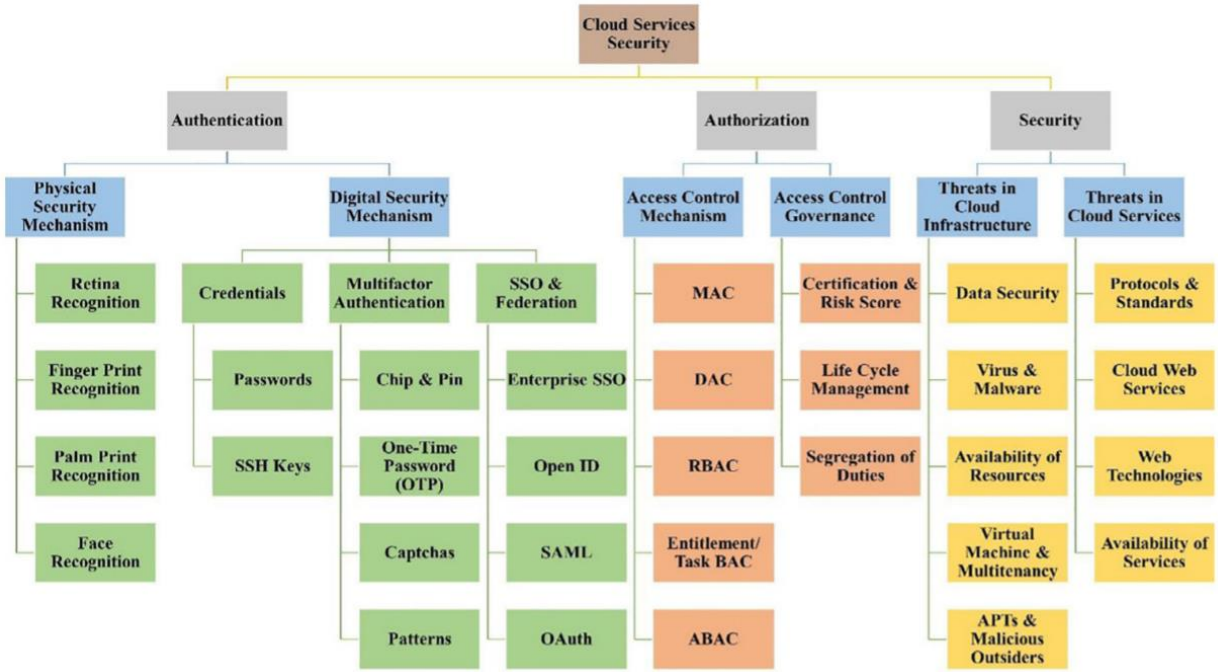


Figure 14: Cloud Service security taxonomy from Indu et al. (2018)

Aside from user accountability and a specific focus on authorisation and authentication, the taxonomy model shown above can help prevent overprivileged users and security threats (Indu et al., 2018). Existing Identity and Access Management (IAM) demonstrates the following practices that provide adequate security in a cloud infrastructure (Sharma et al., 2016), allowing for data integrity, as security and data privacy are inextricably linked (Maniam & Singh, 2020). IAM allows for managing user privileges, information access, authorisation access, and identity control (Indu et al., 2018). From the practises mentioned for a physical site to online digital security, authorisation mechanisms, and privilege governance (Xiong et al., 2020). Serves as a guideline to improve cloud infrastructure's privacy and security aspects, such as an AI-integrated CRM (Karthiban & Smys, 2018).

3. Research Methodology

This chapter will describe the research methodology of research by design used in similar data governance topics. The necessary steps will be outlined with the used approach and what literature has been used. An evaluation will be made describing the deliverables obtained from the research.

3.1. Overview

The appropriate research method that has been used in the past for topics such as data governance and AI applications have been conducted through a design science approach (Cheong & Chang, 2007). Design-based research has gained popularity and is also one of the preferred choices in the field of Science and technology learning (Kennedy-Clark; 2013; Barab & Squire, 2004; Edelson, 2002). Hence, research by design will be used to advance existing theories from the field of security and privacy in AI data governance. A data governance framework emphasising privacy and security will be developed for AI-CRM systems, which leads to the advancement of new theories, a deeper understanding and developments of the topic at hand (Barab & Squire, 2004; Barab et al., 2007; Kennedy-Clark, 2013); focussing on explaining and improving the current situation, compared to the classical explanatory research, which is limited to only explaining that there is a problem (Van Aken, 2005).

The following three phases are used in this study that is based on design-based research (Kennedy-Clark, 2013; Plomp, 2007; Offerman et al., 2009) :

- **Preliminary research** is the first phase where extensive research is conducted from existing literature, understanding the current landscape of the topic at hand and setting up for deriving a prototype framework.
- **Prototyping** is the second phase where an AI-integrated CRM is visualised through the situation of PrexPartners. Afterwards, a prototype framework is being developed through literature review and refined based on qualitative and quantitative data in a survey form, presented to the company's participants and relevant parties involved in the development or subject to the AI-integrated CRM system that is being developed at PrexPartners.
- **Assessment** is the final phase that evaluates if the developed framework answers the proposed research questions through all the gathered deliverables.

3.2. Approach

The approach can be divided into the three aforementioned phases, which will be explained in more detail and lead to the desired result.

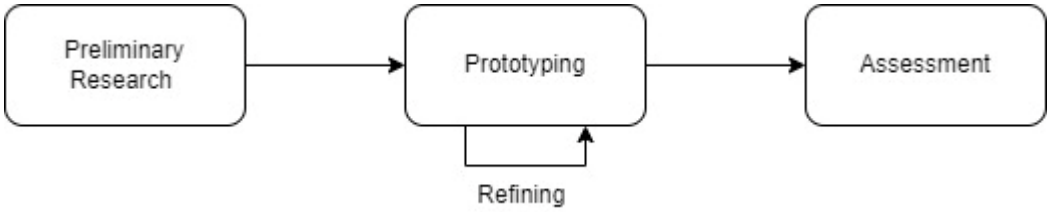


Figure 15: Design-based research approach

3.2.1. Preliminary research

Since there is a lack of existing literature regarding AI-CRM systems in data governance (Chatterjee et al., 2021), an approach to identifying the relevant components was conducted through the *Literature review*. Starting with how a regular CRM system works and what components are relevant regarding privacy and security. After establishing the general process, the differentiation was made on how AI influences the system. As well as critical functions AI has in such a system and the added differences, such as the amount of data handled through ML. Together with explaining the additional data sources instead of a singular ERP source. Having identified all the relevant components and their function in how it works, the structure of an AI-integrated CRM was defined.

Although privacy and security go hand in hand, there are still some differences in their application. Several frameworks were also identified, having found general approaches to data governance and best practices, together with ISO standards specific to data governance and how an AI-integrated CRM is set up, such as IaaS. After thoroughly reviewing existing literature, a base for developing the framework has been found using the framework described by Janssen et al. (2020). Showing a general BDAS approach, which can be adapted to an AI-integrated CRM more specifically. The available literature study spanned *CRM, AI-integrated CRM, data governance, corporate governance, data quality, enterprise application architecture, data stewardship, AI ethics, Machine learning, data collection, Cloud architecture, Big Data, User Access Level*, and other relevant keywords pertaining to these subjects.

Both grey and scientific literature were used to understand the best of both worlds in an academic and actual-world application (Denyer et al., 2008). Such as deriving specific facts from the Cloud Security Alliance to understand the current situation in the field or ISO for their up to date guidelines on data governance. Which is valuable in design science to help understand the importance of data governance and its uses in organisations (Denyer et al., 2008. Data sources were combined, such as the university database and Google Scholar using the mentioned keywords, having sorted through the most relevant and recent papers, and collecting numerous amounts of relevant scientific research. Concluding that enough literature can be found about this topic separately, it still lacks the specifics of data governance from the perspective of security and privacy in an AI-integrated CRM system.

3.2.2. Prototyping

Based on the literature review, we can identify and understand the separate components of an AI-integrated CRM subject to data governance. To illustrate the workings of an AI-integrated CRM system, a high-level process flow can be created based on the development of the system at PrexPartners. Describing each component in detail on how it works and its function, supported with the gathered knowledge of literature review. The processes of discovery allow for understanding the overall structure of an AI-integrated CRM and how data governance can be applied.

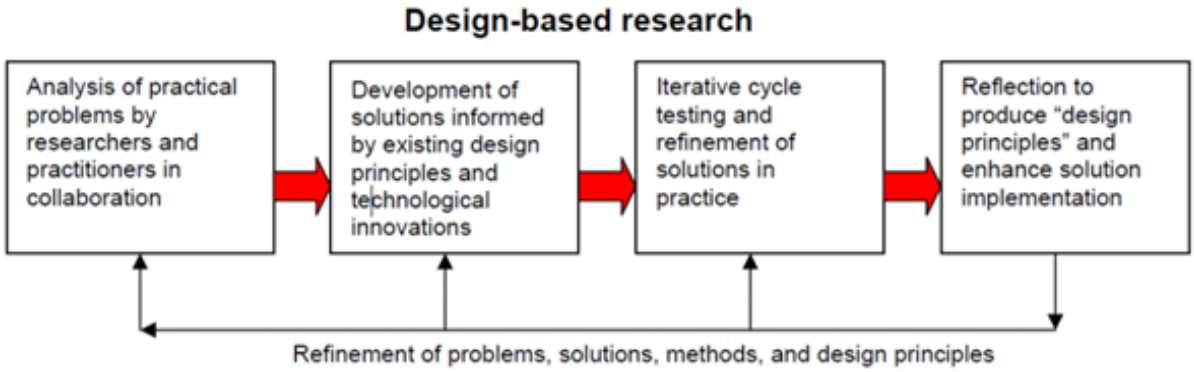


Figure 16: Design-Based research cycle from Kennedy-Clark (2013)

From Kennedy-(2013) Clark's design-based research cycle, we can concentrate on the described development and refinement (iterative) cycle. As previously stated, there is a problem due to a lack of research in data governance in an AI-integrated CRM (Chatterjee et al., 2021). Following the completion of the literature review, a prototype data governance framework can be created based on the existing principles, guidelines, and practices discovered, as well as a base framework from Janssen et al. (2020) that is aligned to an AI-integrated system that relies on BD. The prototype framework will concentrate on each component of an AI-integrated CRM that is crucial in terms of privacy and security.

A survey of 32 PrexPartners participants, including potential 'beta' users from other companies who are also the intended users of the AI-integrated CRM developed will be used to refine the prototype framework. The survey will be qualitative and quantitative, with in-depth descriptive questions and scalable questions such as those on a Likert scale. PrexPartner participants are divided into three groups: an expert development team, PrexPartners' intended end-users of the system, and external parties subject to the AI-integrated CRM system. External parties involved in creating the system will also be surveyed, depending on the system's development process at PrexPartners. The survey data will be analysed and used to fine-tune the prototype framework. Developing a final data governance framework that will be evaluated using the discussion and the proposed research questions.

3.2.3. Assessment

An evaluation will be made based on all the gathered deliverables, which can be separated into visualising an AI-integrated CRM system, a data governance framework and the collected data. These deliverables help answer whether the research questions have been satisfied and reflect on the overall output of deliverables. Aimed to open a discussion in the field of data governance

pertaining to an AI-integrated CRM and further explore new frameworks or models that contribute to privacy and security.

3.2.3.1. Process of an AI-integrated CRM

Since PrexPartners is in the works of developing an AI-integrated CRM, a high-level process flow can be described through its development. Together with the gathered information from the literature review, each component can be described in detail and which are relevant to data governance. The visualisation of an AI-integrated CRM allows for its underlying premise of helping build the design of artefacts such as the prototype framework (Kennedy-Clark, 2013). Allowing for a deeper understanding and learning towards developing new theories for further studies in the future (Barab & Squire, 2004).

3.2.3.2. (Prototype) Data governance framework

Through literature review and the process of how an AI-integrated CRM works in general, all the components can be identified and described in detail, which allows for developing an initial prototype data governance framework. Afterwards, a survey will be held for the relevant parties, and the results will be analysed. The prototype framework will be refined from the gathered results, and a final version will be created. This research aims to create a data governance framework from the perspective of privacy and security when developing or using an AI-integrated CRM in business-related fields that are customer dependent. A visualisation approach will describe the relevant components as well as an in-depth explanation.

3.2.3.3. Collected data

The collected data will include the surveys from the 32 participants at PrexPartners and other relevant parties in the development phase or who are subject to the AI-integrated CRM, such as the customers of PrexPartners. The survey participants can be divided into an expert group, the users at PrexPartners and end-users of other companies. These divided categories are common characteristics when using design-based research (Kennedy-Clark, 2013). The deliverable will include both qualitative and quantitative data from the conducted survey. An in-depth analysis of its results will also be done.

4. Development Process

This chapter covers the development process at PrexPartners, where the study is conducted. The development structure will be described, from which the components and processes can be gathered. Specific examples and implementations are described, but a general approach is taken so that the development process can be applied to any AI-integrated CRM system.

4.1. Development Structure

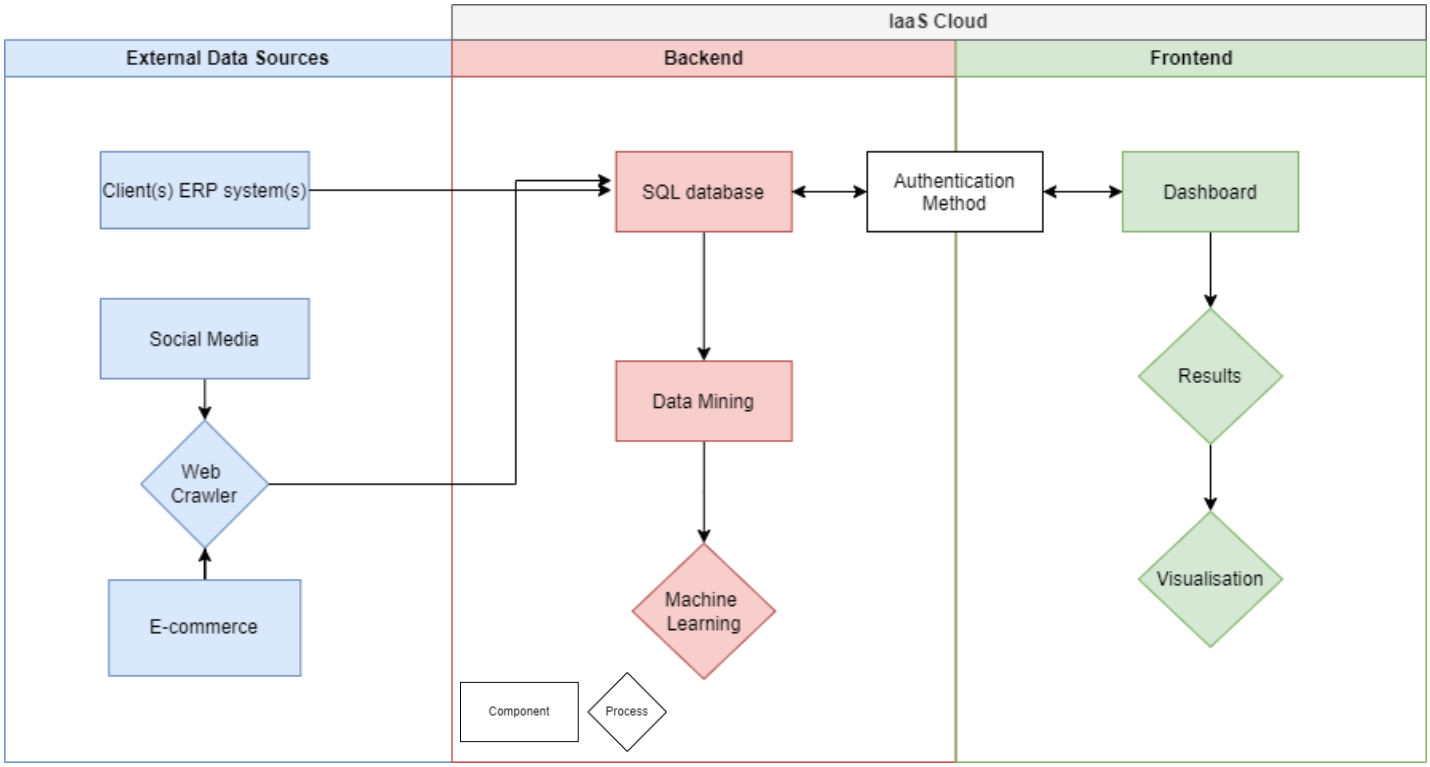


Figure 17: High-level Process Flow structure at PrexPartners

The infrastructure of the AI-integrated CRM system at PrexPartners is hosted on an IaaS provided by a CSP. Due to the many advantages, such as cost efficiency and being future proof due to the flexibility (Anshari et al., 2019). The development structure can be separated into external and internal components. From the client's data sources, such as the ERP system(s) that they are using, to the data scraped from the internet, such as social media. The gathered data will be stored in an SQL database from PrexPartners, and the process of data mining will take place through ML. Data mining results in the SQL database will be shown through a UI dashboard where users can interact and see a detailed overview of the processed data translated to a readable metric, such as graphs. Authentication methods (Indu et al., 2018) that the company provides are in place to either access the development platform in the IaaS or through the UI where users interact with the dashboard.

4.2. Data Sources

The primary data source comes from the client's ERP system(s). Depending on the setup of the client and the number of departments they have, there are multiple ERP systems present. SAP is one of the most commonly found; hence automatic data extraction is set up through the SQL database for that particular setup. In this case, the data results in sorted and categorised data (Yang et al., 2019) directly used for further processing. However, for the remaining ERP systems, a static version is often provided by the IT department of the client. The incoming data can vary from unstructured to semi-structured data, where there is often missing or incomplete information. Ranging from sensitive data pertaining to the client to regular data and prices from products that are important to the client. To gather a better insight, the quality data ***Figure 4: Quality data type pyramid from Zayah et al. (2012)***. can be used as an example. Hence, both periodic and per batch is often the case for data collection, in which the volume (Libai et al., 2020) often varies in petabytes due to the several years.

Social media is a valuable resource that can keep track of the most recent developments and is often gathered. A profile is made on the provided information from the clients through the ERP system(s), together with social media and U-commerce (Liu, 2015). Receiving up-to-date information regarding existing customers and prospects. Contact information together with the relevant connections and its description are collected. Relevance, such as the U-commerce platform (Liu, 2015) of the customer that can be linked back to social media are gathered and scanned of potential data that can be useful to fill in the blanks for specific missing ERP data. Depending on the client, E(U)-commerce analytics are provided, giving an oversight of which customers they are intended to target. All this raw data is being gathered through the SQL database, in which the provided data is given by consent through social media sharing (Anshari et al., 2019). The web crawler has been set up mainly automated but still requires some manual input to change the variables. Variables refer to the company name, certain connections that are well-known but not shown, etc.

Since PrexPartners is a Supply Chain & Procurement consultancy, an important aspect is to focus on bringing the best out of their clients through procuring the best deal with the suppliers. Hence, expressing the need to know the connections and the possible prospects from the focus industry. Therefore, the web crawler is also heavily used through U-commerce by finding the right supplier(s) or its product(s) through set keywords gathered from, e.g. social media. A good example is the Stock Keeping Unit (SKU) number, which can be used to track or identify the correct product online. Depending on the industry, the presence of E(U)-commerce can be limited and is not always easily findable through manual methods (Iqbal & Khan, 2021). The information is usually directly stored in the SQL database through the web crawler process. The gathered data contains both structured and non-structured data (Yang et al., 2019), which can range in the terabyte figures depending on the company, field, connections, etc. Data can only be determined to be valid until going through the ML process from data mining, thus needs to be stored for the moment.

4.3. Backend

All the data gathered through the mentioned sources are stored in an SQL database and can be visualised through the concept of columns and rows (Melton, 1996). The developers have

access through a web interface provided by the CSP. Microsoft Azure and Amazon AWS are often used due to their many advantages (Serrano et al., 2015) and their security practices and standards (Torkura et al., 2020). However, that alone is insufficient due to user errors (CSA, 2019). The structured, semi-structured and non-structured data is gathered and stored in the database per data source. Starting with defining the relevant classes in the field and understanding the business domain by set requirements. Defining classes allows for pre-processing in data mining after identifying the primary customer (Bahara & Sudheep, 2015), building a detailed customer profile, keeping track of the activities, transactions, and relationships, and building a profile for prospect customers of the gathered relations. Data preparation takes place with the help of ML, such as removing duplications and linking missing information throughout the data sources. Which checks for the correctness and completeness of the data (Bahara & Sudheep, 2015) before model building can take place.

