



Utrecht
University



Framing the implementation of data governance platforms

A framework for the organizational implementation of data governance
platforms based on situational factors of the organization

Author: Iris de Wildt
Student number: 6259979

Master thesis
Business Informatics
Department of Information and Computing Sciences
Utrecht University

First examiner: Drs. N.A. (Nico) Brand
Second examiner: Dr. S.J. (Sietse) Overbeek

Internship Avanade
Supervisor: Peter Meij

July, 2022

Abstract

Data governance ensures high data availability, quality and security of data by specifying the decision rights and accountabilities for data. However, a data governance implementation is an extremely complex undertaking as many organizational areas are involved and there is no one-size-fits-all solution. Something that can help improve the data governance at an organization is introducing a data governance platform that automates data governance activities, as this can ensure better availability, consistency and scalability. In order to maximize the benefits a data governance platform can provide, it is important that the organization is mature enough in its data governance and that the implementation is tailored to the organization's specific situation. This research aims to design a framework that supports an organization in implementing a data governance platform by considering the situational factors of the organization, according to the contingency theory which considers that the most effective strategy to implement a system depends on the organization's environment. To design the framework, the research was executed in three phases. Firstly, during the problem investigation phase, current literature of data governance and change implementation was examined. Additionally, subject matter experts on the topic of data governance platform implementations were interviewed about their experience with such a project. Secondly, during the treatment design phase, the success factors and situational factors for the implementation of a data governance platform were derived from the literature and expert interviews. This resulted in two visual frameworks which can be used complementary, with one framework focussing on the process of implementation and the other focussing on picking an implementation strategy. Thirdly, in the treatment validation phase, the frameworks were validated by means of interviews. On the one hand, this took the form of case study in which practitioners from case study organizations were questioned. On the other hand, academics who had affinity with data governance were interviewed. From this, it could be concluded that the frameworks can be used as a communication tool and for deciding how and when to implement a data governance platform. This can give an organization more guidance during the implementation process and contributes to the current body of knowledge on the implementation of data governance platforms.

Keywords: data governance platforms - contingency theory - framework design - change implementation

Table of contents

1. Introduction	5
1.1 Context	5
1.2 Research problem	7
1.3 Scope	9
1.4 Contribution	10
1.5 Outline	10
2. Research approach	11
2.1 Research questions	11
2.2 Research methods	15
2.2.1 Literature review approach	17
2.2.2 Expert interviews with SMEs	21
2.2.3 Framework design	23
2.2.4. Framework validation expert interviews	23
2.2.5 Research overview	26
2.3 Validity threats	27
2.4 Conclusion	28
3. Results related work	29
3.1 Origin of data governance	29
3.2 Data governance fundamentals	30
3.2.1 Data governance definitions	30
3.2.2 Data governance building blocks	31
3.2.3 Related approaches to data governance	33
3.2.4 Data governance frameworks	34
3.2.5 Industry standards on data governance	36
3.3 Implementation fundamentals	36
3.3.1 Implementation definitions	37
3.3.2 Contingency & configuration approach	37
3.3.3 Implementation strategies	38
3.3.4 Change implementation frameworks	39
3.3 Data governance automation	41
3.3.1 Data governance platforms	41
3.3.2. Related tools	42
3.3.2 Vendors	43
3.5 Factors affecting the implementation of data governance	43
3.5.1 Success factors for the implementation of data governance	44
3.5.2 Situational factors for implementation of software systems	48
3.5 Framework design	51
3.6 Conclusion	53

4. Results expert interviews	54
4.1 Interviews overview	54
4.2 Interview analysis	57
4.2.1 Qualitative analysis	57
4.2.2 Quantitative analysis	62
4.3 Conclusion	65
5. Framework design	66
5.1 Success factors	66
5.2 Situational factors	67
5.3 Visualization	67
5.3.1 Framework A	67
5.3.2 Framework B	68
5.4 Conclusion	69
6. Framework validation	71
6.1 Validation interviews overview	71
6.2 Iteration 1	72
6.3 Iteration 2	75
6.4 Iteration 3	79
6.5 Conclusion	81
7. Discussion	82
7.1 Implications	82
7.2 Limitations	82
7.3 Future research	84
8. Conclusion	86
References	89
Appendices	95
Appendix A: Literature review	95
Appendix B: Expert interview questions	97
Appendix C: Checklist from literature for input expert interviews	99
Appendix D: Qualitative results of the interviews	102
Appendix E: Quantitative results expert interviews	111
Appendix F: Citations from interviews	119
Appendix G: Framework validation expert interview questions	145
Appendix H: Survey results validation interviews	148
Appendix I: Visual results of validation interviews	150
Appendix J: Points of improvements from validation interviews	154
Appendix K: Research planning	160