Once the data has been prepped, model building occurs through the mentioned ML algorithms (Vafeiadis et al., 2015). Since manual analysis and processing are not possible anymore due to the amount of data and the complexity of associations (Molinillo & Japutra, 2017). Model building can be separated into several phases that handle and process the data. Starting with classification, where ML not only creates the categories but also learns the meaning and function for further use. Supervised learning can be used as an example, where the output shows its prediction of the classifiers (Nasteski, 2017). Regression is used for analysing the connections between each created category, such as the decision tree example (Vafeiadis et al., 2015). Similar ML techniques, such as association and clustering, can analyse and create classifiers, eventually leading to forecasting. After the model building process, model evaluation allows decision making and predictions through the processed data. Multiple methods of displaying the processed data are created to make the data readable and straightforward for people to interpret. Graphs, summaries, patterns, realistic roadmaps and more are made as well as personalised recommendations towards each customer (Chatterjee, 2019).

4.4. Frontend

A custom API for the AI-integrated CRM is built to interact with the UI for the backend to visualise and display the results of the processed data. Through the environment of the CSP, a dashboard created by the developers is available through a web interface for the end-users to interact with the AI-integrated CRM system. The dashboard contains several results and visualisations through either Microsoft PowerBi or Statistical Analysis System (SAS), which is software that specialises in data visualisation, allowing for Business Intelligence (BI). BI can be described as the end-users, in this case, making intelligent and informed decisions regarding the best outcome for the organisation (Foley & Guillemette, 2010). The help of ML, as pointed out in the advantages in the previous section(s), can lead to a more efficient work environment for its processes and create a competitive advantage over the competition (Rana et al., 2021). An example page from the interactive dashboard is displayed with dummy data. Ranging from the processed data of the customer's ERP data, such as their revenue and products, to the social media connections on other pages.

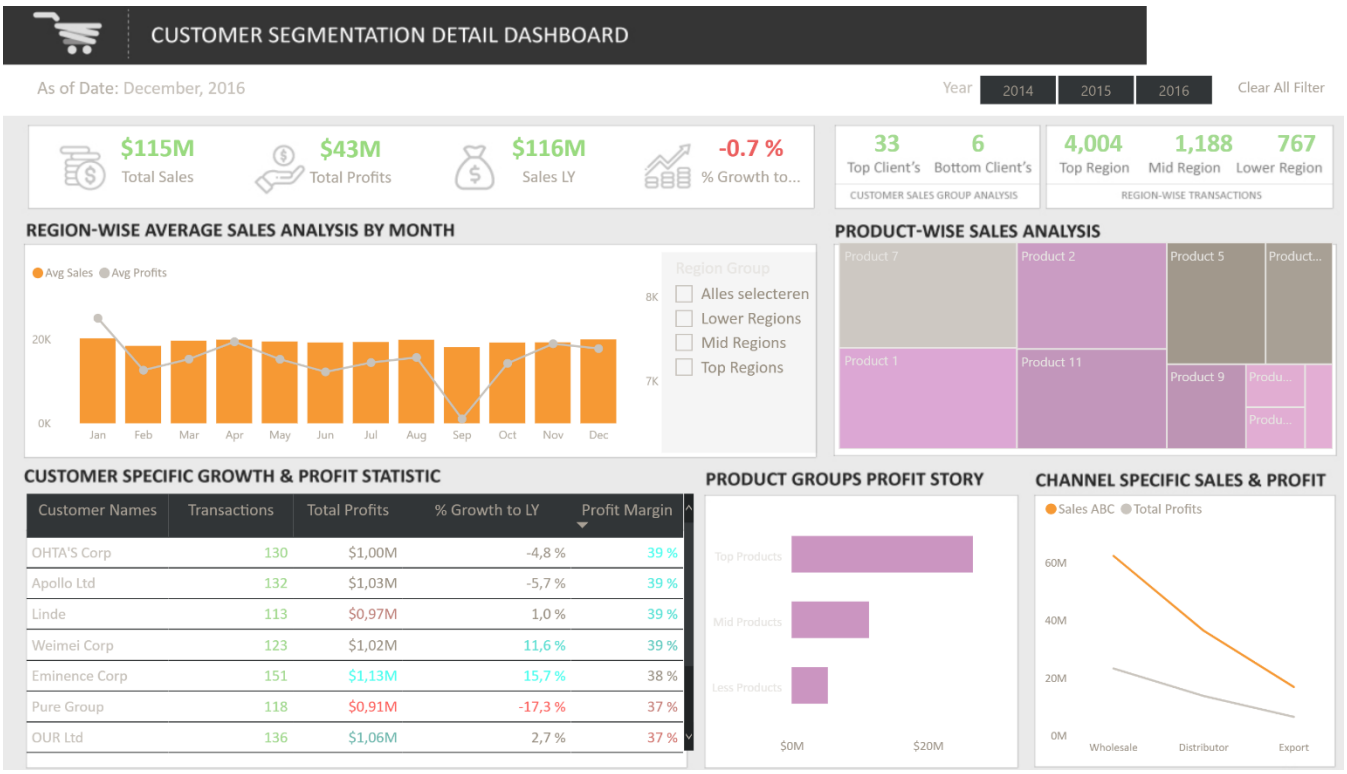


Figure 18: Dashboard example

The dashboard shows a quick overview of the vital data required for an in-depth analysis. However, multiple layers exist in the dashboard, allowing to pinpoint where the exact data comes from through the visuals. A specific individual or organisation can be identified and its transactional history (Zayah et al., 2012). Depending on the industry and which data is provided from the ERP system of the customer, individual customer and even employee data is also included with their personal information. Addresses phone numbers, social media and other sensitive data that can constitute a serious risk to the privacy and security as to the identifiable nature (Mansour, 2016). The critical difference is that due to the AI-integrated CRM system, the mixed data through its various sources are now usable through sorting and linking towards each individual or organisation and can be used for any purpose, including abuse (Mikalef et al., 2019). Opposed to the raw data that varies in size and is usually a large amount of unusable data that cannot be analysed manually, even with regular database tools (Anshari et al., 2019).

4.5. Data handling

The written consent and authorisation approval are acquired before the ERP data can be extracted and used from the organisation analysed. No written consent approval is acquired for the web crawler due to the ambiguity and a grey area of easily accessible data, but ethical practices are followed (Brewer et al., 2021). The data sources are primarily unstructured raw data dumps or gathered unstructured data; hence no level of data sensitivity is given. The same applies to when the data mining process has made identifiable categories and sensitive data can be seen, authorities enforce no specific set requirements due to the fast-paced technological advances (Raab & Szekely, 2017); yet organisations themselves need to establish standards and policies for data usage, processes, management and development that align with, e.g. GDPR (Ogbuke et al., 2020). In this case, all employees from PrexPartners can access the AI-

integrated CRM dashboard without special privilege, where sensitive data can be used throughout the company. Which is applied throughout multiple industries and its organisation to have similar ACP, as data stewardship is also the responsibility of the organisation in question to implement and limit privilege within the structure of command (Rosenbaum, 2010).

Compromises come from the ease of use of over limiting user access, such as when needing access to specific data to accomplish a task (Saltzer, 1974). Leaving the organisation vulnerable to the legal aspect, poor data quality, and abuse of sensitive information (Janssen et al. 2020). A standard login portal is provided for authentication of the employee in question that uses the dashboard and logs, which are enabled to retrace when the user has logged in or out. Standard users have no access to the backend; only the product owner is provided with credentials. Similar applies to the developers who can access the AI-integrated-CRM system's backend and the dashboard for testing purposes. A signed non-disclosure agreement (NDA) is requested from third-party developers as well as in-house developers in regards to the data as well as the tool itself. The customer that is being analysed is also provided login credentials to access a limited interactive dashboard version, which displays all the data similarly to the shown dashboard example.

5. Data Governance Framework

This chapter describes the developed data governance framework into three layers. Describing each layer in detail and pictured with a graph representing one layer. The third layer describes a specific use-case scenario on a system-level design. The described layers work in conjunction with each other to form the data governance framework. Additional best principles to the data governance framework are also described.

5.1. Regulations & Standards

Systems similar to an AI-integrated CRM are regulated under Data Protection Authorities (DPA) enforced in Europe, such as the GDPR, where countries may have their specific authority (Raab & Szekely, 2017). Due to the fast pace of new technologies, challenges and limitations often arise in the DPA. Problems such as comprehending and adapting to the technological field are often flawed, resulting in a lack of competence and unconcise guidelines (Giurgi & Larsen, 2016). Hence, having no specific authoritative requirements that can be followed per system results in organisations needing to adapt their policies to protect customers' data and their own (Ogbuke et al., 2020). However, general guidelines set by the DPA that apply to the AI-integrated CRM system still need to be adhered to, as well as the Freedom of Information (FOI) treaty (Janssen et al., 2020). Having established the first layer of the data governance framework, Janssen et al. (2020) describes the importance of the social norms, values and expectations that go along with the regulations. Bringing up the ethical aspect of data regulation to maintain quality data.

Equality and unbiasedness allow for a trustworthy AI through quality data (Yang et al., 2019). Since certain parts still require human interaction to allow the system to work and end-decisions that are the user's responsibility. Initial setup, such as keywords, manual categories, and the overall data processing and storage, needs to be adhered to by legal and ethical aspects (Benfeldt et al., 2020). No bias toward race, gender and personal preference should interfere with any process (Janssen et al., 2020). Therefore professional norms should be set through data stewardship. Ensuring that information processing and sharing assure data quality through responsibility, security and integrity (Dawes, 2010). Hence, specific standardisation organisations such as ISO exist for policy auditing and setting a standard for data governance. Specific ISO's have been created in order for an organisation to obtain certifications, such as ISO 27001, 38500, 8000, etc. In addition, security policies exist to prevent outside sources of access and data abuse; examples are displayed in ***Figure 14: Cloud Service security taxonomy from Indu et al. (2018)***. Through gathered literature review, as well as incorporating the framework of Janssen et al. (2020) as a basis for the data governance framework, the first layer can be visualised.

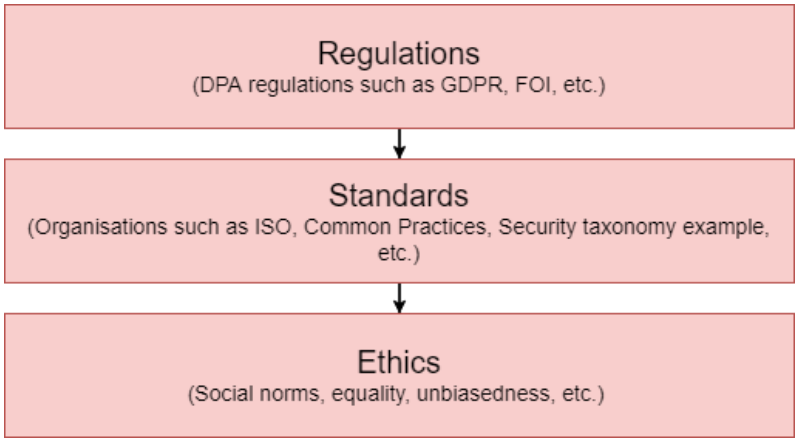


Figure 19: General Data Governance Framework –Layer 1

5.2. Organisation Setup

Having regulations and standards is the first step toward data governance in such a system. However, responsibility and reporting are usually not highlighted in this phase (Mullon & Ngoepe, 2019). The organisational structure can be assigned and designed in a top-level manner through data stewardship. Defining the authority within the organisation that is consistent with the mentioned standards. Recommending decision-making structures in the areas of relevance, in this case, AI, BD, privacy & security, through given roles within the organisation (Mullon & Ngoepe, 2019). Defining roles, such as Chief Privacy Officer (CPO), Chief Security Officer (CSO), Chief Data Officer (CDO), Chief Ethics Officer, Chief Information Officer (CIO) and similar roles within data governance that are the first point of accountability (Janssen et al. 2020). Allowing for data stewardship to take place, which in place allows for data governance throughout the organisation (Rosenbaum, 2010). Accountability elements, as shown in ***Figure 6: Unauthorized access UI design model through user accountability from Vance et al. (2012)*** can also hold the data steward accountable. With these roles come more responsibility and additional costs; hence, planning and control are of the essence, which is also often implemented in the IT-governance framework (De Haes et al., 2013).

Planning and control is often an annual cycle, which includes setting objectives, project scope, budget allocation and following up after the project (Janssen et al., 2020). Departments must often compete for budget allocation and resources for executing their activities. Depending on the size of the company and available resources, priorities must be set. That does not mean that data governance is not essential, as business and technology goals align with their performance (Janssen et al., 2020). An example, in this case, can be maintaining quality data, which will help maintain the AI-integrated CRM system to run correctly and give accurate analysis. Data governance in this approach is carried out through planning and control, by set roles and responsibilities and can be repeated, verified and audited. (Janssen & van der Voort, 2016). Continuous monitoring allows for less needed resource allocation because of the ongoing basis of maintaining quality data. The ongoing recommended security practices will allow for fewer upkeep costs due to already having the infrastructure (Xiong et al., 2019). Maintaining data needs to be within regulations; hence appropriate action plans need to be in place as a precaution for misuse and abuse or at the request of the entity in question (Kroll, 2018).

Due to the rising concern of GDPR in AI (Mogaji et al., 2020), a risk-based approach is often used to identify risks in a BDAS and implement appropriate mitigation tactics (Ladley, 2019). This approach is often the foundation of data governance (Rothstein et al., 2013) and is an effective solution to human error, system faults or influences of outside sources (Janssen & Kuk, 2016). Examples of mitigation efforts through threats can be seen in ***Table 4: Threats and controls from Tariq & Santarcangelo (2016)***. Taking into account a risk-based approach prevents the possibility of the potential misuse of any collected data. Regulations and standards play a governing role in creating practices for this approach. Such as, when an entity requests its information to be deleted, the request needs to be executed by law. Likewise, ISO standards, e.g. ISO 27001, allow practices to follow GDPR and prevent malicious online attacks (Maduka et al., 2017). Since a risk-based approach can be dependent per situation, novel ideas that have been proven efficient need to be followed to keep up with the growing risk developments (Raab & Szekely, 2017). Attempts such as obfuscating accurate data or random anonymisation toward the end-users (Potiguara et al., 2020) without compromising the results are already being implemented.

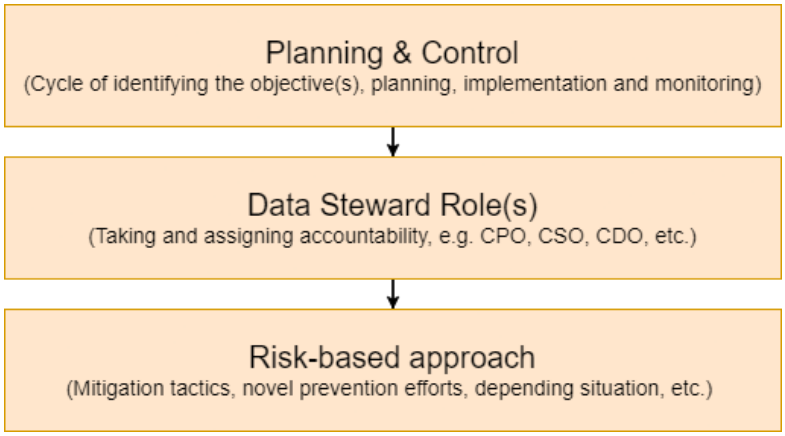


Figure 20: Organisation Structure Data Governance Framework – Layer 2

5.3. AI-integrated CRM system

When developing an AI-integrated CRM system, the platform relies on either an IaaS or PaaS CSP to provide resources (Serrano et al., 2015). However, there is a difference in approach regarding governance when comparing an IaaS and PaaS setup. In an IaaS CSP, the responsibility for maintaining and keeping up-to-date resources such as servers is of the requested entity (Maduka et al., 2017), in this case, an organisation. In a PaaS CSP setup, the responsibility is partially shifted towards the CSP. Due to that, they provide services such as tools and libraries, which allow for rapid development and testing ground, but managing the environment is the responsibility of the CSP (Rani & Ranjan, 2014). Compared to the traditional method (on-premise) in gathering the resources and setting up the development environment, IaaS has the same principle without additional disadvantages such as high costs (Serrano et al., 2015). Since the data governance framework is aimed toward a general implementation, both setups are discussed, and governance methods are adapted. Starting with choosing the right CSP that follows the recommended and sometimes even required standards (in some countries), such as the ISO 27000 family, that is the foremost standard for digital information security (Maduka et al., 2017).

The literature review also mentions the encompassing data governance for organisations such as ISO 38505. However, with the growth of new BD focused organisations (Shahbaz et al.,2020), the need of adapting continuously is highlighted as well as for standardisation organisations, such as the Federal Information Security Modernization Act (FISMA), Service Organisation Control (SOC), CSA, International Standard for Assurance Engagements (ISAE), etc. Mentioned in *Table 5: CSP security Compliance from Tariq & Santarcangelo (2016)* and CSPs that are well known. Linking layer 1 of the data governance framework, regulations and standards must be checked before choosing the right CSP and keeping up-to-date, whether an IaaS, PaaS or on-premise managed and hosted structure. The needed hardware for AI-integrated CRM differs from the setup. However, the bare minimum encompasses an SQL database, a data mining host (e.g. a virtual machine) for ML and a server hosting the UI. The ERP-system(s) from the client(s) is an external component provided by the client and only needs to be connected to the SQL database through an API or manual data extraction. The additional crawler in ***Figure 17: High-level Process Flow structure at PrexPartners*** is an internal component and is company-specific to their needs, hence not included in the data governance framework specifically.

In the ideal scenario, all processes should adhere to the first two described data governance layers, e.g. AI-integrated CRM system that adheres to all the regulations, standards, etc. However, what would happen is ambiguous data ownership, wasted resources due to constant data monitoring, and overly strict standards and compliances, in general, resulting in too much data governance and leading to an inefficient system (Janssen et al., 2020). The same applies when training ML through, e.g. supervised learning (Nasteski, 2017). Faulty forecasts and improper system functioning can occur due to improper data training (Janssen et al., 2020). Hence, more specific implementations of data governance towards the working of an AI-integrated CRM system can help eliminate the ambiguity. Starting with the data sources that can be structured, semi-structured or non-structured in real-time, periodic or per batch (Yang et al., 2019). Transferring data to an SQL database is often an automated process, but that depends on the setup and if resources are allocated to automation. Hence, manually transferring data needs to be executed by a data steward or supervised to preserve quality data and prevent data bias. This can be generally be applied to all scenarios where manual intervention is required when data is involved.

The process of data mining involves cleaning the data, such as removing duplicates, checking the quality of gathered data, etc. (Bahara & Sudheep, 2015) and is often an automated process due to the amount of gathered data (Molinillo & Japutra, 2017). However, data still needs to be manually checked if pattern changes have been detected or discrepancies due to system errors, such as from the algorithms (Janssen & Kuk, 2016). Thus, there is a need for a data steward when manual processing or handling of the data occurs. Once the data has been pre-processed, the ML algorithms allow data analysis through model building. When training the ML algorithms, inconsistencies and bias often occur due to historical data or incorrect training of unforeseen data (Janssen et al., 2020). Accurate data is therefore used and will need to be supervised or done by a data steward. A differentiation needs to be made between input and decision in data stewardship. The engagement towards ML algorithms from human intervention should be able to explain the causality of the data keywords, set rules and not be personally opinionated (Janssen et al., 2020) and can be validated against the outcomes of ML.

The importance of unaltered data cannot be overstated, as any change caused by human intervention can result in an incorrect outcome (Nasteski, 2017). New results should be compared to the trained sample data results for validation and potential pattern changes (Janssens et al., 2020). This is possible via the UI of the AI-integrated CRM system. Either through a web platform or a dedicated application provided by the IaaS or PaaS structure. Access control is critical because the user interface (UI) makes the system usable and provides an easy way to use or modify the results. As a result, by combining layer one standards such as authentication mechanisms (Indu et al., 2018), unauthorised access can be prevented, as can layer two and data stewardship roles. Reducing the risk of data abuse and preventing unauthorised access to a sensitive BD system (Mansour, 2016). Not only sensitive data but also the results and forecasts of the ML algorithms are displayed. Each organisation that uses an AI-integrated CRM system may use different algorithms and thus should be able to be scrutinised by auditors to avoid public concern as well as legal and ethical questions (Janssen et al., 2020).

Even if the inner workings of an AI-integrated CRM system do not need to be understood by the general public, scrutinising such a system allows for transparency (Janssen et al., 2020). Avoiding ethical backlash from the community and increasing system security through increased vulnerability inspection (Malhotra & Malhotra, 2010). Throughout the process, the results of the ML algorithms should be able to be verified and validated. To avoid improper functioning, changes, irregular patterns, and behaviour must be closely monitored. In the event of a failure, improper training, or other factors, any BDAS may produce inaccurate results (Janssens et al., 2020). As a result, open communication should be considered and common sense when implementing governance. More control in the data governance process results in a better working system, but as previously stated, too much governance can result in an improperly functioning BD system (Janssen et al., 2020). Organisations may impose such policies, structures, and so on; the key is to strike the right balance. Each organisation is unique, and there is no one-size-fits-all solution. As a result, data governance frameworks such as those proposed in this section can assist organisations in laying the groundwork (Rana et al., 2021).

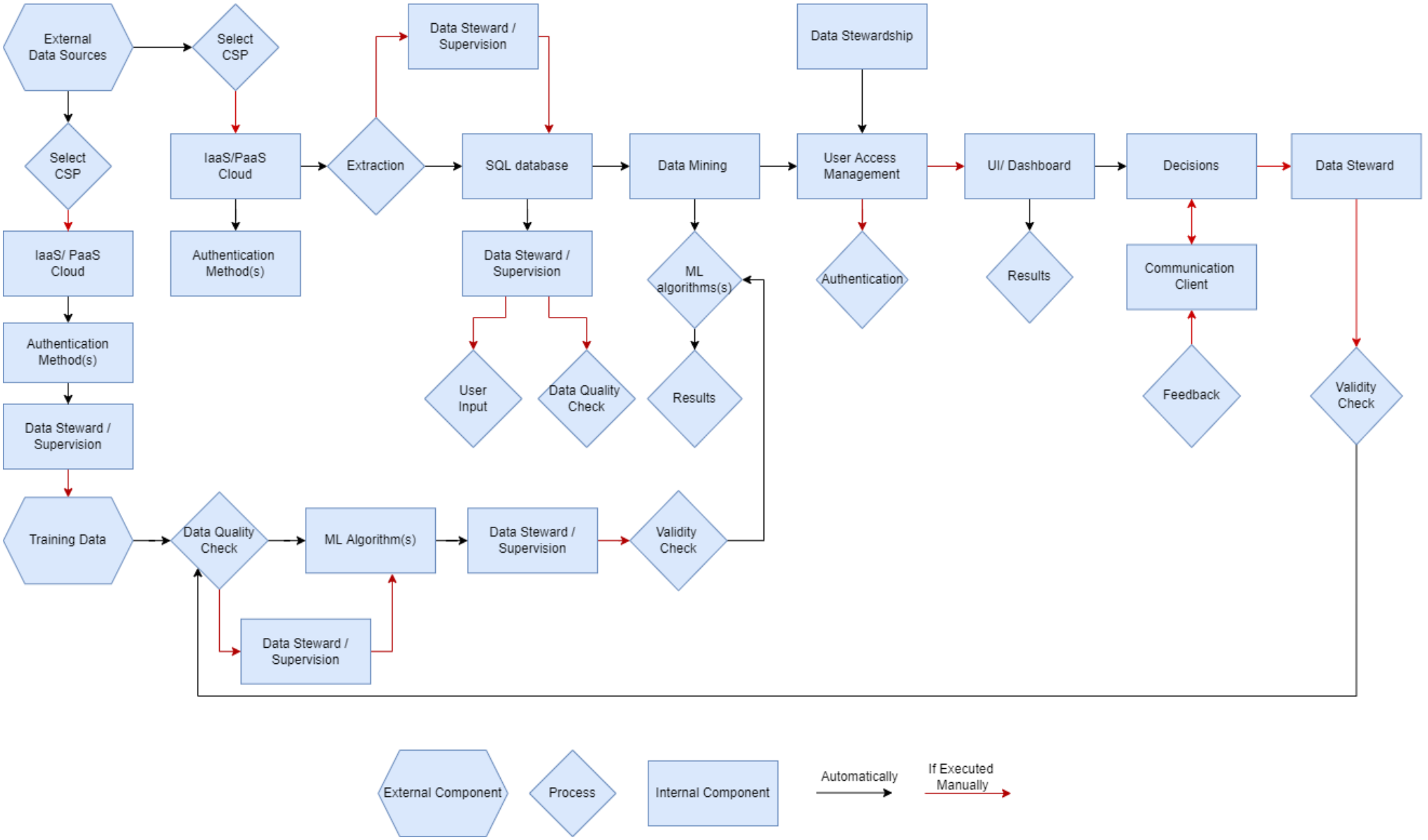


Figure 21: System-level Data Governance Framework – Layer 3

5.4. Data Governance Principles

Although three layers of the data governance framework have been described, certain aspects may still be overlooked due to the complexity of an AI-integrated CRM system (Chatterjee et al., 2019). Achieving sound data governance for a BDAS remains challenging due to a lack of adoption of acceptable practices, a lack of research on a trusted framework, and a lack of consensus (Janssen et al., 2020). When implementing such frameworks, care must be taken to avoid privacy violations, information misuse, discrimination, etc. (Janssen & Kuk, 2016). Data governance promotes accountability, fairness, increased trust, transparency, and eliminating personal inputs that result in discrimination (Al-Badi et al., 2018). Because general applications such as those mentioned in layers 1 and 2 may not be sufficient, a system-level approach may be required to resolve ambiguity. In addition, while each organisation may have a different setup, the foundation remains the same. Additional principles that go along with the mentioned layers allow for more sound data governance in an AI-integrated CRM system and contribute to the scarce research (Jain et al., 2016). (Janssen et al., 2020).

Principle	Description
1. Data quality and validation check	The used data should be checked for bias, inaccuracy or anything that might alter the original content of its data and can be validated (Janssen et al., 2020).
2. Algorithm check	Pattern inconsistencies should be monitored and corrected (Kroll, 2018).
3. Data handling	Data separation of sensitive data if possible, implementing methods such as data obfuscation (Potiguara et al., 2020).
4. Vulnerability analysis	Allow for auditing of the system, programs such as bug bounty (Janssen et al., 2020)
5. Ethics	What might be legal may not be well perceived by social norms such as web crawling (Brewer et al., 2021).
6. Data stewardship	Formal accountability, setting up a structure of responsibility (Dawes, 2010).
7. Authorization	Access should only be given to the relevant users (Janssen et al., 2020).
8. Consent	When sharing data, prior consent needs to be given (Cuganesan et al., 2017)
9. Multi-factor authentication	Up-to-date security measures should be set, preventing abuse of inside and outside factors (Xiong et al., 2019).
10. Data retention	Data that is not being used or not beneficial for the aimed purposes should be safely discarded (Dawes, 2010). Discard information according to regulation (Ogbuke et al., 2020).
11. Data sharing	Data should only be shared with the concerned parties, as well as relevant parties need to be informed of shared usage (Janssen et al., 2020).
12. Feedback	Concerned people should be able to voice their input and adapt accordingly.
13. Delegation	One person should not be able to abuse the system or data, shared responsibility distribution (Janssen et al., 2020).
14. Process Automation	Processes that can be automated to eliminate user input or bias, resource dependant.

Table 6: Data Governance Framework Principles

6. Results

This chapter describes the participants who will be taking the survey and the dimensions that have been used for the survey. The survey can be divided into both open questions and a rating scale, where the values of the rating can be defined and described. The collected results will be shown in a graph with additional gathered feedback from the respondents.

6.1. User Group

The survey will be sent out to 32 people at PrexPartners and three external participants who either use or build the AI-integrated CRM system. The distinction can be made between people with a technology background and people with a business administration related background. Which can indicate the level of comprehension a person might have prior to showing the proposed data governance framework and involving different departments for a proper assessment (Marchildon et al., 2018). The group can be further divided by their function and can be separated into the following:

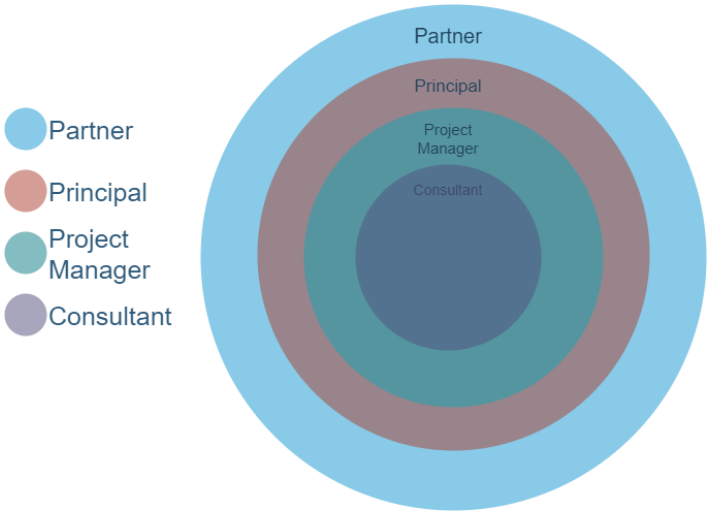


Figure 22: Company structure at PrexPartners

Partners have full access to any data gathered or collected when allowed by the organisation they are consulting. Data access is automatically granted in the data storage when requested from the principals, project managers and consultants. Principals allow for making connections between the organisation and customers, having full access to the information once requested from the project managers and consultants. Project managers are restricted to the projects' data that they are working on, and other data access is based on request or when involved in that specific project. Consultants are granted access to certain data based on the tasks required to perform. There is no restriction per project and can often include multiple datasets from multiple active projects. More data access is granted based on performed tasks or by request.

6.2. Survey

The survey will contain both questions on a rating scale, such as rated and open questions that are based on the research of developing a sound data governance framework, where the specific fields such as Cloud, ML, BD, etc. are taken into account (Khatri & Brown, 2010; Al-Ruithe et al.,2016; Weber et al., 2009; Marchildon et al., 2018, Peffers et al., 2007). A sound data governance framework structure is already partially present due to the used foundation by Janssen et al. (2020). The survey is made through Google Forms and sent out accordingly through the correspondents emails. They will either be directly sent out to the participant or be sent out by the contact that has the person's information. The survey is based on the provided dimensions below that can be categorized in several areas of data governance. Forty questions are numbered throughout the survey. 11 dimensions have been chosen and are listed as follows, which include on average, 3 questions per dimension:

- **Data Risk Management and Compliance**

Aspects that take into account mitigation tactics through control and planning. Methods are in place that can avoid, accept, qualify, quantify and identify risks (Marchildon et al., 2018). Ensuring that all processes have a failsafe in case of the worst possible outcome to an organisation is based on compliance from the authorities relevant to the country and its ordinances, such as GDPR. Similar to Industry standards such as ISO (Al-Badi et al., 2018).

- **Data Value Creation**

Data can be qualified and quantified in an organisation to use for its maximum potential (Marchildon et al., 2018). Where value can be created for multiple domains in the business industry and provides a benefit towards an organisation (Soares, 2010). Ranging from the stakeholders to the users that are the subject of the data in question.

- **Data Organisational Structure and Awareness**

The mutual accountability between different departments, such as the IT and organisational departments. The responsibility and awareness of data governance at different company levels of management are assessed (Marchildon et al., 2018). Data stewardship is closely related, where the functions can be defined, and the responsible party of the data in question can be identified. The quality of internal control, such as compliance reports, can be measured and documented within the organisation's set required standards and rules.

- **Data Policies and Rules**

Establishing practical data principles where the business use of data is defined through specific set policies, guidelines, and appropriate standards. Set policies are guidelines that can be adopted but are not necessarily applied in specific situations (Marchildon et al., 2018). It can be referred to as a rule of thumb, taking into account common sense and the possibility of adaption of specified policies and rules (Soares, 2010).

- **Data Stewardship**

Ensuring custodial care of data assets through control, risk mitigation and risk control (Marchildon et al., 2018). By enabling points of responsibility and access control through the appointed person(s). Limiting unauthorised access to data and overprivileged users that do not

need access to specific data. Preserving data quality and limiting the abuse and misuse of data that might occur in the event of no data supervisor (Soares, 2010; Marchildon et al., 2018).

- **Data Quality Management**

Referring to the usage requirements of the collected data and its ability to fulfil the satisfaction. Data quality has multiple dimensions and can range from accuracy to completeness; thus needs to be defined in how the data is being used (Soares, 2010). Accuracy is aligned with the correctness of the data, which can either be processed through an AI-integrated CRM system or prior to. Completeness refers to that no data is missing, such as an adequate dept are some of the examples (Khatri & Brown, 2010; Marchildon et al., 2018).

- **Data Lifecycle Management**

Data can be moved through different stages in a life cycle and stand central to designing data governance. Changes in data can occur instantly, and records need to be up-to-date to be accurate (Soares, 2010). Understanding the importance of what data is present, criticalness, sources, redundancy, etc., can establish its values through metadata. Besides compliance and legislation related points, retention and archival of data needs to be addressed accordingly, such as data anonymisation (Khatri & Brown, 2010).

- **Data Privacy and Security**

An organisation has set policies, principles, practices, and controls limiting data assets' exposure and potential mitigating risks in either security or privacy (Marchildon et al., 2018). Business practices in confidentially, such as an NDA, are often used, and multiple authentication methods when accessing an AI-integrated CRM system. *Data access* can be defined as assigning a value or categorising data and making it available to the proper beneficiaries (Khatri & Brown, 2010). Ensuring the integrity, availability and confidentiality through risk assessment and monitoring efforts.

- **Data Architecture**

Refers to the design architecture of the data and the system, such as structured, semi-structured or unstructured. The need for such structure is also identified and defined on whether it is necessary. The availability and distribution of data to the appropriate user are highlighted (Marchildon et al., 2018). The necessary data architecture components are assessed based on the established standards of the organisation.

- **Data Classification and Metadata**

Data is arranged in specific ways, such as structured columns and can be identified through defined keywords (Marchildon et al., 2018). A data dictionary contains organisational terms and defines the value to be classified in an AI-integrated CRM system. Different business domains might use other terminology, or organisations might create their terms and keywords for internal use (Soares, 2010). Metadata describes the semantics or data characteristics for further interpretability (Khatri & Brown, 2010; Marchildon et al., 2018). A distinction between physical and domain-independent metadata can be made. Allowing for descriptive data that can be used on all levels and divisions in an organisation, ranging from application data to specific characteristics (Khatri & Brown, 2010).

- **Archiving Information audits and Reporting**

The organisational processes monitor and measure data governance's risk, effectiveness, and data value of data governance (Marchildon et al., 2018). This can include the risk of compliance, such as the GDPR of data retention and allowing the data to be requested and deleted by the instance in question. External assessors, such as auditing companies, can report a positive result on the established data governance practices within the organisation and can be executed when requested.

The survey will briefly introduce the topic at hand and display all the layers of the data governance framework with the mentioned principles, which can be accessed during the survey. There will be one survey, and various respondents' levels of expertise and knowledge will be present. The survey structure compares the current situation at the company and the situation with the proposed data governance framework. At the end of the survey, the confidence level is asked about all the questions and additional comments, concerns or feedback. This allows for an honest assessment in case questions have been misunderstood or how accurate the survey results reflect on the understanding of the proposed data governance framework (Marchildon et al., 2018) or data governance for an AI-integrated CRM system.

6.2.1. Dimensions

Several data governance frameworks have been proposed over time, an example shown in ***Table 2: BD governance frameworks from Al-Badi et al. (2018)***. Resulting in several decision areas that attempt to distinguish a good data governance framework from a bad one (Khatri & Brown, 2010). To this day, the most elaborate listed decision areas and competencies in data governance have been proposed by Soares (2010). They are generally viewed as the most comprehensive for assessing a data governance framework, as they overlap most existing assessment frameworks (Marchildon et al., 2018). Hence, specified decision areas will be based on the most extensive proposed competencies and Industry associations, such as the CSA, ISO, etc. The Data Governance Institute (Thomas, 2006) has created guidelines and frameworks for a general data governance approach, articulating six decision areas (Al-Ruithe et al., 2016). Including an adapted version of the framework from Khatri & Brown (2010) that mentioned five decision areas. Thus, several sources will be used to create the specified competencies.

The mentioned frameworks, categories and competencies are not specific to implementing an AI-integrated CRM. Hence, a combination of the proposed frameworks for implementing a data governance framework can be categorised into several areas to question the proposed data governance framework for maturity. *Maturity* is when an organisation has developed (and deployed) the structures, processes, practices, and policies required to optimise the storage, collection, dissemination, and use of the organisational data assets (Marchildon et al., 2018). The basis of dimensions and their questions will be extracted from Marchildon et al. (2018), built around the decision areas by Soares (2010). Additional questions will be extracted and adapted from other mentioned literature (Khatri & Brown, 2010; Al-Ruithe et al.,2016; Weber et al., 2009; Peffers et al., 2007; Soares, 2010) to be explicitly applied to an AI-integrated CRM system in regards to security and privacy. Overlapping areas will be merged, and a limit to the questions may apply as not to overcomplicate and create a too exhaustive survey.

The extracted questions and respective scales, such as rating or open questions, are based on existing data governance frameworks and methodologies (Marchildon et al., 2018). Compiling the results can give a cumulative score for each of the mentioned data governance dimensions. According to the scoring based on the Capability Maturity Model (CMM), which has become widely accepted as a standard and has been applied to a wide range of identifying problem areas (Lasrado et al., 2015), such as the level of maturity, soundness, etc. that can be represented for specific dimensions on a scale of 5 (Marchildon et al., 2018):

1. Performed, where processes are carried out spontaneously with no prior planning. Processes are rarely implemented across the organisation's different departments and are foremost reactive.
2. Managed, where a policy guides the planning and execution of the organisation's processes. Stakeholders monitor and control the processes through oversight. Specific processes may not be applied in every aspect of the company. The appearance of data governance is present in the organisation with a data integration platform, and a concern for data quality is expressed.
3. Defined, where main processes follow the set of rules from the organisation consistently, and specific processes are adjusted according to these rules. The organisation defines a data governance framework, where the data management service captures the business rules.
4. Measured, where process metrics are used and created, and the performance of processes is managed across the organisation. A data governance framework and centre have been established, with an approach to data quality.
5. Enhanced, where the performance of the process is optimised, best practices are shared, and improvements are identified. Quantitative business process objectives have been in place and are being amended regularly to adjust for the changing business goals. The organisation is confident and adapts continuously to change. Data governance information is shared throughout all departments in complete transparency.

Specific questions will be open for further interpretation, whereas the correspondent has the opportunity to express their comments, concerns or feedback.

6.3. Survey Results

6.3.1.0. Participants

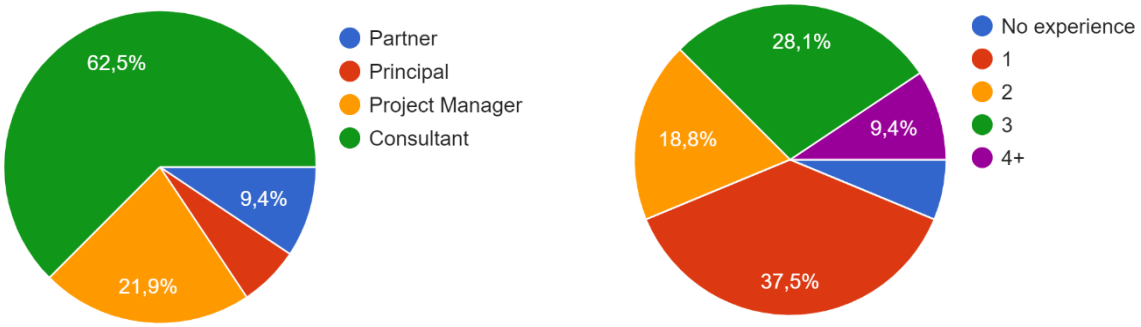


Figure 23: Survey Participants results

The majority of the participants that responded to this survey were 20 consultants. Followed up by seven project managers, three partners and two principals. Everyone has experience with similar tools like the AI-integrated CRM system (PrexDigital Analytics tool). With foremost people having at least one year of experience, ranging to 4+ years. Only 2 participants answered that they have no experience with a similar tool.