1. Introduction

This report starts with the context of the research project in section 1.1 in order to provide background information about the research topic of data governance platforms. In section 1.2, the research problem is described, where guiding frameworks for this research are introduced and the research gap is highlighted. Next, in section 1.3 the scope is addressed to give an idea about the focus of the research project. Moreover, the scientific and practical contributions are highlighted in section 1.4. Finally, section 1.5 gives an overview of the project planning of the research.

1.1 Context

By 2025, International Data Corporation (IDC) predicts that the volume of data being generated each year will grow from 33 to 175 zettabytes (ZB), with more data being stored in the cloud than traditional data centers (Reinsel, Gantz, & Rydning, 2018). With this growing amount of data every day, it is becoming of increasing importance for organizations to have data governance practices in place (Ladley, 2012). Data governance can be defined as the processes, organization, policies, standards, and technologies needed for ensuring the availability, quality, auditability and security of data (Panian, 2011). Data governance is especially of high importance now that organizations are becoming more data-driven and digital transformation is on the forefront of many agendas (Henriette, Feki, & Boughzala, 2016). Organizations need to manage large volumes and increasing complexity of data, which requires new solutions for organizing data (Al-Ruithe, Benkhelifa, & Hameed, 2019). Besides, the growing pressure of regulatory legislation such as the General Data Protection Regulation (GDPR) requires organizations to have compliance measures in place (Abraham, Schneider, & Vom Brocke, 2019). Due to risks such as privacy breaches, errors, and fraud, organizations have to be careful about how they treat their data. According to Ladley (2012), implementing technologies without the appropriate data governance policies can bring potential harm to a great deal of people, including the organization itself, its employees, and its customers. Not only does effective data governance ensure compliance with laws and regulations and ensure data security, it can also lead to an improvement of the quality and use of data (Alhassan, Sammon, & Daly, 2016). Organizations need to manage data as an asset, where data is treated as a resource which can provide business value (Khatri & Brown, 2010). As reported by Ladley (2012), the organizations that do succeed in implementing an effective data governance program gain business value, for example by eliminating compliance risks or reducing data quality issues. Other benefits that can be gained from data governance include improved decision making, more transparency in data and its supporting processes, higher data quality, and a reduction in costs due to more effective and efficient processes (Al-Ruithe et al., 2019). If governed correctly, data governance can provide an organization with a competitive advantage (Russom, 2008).

A survey by the research institute UBM in 2018 showed that 98% of the participating organizations considered data governance to be very important, even though 46% of the participants did not have a formal governance structure in place (Mahanti, 2018). Although many organizations are aware that they need some form of data governance, few make adequate use of the practices and knowledge to put in place a solid data governance program (Ladley, 2012). Therefore, most data governance programs lead to failures, and are stopped before benefits can be gained from it. An important cause of this is limited understanding from leadership about what data governance does

and what it contributes. The value potential in data governance needs to be realized, and there should be support and guidance from top-level management throughout all parts of the organizations (Al-Ruithe et al., 2019). Additionally, a data governance undertaking is costly and therefore requires a clear business case demonstrating the potential value that can be gained from it. If the direct costs outweigh the potential benefits, the data governance program can be stopped by the sponsor (Ladley, 2012). Besides that, Al-Ruithe et al. (2019) give as reason for the failure of data governance efforts that they are mainly driven by IT with rigid processes and fragmented activities that are carried out one step at a time. This is confirmed by Panian (2010), who adds that existing data governance practices mainly revolve around the software applications (e.g. ERP, CRM) instead of the data that is processed by these systems. A unified approach which also considers the business value is often not considered, resulting in the problem of a lack of business-IT alignment. Another factor that can result in the failure of a data governance program is the organizational culture (Tallon, 2013). There can be a lack of support from the rest of the organization if people are not fully committed to the data governance program, and this can be one of the most difficult parts of a data governance program to manage (Petzold, Roggendorf, Rowshankish, & Sporleder, 2020). Finally, Al-Ruithe et al. (2019) emphasize that there is no one-size-fits-all solution, and that implementing a data governance program requires an approach that best fits the characteristics of the organization.