6.3.1.1. Data Risk Management and Compliance

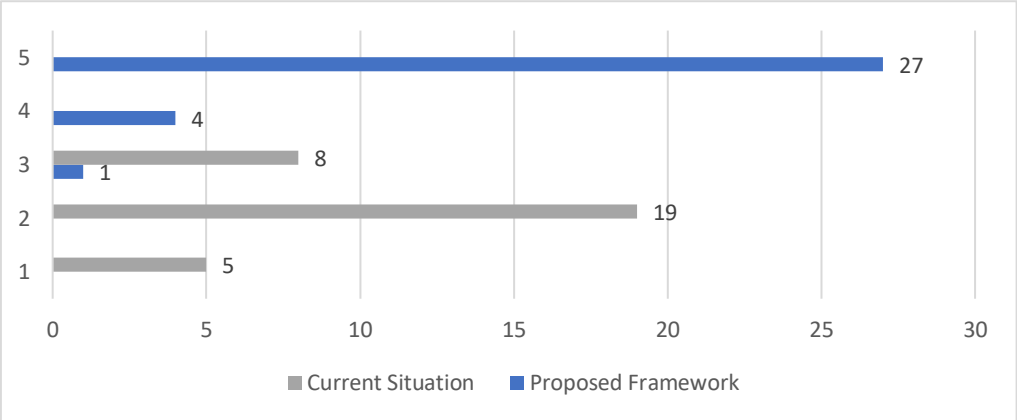


Figure 24: Data Risk Management and Compliance results

For the company's current situation, the majority of the respondents picked 2, with an average score of 2,09. When presented with the situation of using the proposed data governance framework, most of the respondents picked 5, with an average score of 4,81. Additional feedback was given that the company is using no existing data governance framework in the current situation. Although practices and guidelines are followed that abide by industry standards as well as from a legal aspect.

6.3.1.2. Data Value Creation

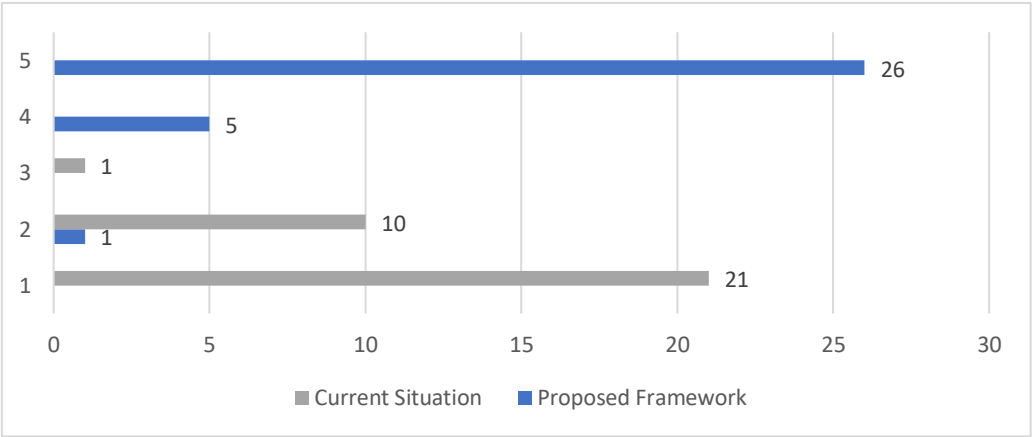
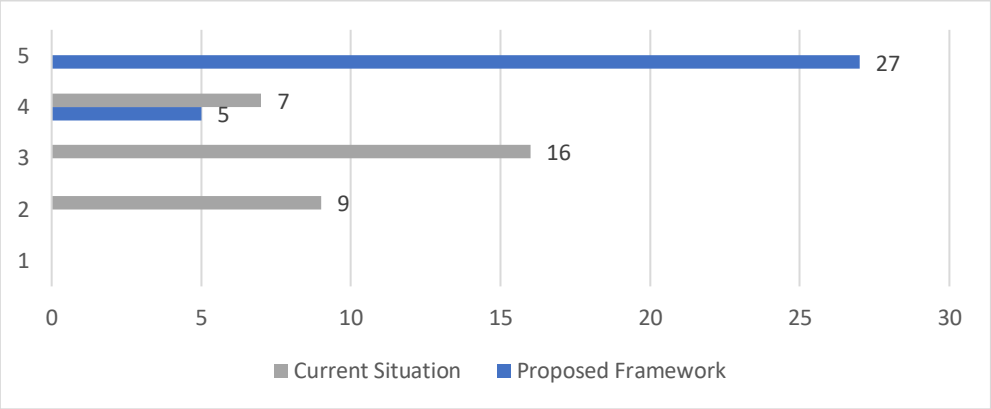


Figure 25: Data Value Creation results

For the company's current situation, the majority of the respondents picked 1, with an average score of 1,38. When presented with the situation of using the proposed data governance framework, the majority of the respondents picked 5, with an average score of 4,75. Additional feedback was given that there is currently a separation between the IT and business domains

involving seamless integration with the set data governance practices. A reference was made between departments operating in their own 'silos'.



6.3.1.3. Data Organisational Structure and Awareness

Figure 26: Data Organisational Structure and Awareness results

For the company's current situation, the majority of the respondents picked 3, with an average score of 2,94. When presented with the situation of using the proposed data governance framework, most respondents picked 5, with an average score of 4,84. Additional feedback was given that in the current situation, the senior management supports the set data governance practices, although there is no actual set structure (framework) that supports the chosen policies from the organisation.

6.3.1.4. Data Policies and Rules

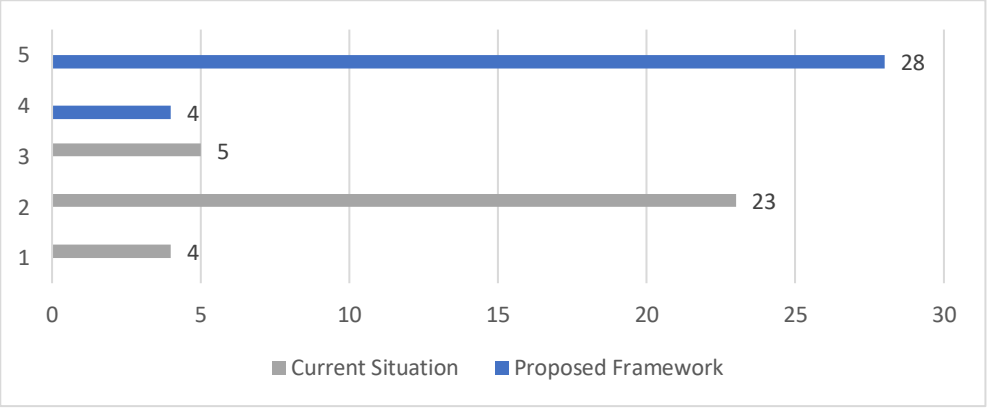
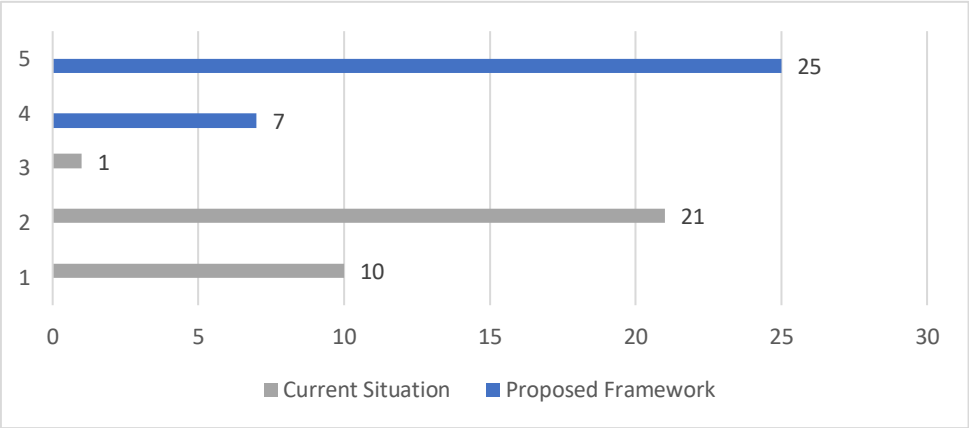


Figure 27: Data Policies and Rules results

For the company's current situation, the majority of the respondents picked 2, with an average score of 2,03. When presented with the situation of using the proposed data governance framework, most respondents picked 5, with an average score of 4,88. Additional feedback was given that if the proposed data governance framework were present, it would be used to strengthen the organisational practices and policies.



6.3.1.5. Data Stewardship

Figure 28: Data Stewardship results

For the company's current situation, the majority of the respondents picked 2, with an average score of 1,72. When presented with the situation of using the proposed data governance framework, the majority of the respondents picked 5, with an average score of 4,78. Additional feedback was given that general data stewardship is barely being implemented and being handled according to who is handling the project in the current situation.

6.3.1.6. Data Quality Management

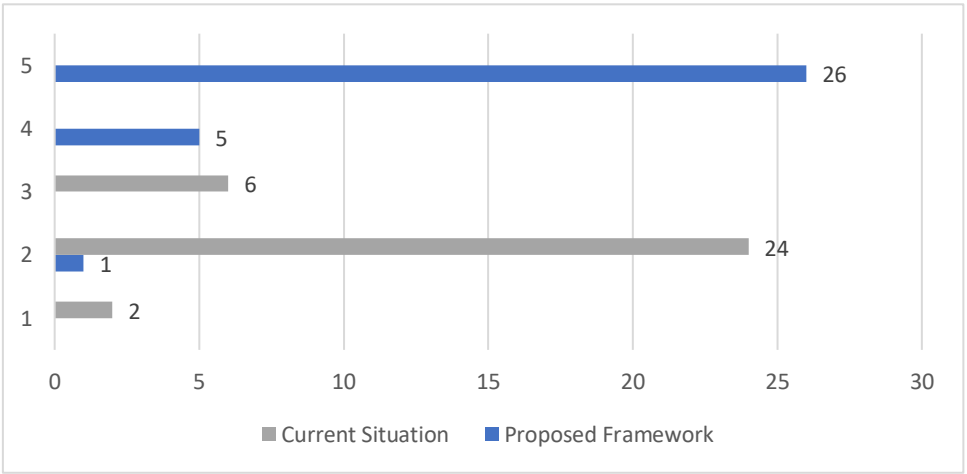


Figure 29: Data Quality Management results

For the company's current situation, the majority of the respondents picked 2, with an average score of 2,13. When presented with the situation of using the proposed data governance framework, the majority of the respondents picked 5, with an average score of 4,75. Additional feedback was given that data quality issues are not being documented in the current situation. Although, data quality issues are being handled accordingly through the right project manager.

6.3.1.7. Data Lifecycle Management

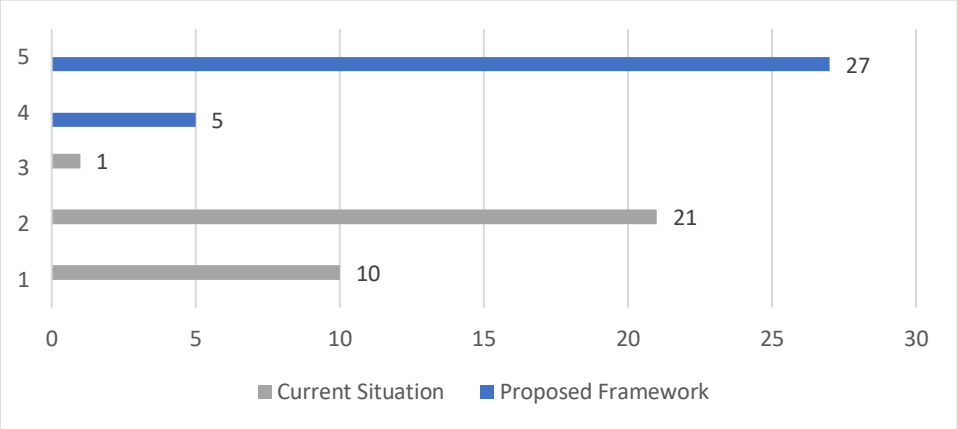


Figure 30: Data Lifecycle Management results

For the company's current situation, the majority of the respondents picked 2, with an average score of 1,72. When presented with the situation of using the proposed data governance framework, most respondents picked 5, with an average score of 4,84. Additional feedback was given that data is being stored and not being deleted in the current situation. There is no plan for data that has been gathered, although efforts have been made to make a centralised database.

6.3.1.8. Data Privacy and Security

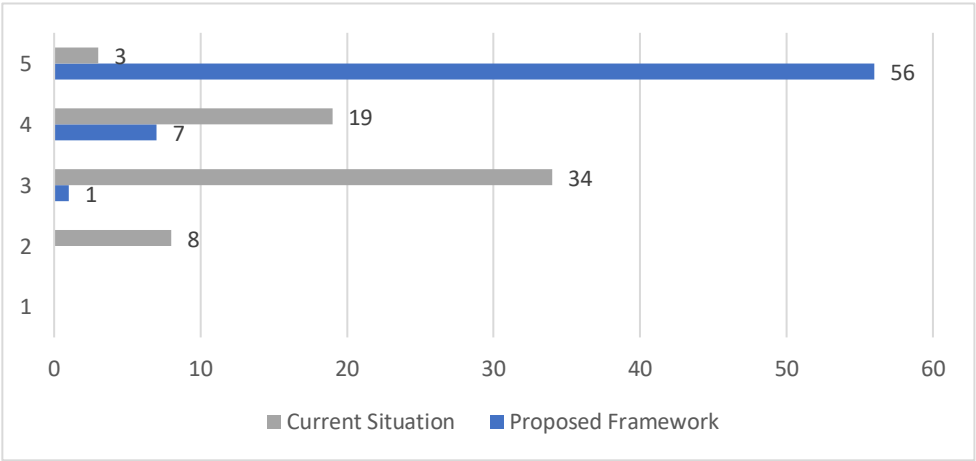


Figure 31: Data Security and Confidentiality results

The respondents were asked an additional question regarding the dimension of privacy and security. For the company's current situation, the majority of the respondents picked 3, with an average score of 3,27. When presented with the situation of using the proposed data governance framework, most respondents picked 5, with an average score of 4,86. Additional feedback was given that data security and security are dependent on the project in the current situation. Access is granted to the developers or involved parties training the AI with data test sets. Higher-ups in the organisation generally do have full access to all the information, regardless if they are involved in the project or not, just by requesting access.

6.3.1.9. Data Architecture

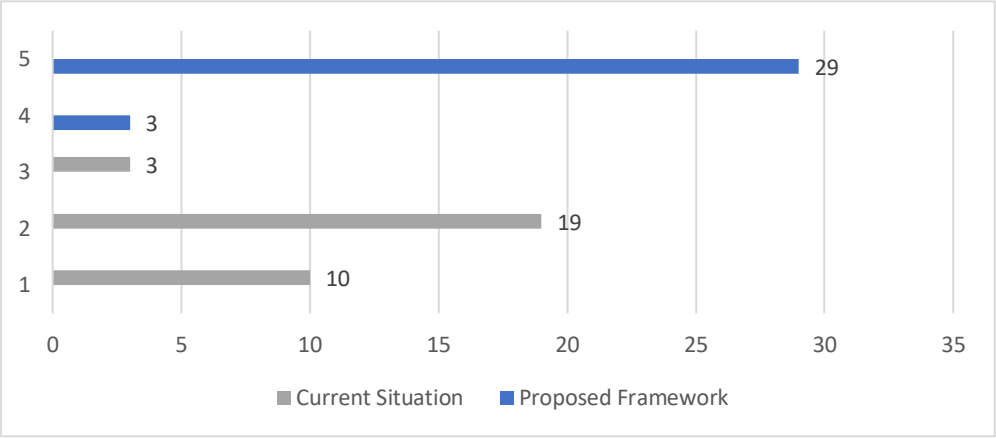


Figure 32: Data Architecture results

For the company's current situation, the majority of the respondents picked 2, with an average score of 1,78. When presented with the situation of using the proposed data governance framework, most respondents picked 5, with an average score of 4,9. Additional feedback was given that there is no existing data architecture as it depends per project due to different data sets gathered per project. Data is often structured when specific tasks need to be accomplished or particular insights need to be created.

6.3.1.10. Data Classification & Metadata

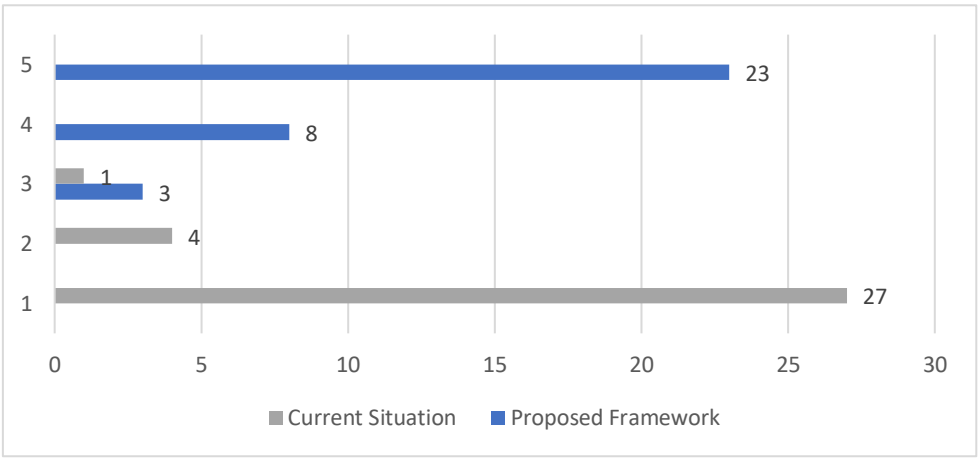


Figure 33: Data Classification & Metadata results

For the company's current situation, the majority of the respondents picked 1, with an average score of 1,19. When presented with the situation of using the proposed data governance framework, most respondents picked 5, with an average score of 4,88. Additional feedback was given that no metadata is stored, but often dependant per project that the person responsible knows the used or standard metadata information.

6.3.1.11. Archiving Information Audits and Reporting

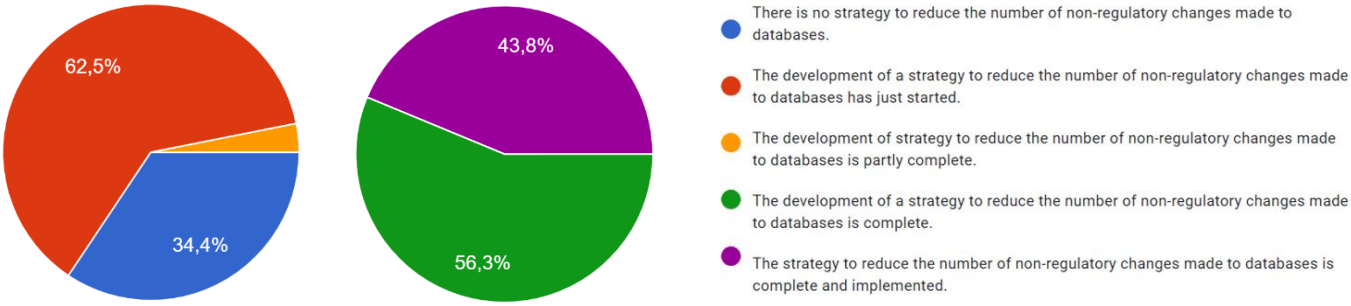


Figure 34: Archiving Information Audits and Reporting results

For the company's current situation, the majority of the respondents picked 2, with an average score of 1,69. When presented with the situation of using the proposed data governance framework, most of the respondents picked 4, with an average score of 4,44. Additional feedback was given that in the current situation, the development of the strategy to reduce the number of non-regulatory changes made to the database has started but is still in its early stages. Compared with the proposed framework, most parts of non-regulatory changes made to the database would be either complete or implemented due to the data governance framework already being present.

6.3.1.12. Additional Information

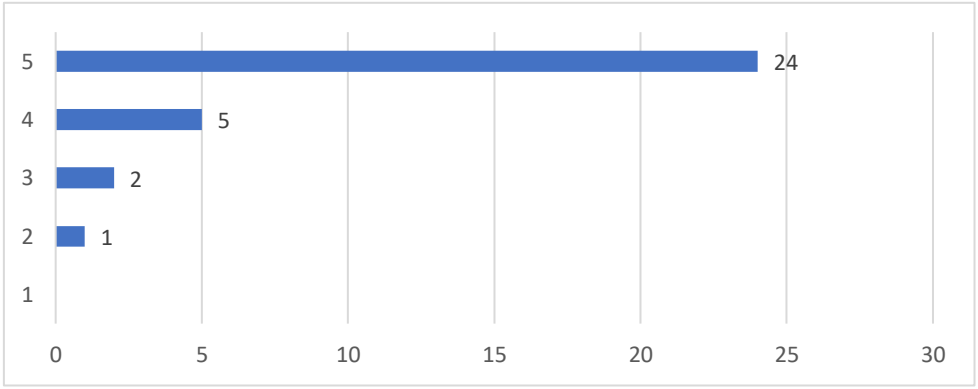


Figure 35: Confidence Level results

The confidence level was tested on how confident the respondents were when answering the survey questions, resulting in an average of 4,63. Additional feedback was given that the PrexDigital Analytics tool is being actively developed, and scoring of the current situation would not reflect the result. The respondents said the survey to be extensive but clear, and the presented data governance framework was, although very in-depth, understandable on how it would be used and implemented. Some parts might be hard to understand or imagine due to the limitations of how a person might perceive the implementation of such a data governance framework by a respondent.

7. Analysis & Discussion

This chapter considers the analysis of the results, their dimensions and the general improvements that can be used to refine the data governance framework. A flowchart with specific scenarios that will improve the privacy & security aspect is mentioned, and a discussion of reasoning for choosing the survey and dimensions.

7.1. Analysis Survey Results

7.1.1. Respondents

The survey sent out to 32 participants was filled out entirely for the questions on a scale, and responses were received within two weeks. Thirty respondents were from PrexPartners, and only two respondents were from an external organisation that helped develop the AI-integrated CRM system (PrexDigital Analytics). The survey was conducted throughout the whole organisation to get the most accurate results, as an expert group alone would create a bias in the results (Marchildon et al., 2018). Hence, all the organisation departments were involved, primarily consultants ranging from the business field to the IT department. The distinction between a more senior position in the company was made to distinguish how data governance might be perceived and created some outliers in the results. Overall, almost all respondents have experience working with an AI-integrated CRM system; as a result, it can be seen that the confidence level in answering the presented questions was high, with 4,63 out of 5. As a result of the comprehensive survey, results were shared with the respondents, which was appreciated. Overall, the survey was positively received, not having received any complaints, and relatively consistent results were gathered.

7.1.2. Overall Dimensions

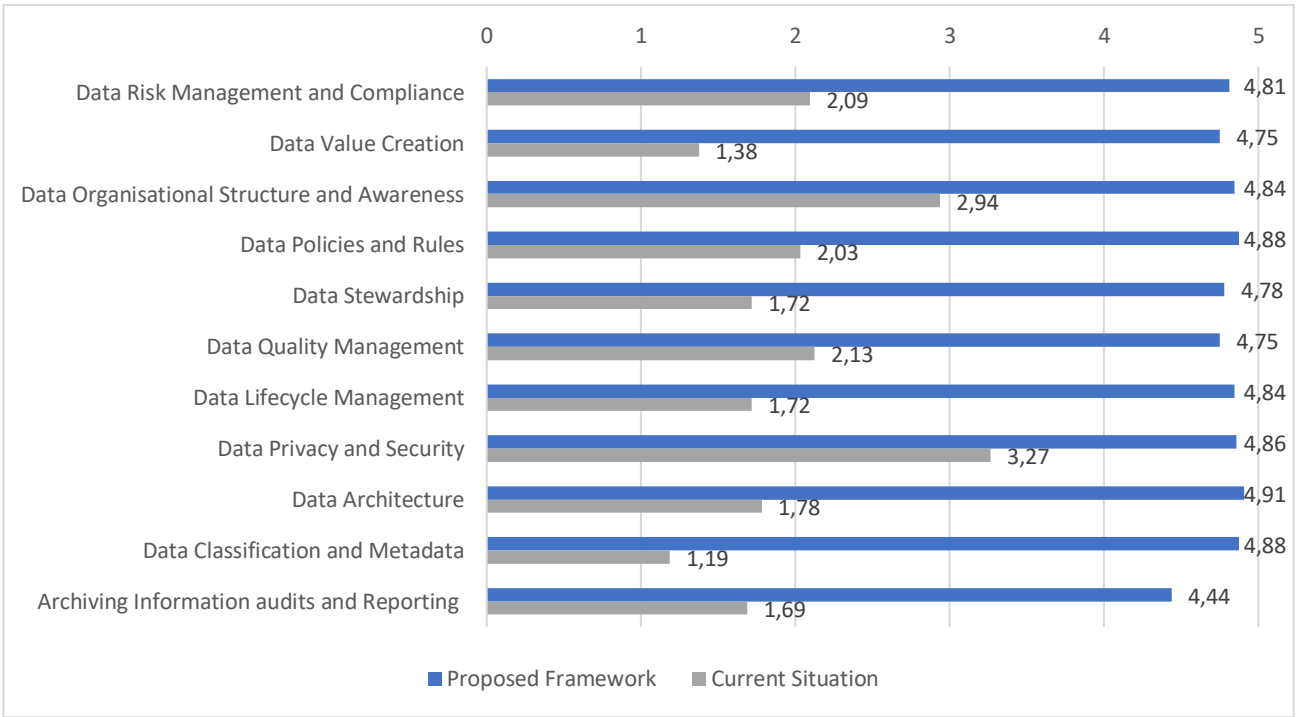


Figure 36: Summarised Results from Data Governance Survey

The overall score for the current situation at the organisation regarding the AI-integrated CRM system falls between the scoring range of 1,19 to 3,27, resulting in an average of 1,99, which according to the CMM, identifies the levels performed, managed and defined (Marchildon et al., 2018). The additional feedback provided with the dimensions pertaining to the current situation confirms that most processes are performed without prior planning and are different per project. Data governance is often not applied across all organisational areas and is primarily reactive. There may be existing approaches or strategies in data governance, but they are often not applied or limited. Processes or projects, in this case, are often controlled and monitored by the relevant stakeholders. The mentioned analysis can be applied to all dimensions, except for two dimensions that scored higher. Data Organisational Structure and Awareness scored close to 3; the same can be said about Data Privacy and Security, which has a slightly higher result. Data governance appears to be present in the organisation for those two dimensions, where data quality is kept in mind. There is a possibility that specific data processes have consistent and tailored organisational guidelines. Sometimes lacking consistency, as provided from the additional feedback, is per case basis.

The overall score when being presented with implementing the data governance framework at the organisation regarding the AI-integrated CRM system has an overall high score from 4,44 to 4,91, resulting in an average of 4,79, which according to the CMM, identifies the levels of measured and optimised (Marchildon et al., 2018). The additional feedback that was provided with the dimensions for the proposed framework was positively received, where there can be said that the layers and principles were clear and understandable. The proposed framework can be implemented as the process metrics are defined and performance is measured throughout the entire organisation (Marchildon et al., 2018). The levels measured and optimised indicate that best practices are often used and optimised consistently, as data quality is being monitored for inconsistencies. A data governance framework centre is established, focussing on data quality. The organisation is agile and confident in the established business objectives, aligned with the IT objectives. Data governance is adopted throughout the company, and communication is fully transparent. Potential improvements are identified and adapted constantly throughout the proposed data governance framework. However, limitations may apply because the proposed framework needs adapting based on specific organisations' setup, and situations can change once the proposed framework has been adopted in an organisation.

7.1.1. Analysis Data Governance Framework

A separate link was provided to the participants to view the proposed data governance framework in addition to the survey. Three layers were displayed on one page to make it easier for the participants to view the framework. An additional page displayed the principles coupled with the data governance framework. From the scoring of the overall dimensions and the additional feedback, it can be said that the respondents clearly understood the data governance framework. The organisation's current situation did not rely on a data governance framework, as practices also differ per project, which is also reflected in the scoring of the dimensions, with either a low or average score that indicates similar to what the feedback from the respondents provides. The proposed data governance framework achieved a high score because the foundation of the data governance framework is based on the model from Janssen et al. (2020). With the proposed data governance framework, all the departments of an organisation are linked together with its performance, in conjunction with following the set rules by country-

specific authorities. The performance of an organisation's process is managed where data is centric, having an overall picture of how data governance is supposed to be in an AI-integrated CRM system, coupled with specific principles that allow for the best practices to be followed.

7.1.1.1. General Improvements

All dimensions for the proposed data governance framework achieved a high score that translates to having either objectively measured or optimised process performance in the organisation, where best practices are being used (Soares, 2010). Improving continuously through setting quantitative business objectives, being a confident organisation with an agile working methodology, where the information sharing process is fully transparent (Marchildon et al., 2018). However, there were certain outliers in four dimensions, and some individual respondents have selected either a score of 3 or 2. Although the overall score was between 4 and 5, the proposed data governance framework aims to achieve the same scoring for individual responses that refer to the level of either quantitively measured or optimised (Soares, 2010). The table below indicates the dimensions that received a lower score from individual respondents.

Dimension	Respondent(s)	Score = Level
Data Risk Management and Compliance	1	3
Data Value Creation	1	2
Data Quality Management	1	2
Data Privacy and Security	1	3
Data Classification and Metadata	3	3

Table 7: Dimensions with lower scoring from individual responses

The Data Risk Management and Compliance dimension consider mitigation tactics through control and planning, ensuring that there are fail-safes in place to avoid risks in general (Marchildon et al., 2018), such as preparing for the worst possible outcome of data leakage. From the respondents' feedback, an improvement point can be assumed to make the proposed data governance framework more specific. As in the organisation's current situation, practices and guidelines are followed that abide by the legal compliance standard from the country, without a framework. Implementing a data governance framework takes more work and needs careful planning/execution, but it is overall more effective, as seen from the score. A data governance framework should work in conjunction with existing processes and an organisation's workflow; hence, specific scenario solutions should be set depending on the organisation (Janssen et al., 2020). Improving this dimension would make a more transparent and concise version that is case-specific per organisation (Kroll, 2018). Defining a straightforward plan for the most common scenarios an organisation can have regarding risk management and compliance with the AI-integrated CRM tool. Hence, depending on the scenario, there is no need to go through all the layers of the data governance framework when an issue arises.

The Data Value Creation dimension considers data to be qualified and quantified in an organisation, where all departments can create or benefit from the gathered data (Marchildon et al., 2018). Feedback from the respondents indicated that data silos were present in the current situation, assuming that the result would still stay the same with the proposed data governance framework. The key to solving data silos is through BD integration to find value, define the

maximum value of the silo and enhance collaboration and communication between the departments that handle the same data (Patel, 2019). A possible way to implement and define data silos is to implement techniques such as Hadoop to allow for BD scalability (Merla & Liang, 2017) and connect similar silos within the AI-integrated CRM system to avoid data silos in general. Although such tools can become complex and challenging to implement (Patel, 2019), enhancing collaboration and communication is the main factor. Hence, avoiding data silos can be tied with specific scenarios that allow for data sharing without risk whilst complying with regulations.

The Dimension of Data Quality Management refers to the ability of the gathered data to fulfil the satisfaction of its usage requirement (Soares, 2010). For the AI-integrated CRM to function correctly, accuracy, usability, and correctness are considered (Marchildon et al.,2018). The feedback foremost focuses on documenting and handling the quality issues. Although achieving a high score, the data governance framework does not have a specific process where data quality issues are documented and handled accordingly. Hence, a simple improvement would be to add a step to the third layer when data is being accessed by the person working with the AI-integrated CRM. Whether being accessed by an organisation employee or a client, data stewards will have to create a specific process that allows for such quality issues to be recorded and handled (Rosenbaum, 2010). Additional resources will likely need to be used and coupled with the AI-integrated CRM to be an infrastructure to store such information. In addition to the improvement, historical data quality issues can be used to learn and adapt, given that similar issues will occur in the future. Resulting in fewer resources being used in the long term for documenting and resolving quality issues.

The Data Classification and Metadata dimension considers the defined keywords to identify structured data, such as columns (Marchildon et al., 2018). Creating keywords for an organisation's internal use can describe semantics or characteristics for further interpretability (Khatri & Brown, 2010). Feedback from the respondents indicated that no database stores the metadata but differs per-project basis. An improvement can be made in the third layer of the data governance framework to include an internal database that can be used for metadata in an AI-integrated CRM. Additional resources will again be used, but having a metadata database can be beneficial in the future. Finding data faster by seeing historical projects that have used similar data, such as the same suppliers. Data will be more accessible (Soares, 2010), as the classification can be used to create valuable insights and linkage within the AI-integrated CRM. ML allows for creating different insights through made connections (Vafeiadis et al., 2015), as the metadata can be used to search for specific data characteristics. Classifications and keywords can be reused for future projects and save time when specific datasets need to be found.

Experts in data management have suggested adding a measurement for the respondents' confidence level (Marchildon et al., 2018), as one person cannot know all the details of an organisation. Thus, the results from the confidence level regarding the answers for the dimensions have achieved an overall high score of 4,63. Only three respondents answered with a score of two or three. The results can be concluded that the participants were confident in their answers. The open questions of additional feedback allowed respondents to elaborate on their scoring. There can be concluded that, although the survey was in-depth, the data governance framework was perceived positively and understandable from even a non-technological standpoint, as multiple departments from the organisation were involved with

the survey. Although limitations appeared, such as a shortened survey adapted, etc., they will be discussed in detail later on. The primary focus is on improving the Privacy and Security aspect of the data governance framework. Hence, analysis and improvement will be elaborated on in-depth in the following section.

7.1.1.2. Improving Privacy and Security

More in-depth questions were asked to gather a more accurate result, focusing on the primary dimension of Data Privacy and Security. Policies, principles and practices allow for limiting the exposure of data assets and mitigation of risks, such as confidentiality practices and regulations of data deletion (Marchildon et al., 2018). The feedback from the respondents provided a clear insight into the current situation where authorisation for accessing the data was determined per project, and higher management can access that data regardless if they are involved in the project. Even though the overall score with the proposed data governance framework has improved, the individual scoring has an outlier. Listing the parts of the framework that are tied to privacy and security, a specific scenario can be created to limit data access. The potential for data abuse will increase (Janssen et al., 2020) if anyone can freely access all data types. Holding an organisation responsible for the legal practices of data usage, access, sharing and distribution (Yang et al., 2019). The framework considers privacy and security in the first layer, where regulations dictate that data usage falls under the correct DPA, e.g. the GDPR (Raab & Szekely, 2017), industry standards, and ethics that eliminate bias.

The second layer of the data governance framework takes into account specific case scenarios (Kroll, 2018) that have been mentioned in the *General Improvements*. Data stewards also are responsible for access restriction and the implementation of security protocols, such as ***Figure 14: Cloud Service security taxonomy from Indu et al. (2018)***. Coupled with a risk-based approach that allows for security protocols, such as mitigation tactics and prevention efforts of outside and internal threats. The third layer of the framework takes a system-level approach toward developing an AI-integrated CRM that considers authentications methods and the supervision of data stewards when data is being handled. Customer communication is also received where data stewards can incorporate additional feedback, such as a data deletion request. Specific *Data Governance Framework Principles* for security and privacy that goes hand in hand with the data governance framework are mentioned, such as data handling, ethics, authorisation, consent, etc. To further improve the privacy and security aspect of the data governance framework and incorporate the respondents' feedback, a flowchart can be made for specific use when requesting access or deletion of the customer's data.

The following ***Figure 37: Data Access & Deletion Decision Flowchart*** has two scenarios that can be incorporated with ***Figure 21: System-level Data Governance Framework – Layer 3***. Both scenarios are simple yet effective and work together with all the layers of the data governance framework, considering the described principles. Data stewards play a crucial role (Rosenbaum, 2010) in both scenarios, where they have the formal responsibility to be liable for data usage through different departments and set up a structure of responsibility (Dawes, 2010) through custodial care of control, risk control & mitigation (Marchildon et al., 2018). Limiting access to data through access control and, even in the case of overprivileged users (Soares, 2010), denying or granting access to specific customer data with the reasoning backed by the data governance framework.

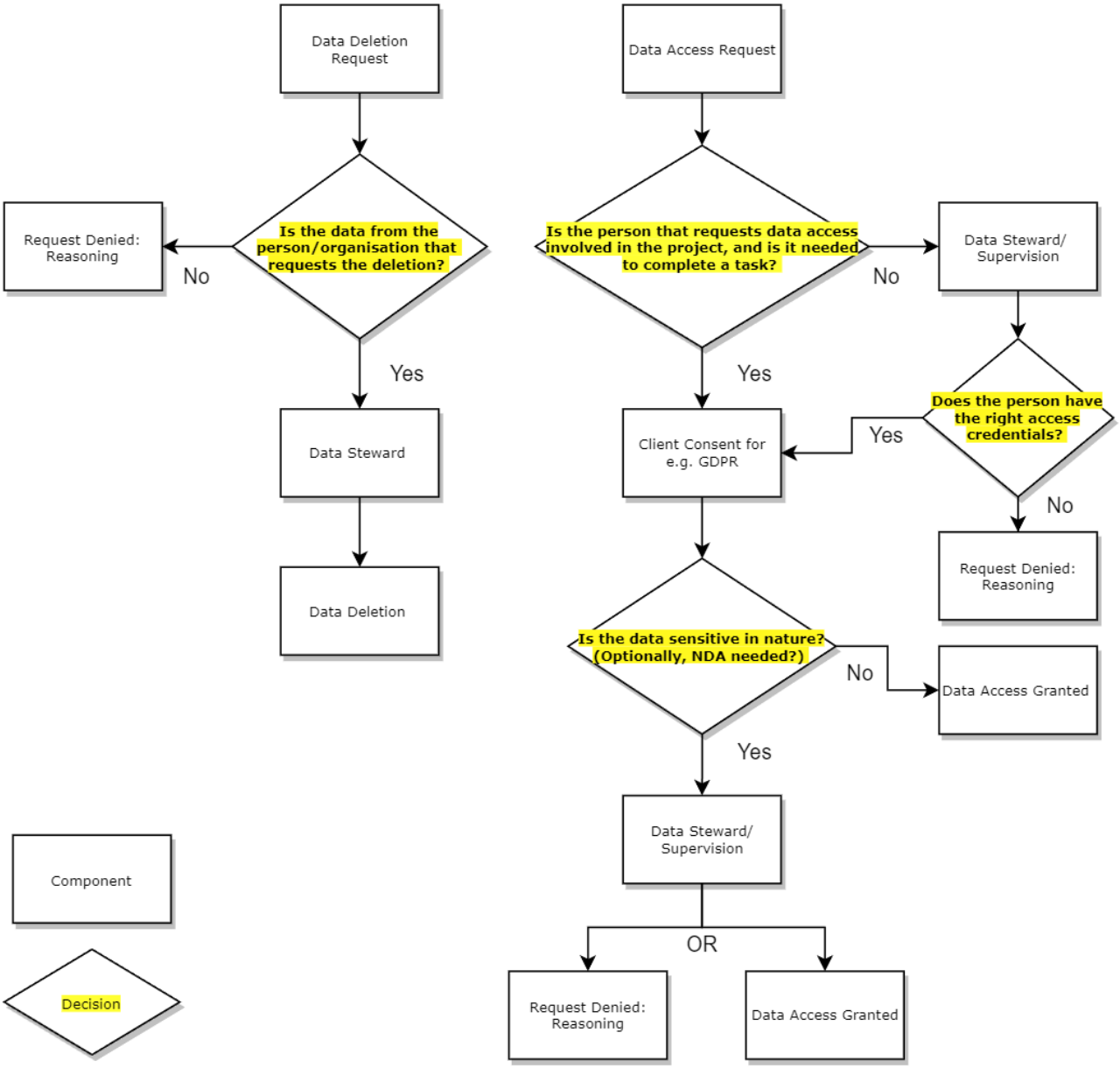


Figure 37: Data Access & Deletion Decision Flowchart

Even though historical data can be beneficial to find prior information, such as transactions, etc. (Zayah et al., 2012) and be used to train the AI-integrated CRM for better functionality, the GDPR requires customers to have the right to delete their data at any given time (EU commission, 2017). Hence, a simple flowchart was created for the situation when a customer requests the deletion of their data. Either the deletion of data will be executed, or reasoning is given why their data could not be deleted from the database through a data steward. The second scenario considers when employees or relevant parties request access to the customers' data. Overprivileged users often occur, and its misconfiguration or lack thereof (Torkura et al., 2020); hence an authentication method such as having the proper credentials is not enough to see all the customers' data. The customer's consent is approved for the organisation's use through the AI-integrated CRM; however, it does not always include the consent of third parties. Hence an additional check is made for consent to comply with the regulations. A distinction is also made between the nature of data, which can be sensitive, such as medical

information (Yang et al., 2019) and an option if an NDA is required. Depending on the answer, either data access is granted or denied with the reasoning.

***Figure* 37: Data Access & Deletion Decision Flowchart** is an addition to the data governance framework and its principles. Where authentication methods, such as the ***Figure 14: Cloud Service security taxonomy from Indu et al.* (2018)**, are taken into account, as well as security towards outside sources through the use of standards from ***Figure 19*: General Data Governance Framework –Layer 1** and ***Figure 20*: Organisation Structure Data Governance Framework – Layer 2**, that takes into account the planning, control and mitigation of risks. The benefit of improving the security & privacy aspect is for clarity and ease of use. Since no specific scenarios for custom implementations have been considered when developing the data governance framework, stipulating a clear flowchart can eliminate any confusion or ambiguity the framework might have. Naming the improvement as a refinement of the data governance framework can be applied universally. However, the disadvantage of creating specific scenarios is that they cannot always be applied to each setup. Hence, implementing a data governance framework needs to be looked at per situation of an organisation, as too much governance can lead to inefficiency and improper functioning (Janssen et al., 2020) of an AI-integrated CRM system.

7.2. Discussion Survey

Choosing a survey instead of regular interviews is to get measurable results that indicate the maturity and soundness of a data governance framework (Marchildon et al., 2018; Soares, 2010). Since the original assessment method from Soares (2010) is based on a survey, as well as the adapted version of Marchildon et al. (2018), with the mentioned models from Khatri & Brown (2010), Al-Ruithe et al. (2016), Weber et al. (2009) and the DGI (Thomas, 2006). The basis for the survey adapted from Marchildon et al. (2018) considers well-established and existing data governance maturity models, methods, and frameworks, which allows for a proper assessment that organisations can use to implement a data governance framework. Only one survey was made for all the involved departments, as one person could not know all the processes in an organisation, depending on the size of a business (Marchildon et al., 2018). Hence, the confidence level question was asked at the end of the survey to ensure a proper understanding of the questions and the presented data governance framework, coupled with the principles.

After each dimension, an open question was used to obtain additional feedback. Allowing for a more precise answer as to why the participants have chosen the scoring to a question (Marchildon et al., 2018). Although, a better approach could have been used to gather additional feedback, as not all participants answered the open questions. The survey did not require the participant to identify themselves, but additional questions relevant to the survey were asked, such as their position in an organisation. Arguably, that question could influence the participant's answer. However, there has been decided that the participants' position could provide a better insight into the provided scoring, as specific questions have had some irregular scores. Finally, the choice was made to use a shortened version of the survey from Marchildon et al. (2018) due to the extensive questions and a comparison that needed to be made between the presented situations. Feedback was given that the presented survey was already extensive. Thus limitations may apply as well, as results could vary, discussed later on.

7.3. Discussion Dimensions

The primary focus of the data governance framework is privacy and security; however, multiple dimensions have been chosen. Because all dimensions are related to data protection and safety to guard valuable customer data from abuse, human errors, or mishaps (Tallon et al., 2013). When developing a data governance framework, several vital questions need to be asked for organisations to understand what is right for them. Hence, the data governance maturity assessment tool from Marchildon et al. (2018) considers all the dimensions relevant to developing a sound framework based on several existing and well-known methods, techniques, and frameworks. There cannot only be focused on one dimension, as a data governance framework does not work if there are flaws in the foundation (Pence, 2014). No dimensions have been changed drastically from the assessment tool; only an emphasis on privacy and security has been made to get a more accurate outcome. Changing the assessment tool could provide for inaccurate results and change the core function of the tool (Marchildon et al., 2018).

Questions needed to be adapted from the assessment tool to the AI-integrated CRM system; since not all questions from the dimensions were relevant to be asked. Not only was an assessment made for the current situation in the organisation, but also with the proposed framework. Even though having focused on the improvements towards a few outliers in the individual scoring, the overall score towards each dimension could be said that they are excellent, with each almost scoring the maximum that translates to having the ideal situation of an optimised data governance framework and its principles (Marchildon et al., 2018). The scoring could improve with the proposed changes, but this varies per organisation. As shown in the individual confidence level scoring, one person alone cannot know all of an organisation's data governance processes (Marchildon et al., 2018). Only a few respondents scored two or three but overall have a high confidence level, possibly because two respondents had never used an AI-integrated CRM system prior.

7.4. Discussion Privacy & Security

Having created specific scenarios for improving the privacy and security aspect of the data governance framework, situations will always differ per organisation. An organisation will have their specific development and implementation (Zerbino et al., 2017), such as PrexPartners with its web crawler. However, the created *Figure* 37: Data Access & Deletion Decision Flowchart with specific scenarios can be used as a general application for data deletion and access processes. It will address users' most common error of data mishandling (Torkura et al., 2020). The data governance framework and its principles handle other issues regarding privacy and security, such as the mentioned ***Figure 14:*** **Cloud Service security taxonomy** from Indu et al. (2018) and in ***Figure 19*****: General Data Governance Framework –Layer 1** that handles the authentication part. There will always be scenarios not explicitly mentioned on how to address an AI-integrated CRM system from an organisation, as the framework is supposed to be adapted based on each setup. There is no perfect solution, only contingencies (Weber et al., 2009), such as using best practices, methods and principles in the data governance framework. There should be taken into account that too much data governance is also a problem (Janssen et al., 2020), and common sense should be used when assessing the need for a data governance framework.

8. Conclusion

This chapter concludes the thesis by answering the research questions introduced in the beginning. Future research will help to further develop the data governance field in an AI-integrated CRM system. Concluding with the limitations connected by this study that will be mentioned.

8.1. Answering Research Questions

8.1.1. Usage

A data governance framework was developed from the perspective of privacy and security. A lack of such a framework was established in the *Introduction* when an organisation plans to use or is using an AI-integrated CRM system. Through an adapted version of the assessment tool of Marchildon et al. (2018), the proposed research questions can be answered that are related to each other. The first research question is formulated as *"How can a data governance framework be used for the security and privacy aspect of an AI-integrated CRM system?"*.

The data governance framework can be used in an organisation to improve and maintain the privacy and security of data, which includes policies, principles, processes, and structures required to manage people such as employees, technologies and optimise data collection, usage, and storing and disseminating, which are mentioned through the three layers of the data governance framework and its principles. The first layer addresses data governance in general, which includes the regulations set by the DPA, which can prevent privacy infractions. Industry standards provided by standardisation organisations, such as ISO and ethics, are attached to social norms and unbiasedness that count towards the security and privacy of the data. The second layer addresses data governance in the structure of an organisation, such as the planning and control of its processes. Data stewards are appointed and identified with the role of accountability. Mitigation tactics and methods are made through a risk-based approach used in a worst-case scenario, such as data leakage.

The system-level data governance layer considers the development and the structure of an AI-integrated CRM and how the first two-layer can be incorporated. Such as the authentication methods and data stewards required at specific processes. Finally, the data governance principles serve as a foundation when implementing the framework. Through the survey, the current situation for privacy and security regarding data governance from an organisation can be improved from an average score of 3,27 to 4,86 on a scale of 5, which is a substantial improvement and can be identified as the highest level of data governance. Specific scenarios for privacy and security are created through the feedback from the respondents in ***Figure 37: Data Access & Deletion Decision Flowchart*** and can be used to improve the score of this dimension even further. The conclusion is that the data governance framework can be used to improve the security and privacy of an organisation that has or is using an AI-integrated CRM system.

8.1.2. Impact

The second research question is formulated as *"How does a data governance framework impact the security and privacy aspect of an AI-integrated CRM system?".*

A partial answer has been given in the first research question, where the current situation for privacy and security regarding data governance in an organisation has an average score of 3,27 on a scale of 5, which translates to the processes of an AI-integrated CRM system that are defined through standardised processes and followed through tailored organisational guidelines. Although, having indicated from the respondents' feedback, it is not always the case, such as third parties having access to sensitive data through the provided credentials. The average score with the data governance framework was 4,86 on a scale of 5. The additional ***Figure 37: Data Access & Deletion Decision Flowchart*** to the data governance framework impacts the process directly when people try to access customer data directly and the general improvement of the privacy and security score.

The provided scenarios directly impact the security and privacy of an AI-integrated CRM system. When someone attempts to access customer's data in an AI-integrated CRM system, the "Data Access Request" is triggered. Taking into account the data governance framework layers, such as data stewards and privacy legislation. Also, sensitive and non-sensitive data are distinguished, either providing access to customers' data or denying access with valid feedback, depending on the situation. The "Data Deletion Request" appears to be a straightforward scenario; however, it avoids misuse of the DPA's requirements and provides a straightforward data retention and deletion process based on the data governance framework. The direct impact is through the processes of how such requests are handled in an AI-integrated CRM. In contrast, the indirect impact improves an AI-integrated CRM system's overall privacy and security in data governance.

8.1. Limitations

The extensiveness of the assessment tool of Marchildon et al. (2018) could not be fully adapted to an AI-integrated CRM system due to the number of in-depth questions that did not apply to the organisation where the study was conducted. Hence, specific questions were not used, and some questions were adapted based on the used literature from the mentioned assessment tool. Also, a comparison was made with the current situation in the organisation and with the proposed data governance, which would double the number of questions asked in the survey and potentially lose the interest of the respondents (Marchildon et al., 2018). Due to the minor changes in the assessment tool, results will not be as accurate as intended.

The focus of this thesis is on privacy and security in data governance. Hence not all aspects of an AI-integrated CRM are discussed in-depth. The broad scope of BD, AI and BDAS has different angles that can be discussed. Hence the limitation in addressing all the related subjects is a possibility. Hence, the focus was given to a particular perspective of an AI-integrated CRM system, resulting in some parts not being adequately discussed. The mentioned dimensions (Soares, 2010; Lasrado et al., 2015; Marchildon et al., 2018) were necessary for the foundation of the data governance framework, but the lacking improvements in the dimensions need to be addressed further.

The data governance framework has a general approach toward organisations from all industries using an AI-integrated CRM system. However, it should be adapted according to the specific implementation and development of such a system. The advantage of the general approach is that it can be used as a foundation in data governance but should not just be implemented blindly. Company-specific developments, such as the web crawler at PrexPartners, should be considered within the data governance framework. Policies should be adapted accordingly, as too much data governance is also possible.

Finally, the data governance framework should be adopted, implemented, and tested to understand its effectiveness. Results may vary per organisation due to their setup and specific development implementations of such features and the purpose of the AI-integrated CRM system. The obtained scores in this thesis merely indicate what such a framework can be capable of when applied in an organisation.

8.2. Future research

The complexity of (new) systems built on AI and BD increases continuously, such as an AI-integrated CRM system. The main focus of this study is the privacy and security aspect of data governance in an AI-integrated CRM system; hence there are still plenty of areas that are as important and need to be explored or developed further.

8.2.1. Further Development

The section on *General Improvements* regarding the data governance framework highlights the improvements that can be made. However, these changes have not been discussed and developed due to the primary focus on privacy and security. Hence, a general focus can be made on the data governance framework and specific implementations regarding the use cases in an organisation. The *Limitations* section mentioned a different outcome that can be achieved, but only when an adapted version is developed and implemented in an organisation. Highlighting the importance through further studies due to the increasing use of an AI-integrated CRM system (Chatterjee et al., 2021). The growing awareness of BD and its use in such AI systems has become more relevant and needs to be explored further.

8.2.2. Trustworthy AI

Utilising an AI system can be difficult, much more so as the technology becomes more sophisticated. The outcomes of an AI system can have a significant impact on individuals or organisations. The next question is whether the AI system can be trusted to draw the correct conclusion based on the available data. As with an AI-integrated CRM system, a trustworthy BDAS requires technologies such as base registries and self-sovereign identities (Janssen et al., 2020). While responsible data collection, citizen control of data, and data stewardship are the foundations of data governance in AI systems, this new research field must continue to advance before it can create a solid research foundation for trusted AI systems using BD.

8.2.3. Informed Consent

Due to the growing concern of organisations misusing or abusing data, such as specific incidents with Facebook and Cambridge Analytica (Mikalef et al., 2019), society has become aware of their data's impact. Not only legal but ethical and moral concerns have been raised with the information usage and sharing between organisations. The missing factor is often not having received the prior consent of data usage from third parties, which is also considered in improving privacy and security. Hence a new concept of 'informed consent' should be explored further, taking into account our emotional state, promoting rational decision making, protecting autonomy, and valuing individuals (Andreotta et al., 2019).

8.2.4. Exploring Further

AI is a disruptive technology and can potentially transform the industry of its application and change how businesses operate (Verma et al., 2021). This can be seen from the usefulness of an AI-integrated CRM system that has only been recently adopted and developed. Psychologically motivated and brain-inspired reasoning algorithms would improve the predictability of consumer behaviour even further. Psychological theories addressing consumers' cognitive and affective needs, combined with engineering tools, will aid in the development of intelligent sentiment mining systems (Verma et al., 2021). Novel applications allow for further exploration and development in the complexity of a system, such as an AI-integrated CRM.

Acknowledgements

I have received a great deal of support and assistance from several people during the process of writing this thesis and would like to thank them profoundly.

Starting with the university supervisors, Professor Dr. Werner has guided me through the processes and given me incredible insights, whilst Professor Dr. Joost has given me structural advice and additional feedback. I want to thank them both for the time and honour to be under their supervision.

Company supervisor Dr. Phillip Harter at PrexPartners has given me the freedom and trust to work on the project with several colleagues. Giving wise feedback and knowledge that they have gained through their extensive experience in several industries.

To all the people involved that helped me through writing this thesis, directly and indirectly, I am grateful for having the opportunity to pursue my master's degree.

References

Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. International Journal of Information Management, 49, 424–438. https://doi.org/10.1016/j.ijinfomgt.2019.07.008.

Akter, S., & Wamba, S. F. (2016). Big data analytics in E-commerce: a systematic review and agenda for future research. Electronic Markets, 26(2), 173–194. https://doi.org/10.1007/s12525-016-0219-0

Al-Badi, A., Tarhini, A., & Khan, A. I. (2018). Exploring Big Data Governance Frameworks. Procedia Computer Science, 141, 271–277. https://doi.org/10.1016/j.procs.2018.10.181

Al-Ruithe, M., Benkhelifa, E., & Hameed, K. (2016). A Conceptual Framework for Designing Data Governance for Cloud Computing. Procedia Computer Science, 94, 160–167. https://doi.org/10.1016/j.procs.2016.08.025

Akoka, J., Comyn-Wattiau, I. and Laoufi, N., (2017), Research on Big Data–A systematic mapping study, Computer Standards & Interfaces, vol. 54, pp. 105-115.

Andreotta, A.J., Kirkham, N. & Rizzi, M. AI, big data, and the future of consent. AI & Soc (2021). https://doi.org/10.1007/s00146-021-01262-5

Anshari, M., Almunawar, M. N., Lim, S. A., & Al-Mudimigh, A. (2019). Customer relationship management and big data enabled: Personalization & customization of services. Applied Computing and Informatics, 15(2), 94–101. https://doi.org/10.1016/j.aci.2018.05.004

Amiel, T., & Reeves, T. C. (2008). Design-based research and educational technology: Rethinking technology and the research agenda. Educational Technology & Society, 11(4), 29-40.

Ardagna, C. A., Bellandi, V., Ceravolo, P., Damiani, E., Bezzi, M., & Hebert, C. (2017). A Model-Driven Methodology for Big Data Analytics-as-a-Service. 2017 IEEE International Congress on Big Data (BigData Congress). Published. https://doi.org/10.1109/bigdatacongress.2017.23

Badger L, Grance T, Patt-Corner R, Voas J (2012) Cloud computing synopsis and recommendations. Recommendations of the National Institute of Standards and Technology. NIST Special, Publication, pp. 800–146

Bahari, T. F., & Elayidom, M. S. (2015). An Efficient CRM-Data Mining Framework for the Prediction of Customer Behaviour. Procedia Computer Science, 46, 725–731. https://doi.org/10.1016/j.procs.2015.02.136

Barab, S. A., & Squire, K. (2004). Design-based research: Putting a stake in the ground. Journal of the Learning Sciences, 13(1), 1 - 14.

Barab, S. A., Dodge, T., Thomas, M. K., Jackson, C., & Tuzun, H. (2007). Our designs and the social agendas they carry. Journal of the Learning Sciences, 16(2), 263-305.

Bart, Y., Shankar, V., Sultan, F., & Urban, G. L. (2005). Are the Drivers and Role of Online Trust the Same for All Web Sites and Consumers? A Large-Scale Exploratory Empirical Study. Journal of Marketing, 69(4), 133–152. https://doi.org/10.1509/jmkg.2005.69.4.133

Bean, R., & Kiron, D. (2013). Organizational alignment is key to big data success. MIT Sloan Management Review, 54(3) o. S.

Belsis, P., and Pantziou, G.E. A k-Anonymity Privacy-Preserving Approach in Wireless Medical Monitoring Environments. Personal and Ubiquitous Computing 18(1), 2014.

Benfeldt, O., Persson, J. S., & Madsen, S. (2020). Data governance as a collective action problem. Information Systems Frontiers, 22(2), 299–313. https://doi.org/10.1007/s10796-019-09923-z.

Beretta, E., Vetrò, A., Lepri, B., & De Martin, J. C. (2018). Ethical and socially-aware data labels. Paper presented at the Annual International Symposium on Information Management and Big Data.

Bibiano, L. H., Marco-Simo, J. M., & Pastor, J. A. (2014). An initial approach for Improving CRM systems implementation projects. *2014 9th Iberian Conference on Information Systems and Technologies (CISTI)*. Published. https://doi.org/10.1109/cisti.2014.6876945

Bose, R. (2002). Customer relationship management: key components for IT success. Industrial Management & Data Systems, 102(2), 89–97. https://doi.org/10.1108/02635570210419636

Brewer, R., Westlake, B., Hart, T., & Arauza, O. (2021). The Ethics of Web Crawling and Web Scraping in Cybercrime Research: Navigating Issues of Consent, Privacy, and Other Potential Harms Associated with Automated Data Collection. Researching Cybercrimes, 435–456. https://doi.org/10.1007/978-3-030-74837-1_22

C.J. Stefanou, C. Sarmaniotis, and A. Stafyla, "CRM and customercentric knowledge management: an empirical research". Business Process Management Journal, vol. 9, no 5, pp. 617–634, 2003.

Chatterjee, S., Chaudhuri, R., Vrontis, D., Thrassou, A., & Ghosh, S. K. (2020). ICT-enabled CRM system adoption: A Dual Indian qualitative case study and Conceptual Framework Development. Journal of Asia Business Studies, 15(2), 257–277. https://doi.org/10.1108/jabs-05-2020-0198

Chatterjee, S., Ghosh, S. K., Chaudhuri, R., & Chaudhuri, S. (2020). Adoption of AI-integrated CRM system by Indian industry: from security and privacy perspective. *Information & Computer Security*, *29*(1), 1–24. https://doi.org/10.1108/ics-02-2019-0029

Chatterjee, S., Rana, N. P., Khorana, S., Mikalef, P., & Sharma, A. (2021). Assessing Organizational Users' Intentions and Behavior to AI Integrated CRM Systems: a Meta-UTAUT Approach. *Information Systems Frontiers*. Published. https://doi.org/10.1007/s10796-021-10181-1

Chatterjee, S., Tamilmani, K., Rana, N. P., & Dwivedi, Y. K. (2020). Employees' Acceptance of AI Integrated CRM System: Development of a Conceptual Model. *Re-imagining Diffusion and Adoption of Information Technology and Systems: A Continuing Conversation*, 679–687. https://doi.org/10.1007/978-3-030-64861-9_59

Chen, J., Jiang, W., & Yan, J. (2010). Understanding the Approach for Auditing of Cloud Computing System. 2010 Second International Conference on Information Technology and Computer Science. Published. https://doi.org/10.1109/itcs.2010.149

Chen, Z., & Yoon, J. (2010). IT Auditing to Assure a Secure Cloud Computing. 2010 6th World Congress on Services. Published. https://doi.org/10.1109/services.2010.118

Cheong, L.K. & Chang, V., 2007. The Need for Data Governance : A Case Study. ACIS 2007 Proceedings, pp.999–1008.

Cilimkovic, M. (2015). Neural networks and back propagation algorithm. Institute of Technology Blanchardstown, Blanchardstown Road North Dublin, 15, 1-12.

Coltman, T., Devinney, T. M., & Midgley, D. F. (2011). Customer Relationship Management and Firm Performance. *Journal of Information Technology*, *26*(3), 205–219. https://doi.org/10.1057/jit.2010.39

Cooley, R., Mobasher, B., & Srivastava, J. (1997). Web mining: information and pattern discovery on the World Wide Web. Proceedings Ninth IEEE International Conference

on Tools with Artificial Intelligence. Published. https://doi.org/10.1109/tai.1997.632303
Cooley, R. W.: Web Usage Mining: Discovery and Applications of Interesting Patterns from Web Data. 2000.
CSA. Alliance, Cloud Penetration Testing Playbook. Seattle, WA, USA: Cloud Security Alliance, 2019.
CSA. (2019). Cloud Controls Matrix Working Group | CSA. Cloud Security Alliance. Geraadpleegd op 1 november 2021, van https://cloudsecurityalliance.org/research/working-groups/cloud-controls-matrix/
Cuganesan, S., Hart, A., & Steele, C. (2017). Managing information sharing and stewardship for public-sector collaboration: A management control approach. Public Management Review, 19(6), 862–879.
Cuzzocrea, A. (2014). Privacy and Security of Big Data. Proceedings of the First International Workshop on Privacy and Secuirty of Big Data - PSBD '14. Published. https://doi.org/10.1145/2663715.2669614
Dai, W., Wardlaw, I., Cui, Y., Mehdi, K., Li, Y. and Long, J. (2016), Data profiling technology of data governance regarding big data: Review and rethinking, In Information Technology: New GenerationsSpringer, pp. 439-450.
Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2019). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 24-42.
Dawes, S. S. (2010). Stewardship and usefulness: Policy principles for information-based transparency. Government Information Quarterly, 27(4), 377–383.
De Haes, S., Van Grembergen, W., & Debreceny, R. S. (2013). COBIT 5 and Enterprise Governance of Information Technology: Building Blocks and Research Opportunities. Journal of Information Systems, 27(1), 307–324. https://doi.org/10.2308/isys-50422
Deighton, J. (2018). Big data. Consumption Markets & Culture, 22(1), 68–73. https://doi.org/10.1080/10253866.2017.1422902
Denyer, D., Tranfield, D. & van Aken, J.E., 2008. Developing Design Propositions through Research Synthesis. Organization Studies, 29(3), pp.393– 413.
Edelson, D. C. (2002). Commentary: Design-based research: What we learn when we engage in design. Journal of the Learning Sciences, 11(1), 105 - 121.
Elish, M. C., & Boyd, D. (2017). Situating methods in the magic of Big Data and AI. Communication Monographs, 85(1), 57–80. https://doi.org/10.1080/03637751.2017.1375130
ENISA. (2009). Cloud Computing Risk Assessment. Geraadpleegd op 1 november 2021, van https://www.enisa.europa.eu/publications/cloud-computing-risk-assessment
European Commission (2017). New European Interoperability Framework - Promoting seamless services and data flows for European public administrations. Retrieved from https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf.
Ewing , M . T .(2009) Integrated marketing communications measurement and evaluation . Journal of Marketing Communications 15 (2) : 103 – 117 .
EWT Ngai. Customer relationship management research (1992–2002): An academic literature review and classification, Marketing
Intelligence, Planning; 23, 2005. p. 582–605.
EWT Ngai, L Xiu, DCK.Chau. Application of Data Mining Techniques in Customer Relationship Management: A Literature Review on
Classification, Expert Systems with Applications; 36- 2, 2009. p. 2592-2602.
Foley, R., & Guillemette, M. G. (2010). What is Business Intelligence? International Journal of Business Intelligence Research, 1(4), 1–28. https://doi.org/10.4018/jbir.2010100101

Foss, B., Stone, M., & Ekinci, Y. (2008). What makes for CRM system success — Or failure? *Journal of Database Marketing & Customer Strategy Management, 15*(2), 68–78. https://doi.org/10.1057/dbm.2008.5

Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How "big data" can make big impact: Findings from a systematic review and a longitudinal case study. International Journal of Production Economics, 165, 234–246. http://doi.org/10.1016/j.ijpe.2014.12.031.

Garrido-Moreno, Aurora & Padilla-Meléndez, Antonio & Del Aguila-Obra, Ana Rosa. (2010). Exploring the Importance of Knowledge Management for CRM Success. World Academy of Science, Engineering and Technology. 66.

Gartner, M., and R. Brocca. 2015. "Deconstructing Supply Chain Analytics." Journal of Supply Chain Management 25 (2015): 200–326.

George, G., Haas, M. R., & Pentland, A. (2014). Big Data and management. Academy of Management Journal, 57(2), 321–326. http://doi.org/10.1111/risa.12257.

Graca, S.S., Barry, J.M. and Doney, P.M. (2015), "Performance outcomes of behavioral attributes in buyersupplier relationships", Journal of Business and Industrial Marketing, Vol. 30 No. 7, pp. 805-816.

Giurgiu, A., & A Larsen, T. (2016). Roles and Powers of National Data Protection Authorities. European Data Protection Law Review, 2(3), 342–352. https://doi.org/10.21552/edpl/2016/3/9

Gupta, N., Blair, S., & Nicholas, R. (2020). What We See, What We Don't See: Data Governance, Archaeological Spatial Databases and the Rights of Indigenous Peoples in an Age of Big Data. In Journal of Field Archaeology (Vol. 45, Issue sup1, pp. S39–S50). Informa UK Limited. https://doi.org/10.1080/00934690.2020.1713969

Hadden J., Tiwari A., Roy R., Ruta D., Churn prediction: does technology matter, Int. J. Intell. Technol. 1 (2) (2006) 104–110.

Hany AE. Bank Direct Marketing Analysis of Data Mining Techniques, International Journal of Computer Applications; 85-7, 2014.

Hofmann, E. (2015). Big data and supply chain decisions: the impact of volume, variety and velocity properties on the bullwhip effect. International Journal of Production Research, 55(17), 5108–5126. https://doi.org/10.1080/00207543.2015.1061222

Holzinger, A. (2018). From Machine Learning to Explainable AI. 2018 World Symposium on Digital Intelligence for Systems and Machines (DISA). Published. https://doi.org/10.1109/disa.2018.8490530

Hossain, S. F. A., Xi, Z., Nurunnabi, M., & Hussain, K. (2020). Ubiquitous Role of Social Networking in Driving M-Commerce: Evaluating the Use of Mobile Phones for Online Shopping and Payment in the Context of Trust. SAGE Open, 10(3), 215824402093953. https://doi.org/10.1177/2158244020939536

Hoyer, W. D., Kroschke, M., Schmitt, B., Kraume, K., & Shankar, V. (2020). Transforming the Customer Experience Through New Technologies. *Journal of Interactive Marketing, 51*, 57–71. https://doi.org/10.1016/j.intmar.2020.04.001

Huang, W., Ganjali, A., Kim, B. H., Oh, S., & Lie, D. (2015). The State of Public Infrastructure-as-a-Service Cloud Security. ACM Computing Surveys, 47(4), 1–31. https://doi.org/10.1145/2767181

Indu, I., Anand, P. R., & Bhaskar, V. (2018). Identity and access management in cloud environment: Mechanisms and challenges. Engineering Science and Technology, an International Journal, 21(4), 574–588. https://doi.org/10.1016/j.jestch.2018.05.010

Iosup, A., Prodan, R., & Epema, D. (2014). IaaS Cloud Benchmarking: Approaches, Challenges, and Experience. Cloud Computing for Data-Intensive Applications, 83–104. https://doi.org/10.1007/978-1-4939-1905-5_4

Ishibuchi, H., Yamane, M., and Nojima, Y. Learning from Multiple Data Sets with Different Missing Attributes and Privacy Policies: Parallel Distributed Fuzzy Genetics-based Machine Learning Approach. Proc. of BigData Conference, 2013.

ISO. (2017). ISO/IEC 38505–1:2017. ISO - International Organization for Standardization. Geraadpleegd op 25 november 2021, van https://www.iso.org/standard/56639.html

Iqbal, T., & Khan, M. (2021, 1 januari). The Impact of Artificial Intelligence (AI) on CRM and Role of Marketing Managers. DIVA Portal. Geraadpleegd op 1 oktober 2021, van http://hig.diva-portal.org/smash/record.jsf?pid=diva2%3A1526230&dswid=4810

ISACA. (2017). Controls and Assurance in the Cloud: Using COBIT 5. Geraadpleegd op 1 november 2021, van https://www.isaca.org/resources/isaca-journal/issues/2017/volume-3/controls-and-assurance-in-the-cloud-using-cobit-5

Jain, P., Gyanchandani, M., & Khare, N. (2016). Big data privacy: a technological perspective and review. Journal of Big Data, 3(1). https://doi.org/10.1186/s40537-016-0059-y

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. *Government Information Quarterly*, *37*(3), 101493. https://doi.org/10.1016/j.giq.2020.101493

Janssen, M., Kuk, G. (2016). The challenges and limits of big data algorithms in technocratic governance. Government Information Quarterly, 33(3), 371–377. https:// doi.org/10.1016/j.giq.2016.08.011.

Janssen, M., Matheus, R., & Zuiderwijk, A. (2015). Big and open linked data (BOLD) to create smart cities and citizens: Insights from smart energy and mobility cases. In E. Tambouris, M. Janssen, H. J. Scholl, M. A. Wimmer, K. Tarabanis, M. Gascó, ... P. Parycek (Vol. Eds.), Electronic Government. 9248. Electronic Government (pp. 79–90). Springer International Publishing.

Janssen, M., Matheus, R., Longo, J., & Weerakkody, V. (2017). Transparency-by-design as a foundation for open government. Transforming Government: People, Process and Policy.

Janssen, M., Van der Voort, H. (2016). Adaptive governance: Towards a stable, accountable and responsive government. Government Information Quarterly, 33(1), 1–5. https://doi.org/10.1016/j.giq.2016.02.003.

Jing, Q., Vasilakos, A. V., Wan, J., Lu, J., & Qiu, D. (2014). Security of the Internet of Things: perspectives and challenges. Wireless Networks, 20(8), 2481–2501. https://doi.org/10.1007/s11276-014-0761-7

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, *62*(1), 15–25. https://doi.org/10.1016/j.bushor.2018.08.004

Kardon, B. (2019, February 13). *Five AI Solutions Transforming B2B Marketing*. Retrieved from MIT Sloan Management Review: https://sloanreview.mit.edu/article/five-ai-solutions-transforming-b2b-marketing/

Karthiban, K., & Smys, S. (2018). Privacy preserving approaches in cloud computing. 2018 2nd International Conference on Inventive Systems and Control (ICISC). Published. https://doi.org/10.1109/icisc.2018.8399115

Katal A., Wazid M., and R. H. Goudar, "Big data: Issues, challenges, tools and Good practices," 2013 Sixth International Conference on Contemporary Computing (IC3), 2013, pp. 404-409, doi: 10.1109/IC3.2013.6612229.

Kennedy, A. (2006), "Electronic customer relationship management (eCRM): opportunities and challenges", Irish Marketing Review, Vol. 18 No. 1/2, pp. 58-69.

Kennedy-Clark, S. (2013). Research by Design: Design-Based Research and the Higher Degree Research student. *Journal of Learning Design*, *6*(2). https://doi.org/10.5204/jld.v6i2.128

Keramati, A., Mehrabi, H., & Mojir, N. (2010). A process-oriented perspective on customer relationship management and organizational performance: An empirical investigation. *Industrial Marketing Management, 39*(7), 1170–1185. https://doi.org/10.1016/j.indmarman.2010.02.001

Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM, 53*(1), 148–152. https://doi.org/10.1145/1629175.1629210

Kim, H. Y., & Cho, J. S. (2017). Data Governance Framework for Big Data Implementation with a Case of Korea. 2017 IEEE International Congress on Big Data (BigData Congress). Published. https://doi.org/10.1109/bigdatacongress.2017.56

Kirui C., Hong L., Cheruiyot W., Kirui H., Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining, Int. J. Comput. Sci. Iss. (IJCSI) 10 (2) (2013) 165–172.

Kraljevic G., Gotovac S., Modeling data mining applications for prediction of prepaid churn in telecommunication services, AUTOMATIKA: cˇasopis za automatiku, mjerenje, elektroniku, racˇunarstvo i komunikacije 51 (3) (2010) 275–283.

Koltay, T. (2016). Data governance, data literacy and the management of data quality. IFLA Journal, 42(4), 303–312. https://doi.org/10.1177/0340035216672238

Kroll, J. A. (2018). Data Science Data Governance [AI Ethics]. *IEEE Security & Privacy, 16*(6), 61–70. https://doi.org/10.1109/msec.2018.2875329

Krumm, J. "Ubiquitous Advertising: The Killer Application for the 21st Century," IEEE Pervasive Computing, 10(1), 66-73, 2011.

Kshetri, N. 2014. "Big Data's Impact on Privacy, Security and Consumer Welfare." Telecommunications Policy 38 (11): 1134–1145. doi:10.1016/j. telpol.2014.10.002.

Ladley, J. (2019). Data governance: How to design, deploy, and sustain an effective data governance program. Academic Press.

Lasrado, L. A., Vatrapu, R., & Andersen, K. N. (2015, August). Maturity models development in is research: a literature review. In IRIS Selected Papers of the Information Systems Research Seminar in Scandinavia (Vol. 6, No. 6). New York: IRIS.

Lee S.J., Siau K., A review of data mining techniques, Ind. Manage. Data Syst. 101 (1) (2001) 41–46

Li, J., Qiu, M., Ming, Z., Quan, G., Qin, X., & Gu, Z. (2012). Online optimization for scheduling preemptable tasks on IaaS cloud systems. Journal of Parallel and Distributed Computing, 72(5), 666–677. https://doi.org/10.1016/j.jpdc.2012.02.002

Li, M. and Nguyen, B. (2016), "When will firms share information and collaborate to achieve innovation? A review of collaboration strategies", The Bottom Line, Vol. 30 No. 1, pp. 65-86.

Libai, B., Bart, Y., Gensler, S., Hofacker, C. F., Kaplan, A., Kötterheinrich, K., & Kroll, E. B. (2020). Brave New World? On AI and the Management of Customer Relationships. *Journal of Interactive Marketing, 51*, 44–56. https://doi.org/10.1016/j.intmar.2020.04.002

Linoff G.S., Berry M.J, Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, John Wiley & Sons, 2011.

Liu , C .- H .and Wang , C .- C .(2010) Formulating service business strategies with integrative services model from customer and provider perspectives . European Journal of Marketing 44 (9/10) : 1500 – 1527 .

Liu, C. H. (2015). A Conceptual Framework of Analytical CRM in Big Data Age. International Journal of Advanced Computer Science and Applications, 6(6). https://doi.org/10.14569/ijacsa.2015.060620

Maduka A. J., Aghili S., and Butakorv S., ―A Proposed Assurance model to Assess Security and Privacy risks in IaaS and PaaS Environments,‖ Annu. Symp. Inf. Assur. (ASIA '17), pp. 61–67, 2017.

Mahdavinejad, M. S., Rezvan, M., Barekatain, M., Adibi, P., Barnaghi, P., & Sheth, A. (2018). Machine learning for internet of things data analysis: a survey. Digital Communications and Networks, 161-175.

Malhotra, A., & Kubowicz Malhotra, C. (2010). Evaluating Customer Information Breaches as Service Failures: An Event Study Approach. Journal of Service Research, 14(1), 44–59. https://doi.org/10.1177/1094670510383409

Mansour, R. F. (2016). Understanding how big data leads to social networking vulnerability. *Computers in Human Behavior*, *57*, 348–351. https://doi.org/10.1016/j.chb.2015.12.055

Maniam, J., & Singh, D. (2020). TOWARDS DATA PRIVACY AND SECURITY FRAMEWORK IN BIG DATA GOVERNANCE. International Journal of Software Engineering and Computer Systems, 6(1), 41–51. https://doi.org/10.15282/ijsecs.6.1.2020.5.0068

Marchildon, P., Bourdeau, S., Hadaya, P., & Labissière, A. (2018). Data governance maturity assessment tool: A design science approach. Projectics / Proyéctica / Projectique, n°20(2), 155. https://doi.org/10.3917/proj.020.0155

Marsh , R .(2005) Drowning in dirty data? It's time to sink or swim: A four-stage methodology for total data quality management . Journal of Database Marketing &Customer Strategy Management 12 (2) : 105 – 112 .

Melton, J. Sql language summary. ACM Comput. Surv., 28(1):141–143, 1996.

Merla, P., & Liang, Y. (2017). Data analysis using hadoop MapReduce environment. In 2017 IEEE International Conference on Big Data (Big Data). 2017 IEEE International Conference on Big Data (Big Data). IEEE. https://doi.org/10.1109/bigdata.2017.8258541

Mi tra S, Pal SK, Mitra P. Data mining in soft computing framework: A survey, IEEE Transactions on Neural Networks; 13, 2002. p. 3–14.

Mikalef, P., M. Boura, G. Lekakos, and J. Krogstie. 2019. "Big Data Analytics and Firm Performance: Findings from a Mixed-Method Approach." Journal of Business Research 98: 261–276. doi:10.1016/j. jbusres.2019.01.044.

Min, H. (2009). Artificial intelligence in supply chain management: theory and applications. *International Journal of Logistics Research and Applications*, *13*(1), 13–39. https://doi.org/10.1080/13675560902736537

Mogaji, E., Soetan, T. O., & Kieu, T. A. (2020). The implications of artificial intelligence on the digital marketing of financial services to vulnerable customers. *Australasian Marketing Journal*, *29*(3), 235–242. https://doi.org/10.1016/j.ausmj.2020.05.003

Molinillo, S., & Japutra, A. (2017). Organizational adoption of digital information and technology: a theoretical review. *The Bottom Line*, *30*(01), 33–46. https://doi.org/10.1108/bl-01-2017-0002

Mullon, P. A., & Ngoepe, M. (2019). An integrated framework to elevate information governance to a national level in South Africa. Records Management Journal, 29(1/2), 103–116. https://doi.org/10.1108/rmj-09-2018-0030

Morabito, V. (2015), Big data governance, In Big data and analyticsSpringer, pp. 83-104

Musa, A., and A. A. A. Dabo. 2016. "A Review of RFID in Supply Chain Management: 2000–2015." Global Journal of Flexible Systems Management 17 (2): 189–228. doi:10.1007/s40171-016-0136-2

Nasteski, V. (2017). An overview of the supervised machine learning methods. HORIZONS.B, 4, 51–62. https://doi.org/10.20544/horizons.b.04.1.17.p05

Ng, A. (2017, September 21). *What AI Can and Can't Do*. Retrieved from Harvard Business Review: https://hbr.org/2016/11/what-artificial-intelligence-can-and-cant-do-right-now

Nguyen, B., & Mutum, D. S. (2012). A review of customer relationship management: successes, advances, pitfalls and futures. *Business Process Management Journal*, *18*(3), 400–419. https://doi.org/10.1108/14637151211232614

Nguyen, T. H., Sherif, J. S., & Newby, M. (2007). Strategies for successful CRM implementation. Information Management & Computer Security, 15(2), 102–115. https://doi.org/10.1108/09685220710748001

Niranjanamurthy, M., Kavyashree, N., 2013. Analysis of e-commerce and m-commerce: advantages, limitations and security issues. Int. J. Adv. Res. Comput. Commun. Eng. 2, 2360–2370.

Offermann, P., Levina, O., Schönherr, M., & Bub, U. (2009). Outline of a design science research process. Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology - DESRIST '09. Published. https://doi.org/10.1145/1555619.1555629

Ogbuke, N. J., Yusuf, Y. Y., Dharma, K., & Mercangoz, B. A. (2020). Big data supply chain analytics: ethical, privacy and security challenges posed to business, industries and society. Production Planning & Control, 1–15. https://doi.org/10.1080/09537287.2020.1810764

Oussous, A., Benjelloun, F. Z., Ait Lahcen, A., & Belfkih, S. (2018). Big Data technologies: A survey. Journal of King Saud University - Computer and Information Sciences, 30(4), 431–448. https://doi.org/10.1016/j.jksuci.2017.06.001

Patel, J. (2019). Bridging data silos using Big Data integration. International Journal of Database Management Systems, 11(3), 01-06.

Payton , F . C .and Zahay , D .(2003) Understanding why marketing does not use the corporate data warehouse for CRM applications . Journal of Database Marketing 10 (4) : 315 – 326 .

Pence, H. E. (2014). What is Big Data and Why is it Important? Journal of Educational Technology Systems, 43(2), 159–171. https://doi.org/10.2190/et.43.2.d

Pencheva, I., Esteve, M., & Mikhaylov, S. J. (2018). Big Data and AI – A transformational shift for government: So, what next for research? Public Policy and Administration, 35(1), 24–44. https://doi.org/10.1177/0952076718780537

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(3), 45–77. https://doi.org/10.2753/mis0742-1222240302

Phillips-Wren, G., & Hoskisson, A. (2015). An analytical journey towards big data. Journal of Decision Systems, 24(1), 87–102. http://doi.org/10.1080/12460125.2015. 994333.

Plomp, T. (2007). Educational design-based research: An introduction. In T. Plomp & N. Nieveen (Eds.), An Introduction to Educational Design-based research. Proceedings of the seminar conducted at the East China Normal University, Shangai (PR China), November 23-26, 2007 (pp. 9-33): SLO Netherlands institute for curriculum development.

Porambage, P., Ylianttila, M., Schmitt, C., Kumar, P., Gurtov, A., & Vasilakos, A. V. (2016). The Quest for Privacy in the Internet of Things. IEEE Cloud Computing, 3(2), 36–45. https://doi.org/10.1109/mcc.2016.28

Potiguara Carvalho, A., Potiguara Carvalho, F., Dias Canedo, E., & Potiguara Carvalho, P. H. (2020). Big Data, Anonymisation and Governance to Personal Data Protection. The 21st Annual International Conference on Digital Government Research. Published. https://doi.org/10.1145/3396956.3398253

Raab, C., & Szekely, I. (2017). Data protection authorities and information technology. Computer Law & Security Review, 33(4), 421–433. https://doi.org/10.1016/j.clsr.2017.05.002

Radosavljevik D., van der Putten P., K.K. Larsen, The impact of experimental setup in prepaid churn prediction for mobile telecommunications: what to predict, for whom and does the customer experience matter?, Trans MLDM 3 (2) (2010) 80–99.

Rana, N. P., Chatterjee, S., Dwivedi, Y. K., & Akter, S. (2021). Understanding dark side of artificial intelligence (AI) integrated business analytics: assessing firm's operational inefficiency and competitiveness. *European Journal of Information Systems*, 1–24. https://doi.org/10.1080/0960085x.2021.1955628

Rani, D., & Ranjan, R. K. (2014). A comparative study of SaaS, PaaS and IaaS in cloud computing. International Journal of Advanced Research in Computer Science and Software Engineering, 4(6).

Ransbotham, S., Gerbert, P., Reeves, M., Kiron, D., & Spira, M. (2018, September 17). *Artificial Intelligence in Business Gets Real*. Retrieved from MIT Sloan Management Review: https://sloanreview.mit.edu/projects/artificial-intelligence-in-business-gets-real/

Reddy, S., Allan, S., Coghlan, S., & Cooper, P. (2019). A governance model for the application of AI in health care. *Journal of the American Medical Informatics Association*, *27*(3), 491–497. https://doi.org/10.1093/jamia/ocz192

Riggins, F. J., & Klamm, B. K. (2017). Data governance case at krausemcmahon LLP in an era of self-service BI and Big Data. Journal of Accounting Education, 38, 23–36. https://doi.org/10.1016/j.jaccedu.2016.12.002

Rosenbaum, S. (2010). Data Governance and Stewardship: Designing Data Stewardship Entities and Advancing Data Access. Health Services Research, 45(5p2), 1442–1455. https://doi.org/10.1111/j.1475-6773.2010.01140.x

Rothstein, H., Borraz, O., & Huber, M. (2013). Risk and the limits of governance: Exploring varied patterns of risk-based governance across Europe. Regulation & Governance, 7(2), 215–235. https://doi.org/10.1111/j.1748-5991.2012.01153.x

Saltzer, "Protection and the control of information sharing in multics", Communications of the 5 ACM, 17 (1974), pp. 388-402.

Sarmah, S. (2019). Database Security –Threats & Prevention. International Journal of Computer Trends and Technology. 67. 46-53. https://doi.org/10.14445/22312803/IJCTT-V67I5P108.

Schneider, J., Abraham, R., & Meske, C. (2020, 20 november). *AI Governance for Businesses*. ArXiv.Org. https://arxiv.org/abs/2011.10672

Schrage, M., & Kiron, D. (2018, August 21). *Improving Strategic Execution With Machine Learning*. Retrieved from MIT Sloan Management Review: https://sloanreview.mit.edu/article/improving-strategic-execution-with-machine-learning/

Serrano, N., Gallardo, G., & Hernantes, J. (2015). Infrastructure as a Service and Cloud Technologies. IEEE Software, 32(2), 30–36. https://doi.org/10.1109/ms.2015.43

Shahbaz, M., Gao, C., Zhai, L., Shahzad, F., Abbas, A., & Zahid, R. (2020). Investigating the Impact of Big Data Analytics on Perceived Sales Performance: The Mediating Role of Customer Relationship Management Capabilities. Complexity, 2020, 1–17. https://doi.org/10.1155/2020/5186870

Sharma, D. H., Dhote, C., & Potey, M. M. (2016). Identity and Access Management as Security-as-a-Service from Clouds. Procedia Computer Science, 79, 170–174. https://doi.org/10.1016/j.procs.2016.03.117

Soares, S. (2010). The IBM Data Governance Unified Process: Driving Business Value with IBM Software and Best Practices. Boise, ID: MC Press.

Soares, S. (2012). Big Data Governance: An Emerging Imperative. Boise, ID: MC Press.

Soares, S. (2014). Data Governance Tools: Evaluation Criteria, Big Data Governance, and Alignment with Enterprise Data Management. Boise, ID: MC Press.

Taft, R., Sharif, I., Matei, A., VanBenschoten, N., Lewis, J., Grieger, T., Niemi, K., Woods, A., Birzin, A., Poss, R., Bardea, P., Ranade, A., Darnell, B., Gruneir, B., Jaffray, J., Zhang, L., & Mattis, P. (2020). CockroachDB: The Resilient Geo-Distributed SQL Database. In Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data. SIGMOD/PODS '20: International Conference on Management of Data. ACM. https://doi.org/10.1145/3318464.3386134

Tallon, P. P., Ramirez, R. V. & Short, J. E. (2013). The information artifact in IT governance: Toward a theory of information governance. Journal of Management Information Systems, 30(3), 141-178.

Tariq, M. I., & Santarcangelo, V. (2016). Analysis of ISO 27001:2013 Controls Effectiveness for Cloud Computing. Proceedings of the 2nd International Conference on Information Systems Security and Privacy. Published. https://doi.org/10.5220/0005648702010208

Tawalbeh, L. A., & Saldamli, G. (2021). Reconsidering big data security and privacy in cloud and mobile cloud systems. Journal of King Saud University - Computer and Information Sciences, 33(7), 810–819. https://doi.org/10.1016/j.jksuci.2019.05.007

Thomas, G. (2006). The DGI data governance framework. The Data Governance Institute (DGI), Orlando, FL (USA), 20.

Torkura, K. A., Sukmana, M. I. H., Cheng, F., & Meinel, C. (2020). CloudStrike: Chaos Engineering for Security and Resiliency in Cloud Infrastructure. IEEE Access, 8, 123044–123060. https://doi.org/10.1109/access.2020.3007338

Tom M Mitchell. Machine Learning, 2nd ed. McGraw Hill; 2010.

Toth, K. C., & Anderson-Priddy, A. (2019). Self-Sovereign Digital Identity: A Paradigm Shift for Identity. IEEE Security & Privacy, 17(3), 17–27. https://doi.org/10.1109/msec.2018.2888782

Ularu, E. G., Puican, F. C., Apostu, A. and Velicanu, M., (2012), Perspectives on big data and big data analytics, Database Systems Journal, vol. 3 (4), pp. 3-14.

Vafeiadis, T., Diamantaras, K., Sarigiannidis, G., & Chatzisavvas, K. (2015). A comparison of machine learning techniques for customer churn prediction. Simulation Modelling Practice and Theory, 55, 1–9. https://doi.org/10.1016/j.simpat.2015.03.003

Van Aken, J.E., 2005. Management research as a design science: Articulating the research products of mode 2 knowledge production in management. British Journal of Management, pp.19–36.

Vance, A., Molyneux, B., & Lowry, P. B. (2012). Reducing Unauthorized Access by Insiders through User Interface Design: Making End Users Accountable. 2012 45th Hawaii International Conference on System Sciences. Published. https://doi.org/10.1109/hicss.2012.499

Vaquero, L. M. (2011). EduCloud: PaaS versus IaaS Cloud Usage for an Advanced Computer Science Course. IEEE Transactions on Education, 54(4), 590–598. https://doi.org/10.1109/te.2010.2100097

Verma, D. and Verma, D.S. (2013), "Managing customer relationships through mobile CRM in organized retail outlets", International Journal of Engineering Trends and Technology, Vol. 4 No. 5, pp. 1696-1701.

Verma, S., Sharma, R., Deb, S., & Maitra, D. (2021). Artificial intelligence in marketing: Systematic review and future research direction. International Journal of Information Management Data Insights, 1(1), 100002. https://doi.org/10.1016/j.jjimei.2020.100002

Wen, K.W. and Chen, Y. (2010),"E-business value creation in small and medium enterprises: a US study using the TOE framework", International Journal of Electronic Business, Vol. 8 No. 1, pp. 80-100.

Wai. H. Au, K.C. Chan, X. Yao, A novel evolutionary data mining algorithm with applications to churn prediction, IEEE Trans. Evol. Comput. 7 (6) (2003) 532–545

Wang, L. (2017). Big Data in Intrusion Detection Systems and Intrusion Prevention Systems. Journal of Computer Networks, 4(1), 48–55. https://doi.org/10.12691/jcn-4-1-5

Weber, K., Otto, B., & Österle, H. (2009). One Size Does Not Fit All---A Contingency Approach to Data Governance. Journal of Data and Information Quality, 1(1), 1–27. https://doi.org/10.1145/1515693.1515696

Wihlborg, E., Larsson, H., & Hedström, K. (2016). "The Computer Says No!"–A Case Study on Automated Decision-Making in Public Authorities. Paper presented at the 2016 49th Hawaii International Conference on System Sciences (HICSS).

Xiong, J., Zhang, Y., Tang, S., Liu, X., & Yao, Z. (2019). Secure Encrypted Data With Authorized Deduplication in Cloud. IEEE Access, 7, 75090–75104. https://doi.org/10.1109/access.2019.2920998

Wu, C., and Guo, Y. Enhanced User Data Privacy with Pay-By-Data Model. Proc. of BigData Conference, 2013.

Xu, M., & Walton, J. (2005). Gaining customer knowledge through analytical CRM. Industrial Management & Data Systems, 105(7), 955–971. https://doi.org/10.1108/02635570510616139

Yang, L., Li, J., Elisa, N., Prickett, T., & Chao, F. (2019). Towards Big data Governance in Cybersecurity. Data-Enabled Discovery and Applications, 3(1). https://doi.org/10.1007/s41688-019-0034-9

Yong Ahn, J., Ki Kim, S., & Soo Han, K. (2003). On the design concepts for CRM system. Industrial Management & Data Systems, 103(5), 324–331. https://doi.org/10.1108/02635570310477370

Youn, S., & Jin, S. V. (2021). "In A.I. we trust?" The effects of parasocial interaction and technopian versus luddite ideological views on chatbot-based customer relationship management in the emerging "feeling economy". *Computers in Human Behavior*, *119*, 106721. https://doi.org/10.1016/j.chb.2021.106721

Yuri D., Zhiming Z., Paola G., Adianto W., Cees d.L., "Addressing Big Data Challenges for Scientific Data Infrastructure", IEEE , 4th International Conference on Cloud Computing Technology and Science, 2012.

Zahay, D., Peltier, J., & Krishen, A. S. (2012). Building the foundation for customer data quality in CRM systems for financial services firms. Journal of Database Marketing & Customer Strategy Management, 19(1), 5–16. https://doi.org/10.1057/dbm.2012.

Zaki, A.K. (2014). NoSQL DATABASES: NEW MILLENNIUM DATABASE FOR BIG DATA, BIG USERS, CLOUD COMPUTING AND ITS SECURITY CHALLENGES. International Journal of Research in Engineering and Technology, 03, 403-409.

Zerbino, P., Aloini, D., Dulmin, R., & Mininno, V. (2018). Big Data-enabled Customer Relationship Management: A holistic approach. Information Processing & Management, 54(5), 818–846. https://doi.org/10.1016/j.ipm.2017.10.005

Appendix

Appendix I – Blank Survey

Data Governance Framework Survey

Hi,

I would like to ask you to fill in this survey because you're working with the PrexDigital Analytics tool.

Questions on a rating scale and open questions will be asked on how the current situation of data governance is at PrexPartners and how the proposed framework could (not) address the issue for the tool.

I will be sharing the results with you after collecting all the answers.

Thank you for your time!
J.Hu

Approx. 10 minutes.
40 questions total

1) What is your job title at the company?

- ◯ Partner
- ◯ Principal
- ◯ Project Manager
- ◯ Consultant
- ◯ Anders: _____

2) How many years of experience do you have with similar tools like PrexDigital Analytics?

- ◯ No experience
- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4+
- ◯ Anders: _____

3) Currently at my organisation, there is (...) alignment between the organisation's data governance framework and risk management & compliance for the PrexDigital Analytics tool.

	1	2	3	4	5	
no	○	○	○	○	○	complete

4) Assume that your organisation uses the proposed data governance framework in this situation. There is (...) alignment between the organisation's data governance framework and risk management & compliance for the PrexDigital Analytics tool.

	1	2	3	4	5	
no	○	○	○	○	○	complete

5) Explain scoring if needed.

Jouw antwoord

6) Currently at my organisation, there are key stakeholders from (...) business and IT domains involved in the data governance practices set by the organisation for the PrexDigital Analytics tool.

	1	2	3	4	5	
no	○	○	○	○	○	all

7) Assume that your organisation uses the proposed data governance framework in this situation. There are key stakeholders from (...) business and IT domains involved in the data governance practices set by the organisation for the PrexDigital Analytics tool.

	1	2	3	4	5	
no	○	○	○	○	○	all

8) Explain scoring if needed.

Jouw antwoord

9) (...) senior managers (project manager and up) support the data governance practices set by the organisation and consider customer data as an organisational asset.

	1	2	3	4	5	
No	○	○	○	○	○	All

10) Assume that your organisation uses the proposed data governance framework in this situation. (...) senior managers (project manager and up) support the data governance practices set by the organisation and consider customer data as an organisational asset.

	1	2	3	4	5	
No	○	○	○	○	○	All

11) Explain scoring if needed.

Jouw antwoord

12) When strengthening organisational practices and policies, a data governance framework is (...) considered.

	1	2	3	4	5	
never	○	○	○	○	○	always

13) Assume that your organisation uses the proposed data governance framework in this situation. When strengthening organisational practices and policies, the data governance framework is (...) considered.

	1	2	3	4	5	
never	○	○	○	○	○	always

14) Explain scoring if needed.

Jouw antwoord

15) Currently in your organisation, data stewards are (...) responsible for defining the attributes from the data of customers that is being used under their area of responsibility.

	1	2	3	4	5	
not	○	○	○	○	○	completely

16) Assume that your organisation uses the proposed data governance framework in this situation. Data stewards are (...) responsible for defining the attributes from the data of customers that is being used under their area of responsibility.

	1	2	3	4	5	
not	○	○	○	○	○	completely

17) Explain scoring if needed.

Jouw antwoord

18) Currently in your organisation, data quality issues are (...) documented and handled according to the specific data quality issue.

	1	2	3	4	5	
never	○	○	○	○	○	always

19) Assume that your organisation uses the proposed data governance framework in this situation. Data quality issues are (...) documented and handled according to the specific data quality issue.

	1	2	3	4	5	
never	○	○	○	○	○	always

20) Explain scoring if needed.

Jouw antwoord

21) Currently in your organisation, how do you rate the quality of the set metrics by your organisation used to evaluate the collection, use, storage and deletion of data from all sources of the PrexDigital Analytics tool?

	1	2	3	4	5	
Poor	○	○	○	○	○	Excellent

22) Assume that your organisation uses the proposed data governance framework in this situation. How do you rate the quality of the metrics used to evaluate the collection, use, storage and deletion of data from all sources of the PrexDigital Analytics tool?

	1	2	3	4	5	
Poor	○	○	○	○	○	Excellent

23) Explain scoring if needed.

Jouw antwoord

24) Currently in your organisation, sensitive customer data that the PrexDigital Analytics tool uses is (...) encrypted.

	1	2	3	4	5	
never	○	○	○	○	○	always

25) Currently in your organisation, do administrators, subcontractors or third parties have access to unencrypted sensitive customer data from the Prexdigital Analytics tool?

	1	2	3	4	5	
always	○	○	○	○	○	never

26) Assume that your organisation uses the proposed data governance framework in this situation. Sensitive customer data that the PrexDigital Analytics tool uses is (...) encrypted.

	1	2	3	4	5	
never	○	○	○	○	○	always

27) Assume that your organisation uses the proposed data governance framework in this situation. Do administrators, subcontractors or third parties have access to unencrypted sensitive customer data from the Prexdigital Analytics tool?

	1	2	3	4	5	
always	○	○	○	○	○	never

28) Explain scoring if needed.

Jouw antwoord

29) Currently in your organisation, the architecture of the (un)structured data systems in the PrexDigital Analytics tool that contains customer information is (...) on established standards.

	1	2	3	4	5	
not based	○	○	○	○	○	completely based

30) Assume that your organisation uses the proposed data governance framework in this situation. The architecture of the (un)structured data systems in the PrexDigital Analytics tool that contain customer information is (...) on established standards.

	1	2	3	4	5	
not based	○	○	○	○	○	completely based

31) Explain scoring if needed.

Jouw antwoord

32) Currently in your organisation, a data warehouse is (...) used to store technical metadata about the customer.

	1	2	3	4	5	
never	○	○	○	○	○	always

33) Assume that your organisation uses the proposed data governance framework in this situation. A data warehouse is (...) used to store technical metadata about the customer.

	1	2	3	4	5	
never	○	○	○	○	○	always

34) Explain scoring if needed.

Jouw antwoord

35) Currently in your organisation, is there a strategy to reduce the number of non-regulatory changes made to databases from the Prexdigital Analytics tool?

- ○ There is no strategy to reduce the number of non-regulatory changes made to databases.
- ○ The development of a strategy to reduce the number of non-regulatory changes made to databases has just started.
- ○ The development of strategy to reduce the number of non-regulatory changes made to databases is partly complete.
- ○ The development of a strategy to reduce the number of non-regulatory changes made to databases is complete.
- ○ The strategy to reduce the number of non-regulatory changes made to databases is complete and implemented.

36) Assume that your organisation uses the proposed data governance framework in this situation. Is there a strategy to reduce the number of non-regulatory changes made to databases from the Prexdigital Analytics tool?

- ○ There is no strategy to reduce the number of non-regulatory changes made to databases.
- ○ The development of a strategy to reduce the number of non-regulatory changes made to databases has just started.
- ○ The development of strategy to reduce the number of non-regulatory changes made to databases is partly complete.
- ○ The development of a strategy to reduce the number of non-regulatory changes made to databases is complete.
- ○ The strategy to reduce the number of non-regulatory changes made to databases is complete and implemented.

37) Explain scoring if needed.

Jouw antwoord

38) How do you evaluate your confidence level in answering all of the above questions?

	1	2	3	4	5	
Not confident at all	○	○	○	○	○	Very confident

39) Explain scoring if needed.

Jouw antwoord

40) Additional comments, feedback, concerns that might be usefull for this survey?

Jouw antwoord