Russom (2008) states that software automation of data governance by means of technology can potentially result in greater consistency, speed, transparency, accuracy, and scalability of data governance related processes. This can result in more effective and efficient data lifecycle management, better informed decision-making, and improved communication of information. Similarly, Petzold et al. (2020) describe that a growing set of data governance tool suites are beginning to automate data governance, and that due to the importance of data governance these solutions will be improved over time by having higher coverage and by being more cost-effective. However, Ladley (2012) emphasizes that one of the key problems of a failed data governance program is that technology is bought even though no clear plans have been made for how to use it. Numerous organizations try to address issues related to data governance simply by implementing systems with data governance functionalities. Nevertheless, studies have shown that organizational rather than technical implementation issues play an important role in the success of a data governance program (Wende & Otto, 2007). The system acquisition needs to be justified, meaning that it should be confirmed that the data governance program benefits from a system, and that the system can actually be implemented and effectively operated. An important factor that plays a role in this is the readiness of the organization for using a data governance solution (Ladley, 2012). The most straight-forward choice in a solution depends to a large extent on the applications that are already used within the organization. It is important that data governance solutions monitor cross-system data and discover cross-system relationships in the systems that the organization already uses (Russom, 2008). Only then can it bring the data governance benefits such as ensuring compliance by discovering exceptions to business rules across the organization's data.

To conclude, it is critical for organizations to have data governance practices in place due to the growing emphasis on compliance, data security, transparent processes, data quality, and the use of data as an organizational asset. Even though many organizations are aware of the importance of data governance, it often proves to be difficult to implement a successful data governance program. This is because of a lack of data governance understanding among leadership, limited sponsorship, a high IT focus leading to a lack of business-IT alignment, resistance to change, and no one-size-fits-all

solution. Data governance platforms can help make a data governance program successful, as the automation of data governance activities can ensure consistent and transparent use of data. However, in order to realize the benefits data governance platforms provide, the organization should be mature enough in its data governance practices to implement it and the organizational implementation should be tailored to the organization's situation. Thus, there is a clear need for adjusting the organizational implementation of a data governance platform to the context of the organization in question, which leads to the research problem as described below.

1.2 Research problem

Instead of only considering the IT perspective, data governance needs to be related to concepts from IT governance where the entire enterprise is considered together with organizational theories for how to implement such a data governance undertaking (Weber, Otto, & Österle, 2009). Cheong and Chang (2007) recognize the need for a data governance framework, as it supports collaboration between levels of the organization to manage enterprise-wide data, and it makes sure there is alignment between data-related programs and corporate objectives. Current data governance studies have looked into structuring or organizing data governance and giving an overview of its building blocks (Brous, Janssen, & Vilminko-Heikkinen, 2016). However, according to Otto (2011), few studies give an overview of what organizational aspects should be organized when implementing data governance, and what data governance processes entail when it is implemented. Most scientific studies have only looked at a single aspect of data governance such as decision rights and roles, but what currently seems to be missing is an approach that links multiple aspects for considering data governance best practices (Abraham et al., 2019). Moreover, Weber et al. (2009) emphasize that current data governance models presume data governance to be a one size fits all solution even though configurations can be specific to an organization, where factors such as competitive strategy and decision-making style should also be considered. Petzold et al. (2020) suggest that data governance needs to be applied iteratively and that there should be a focused implementation based on the domain. Governance priorities need to be tailored to the domain and data quality issues should be prioritized so that the most pressing issues are addressed first. Rather than having a fully integrated data set across the entire enterprise, some domains require a tailored approach where the data governance is configured to a dedicated platform (Petzold et al., 2020). Consequently, the intersection between the data governance undertaking and the already existing business initiatives and IT implementations should be considered (Russom, 2008). Russom (2008) also highlights that these intersections should be prioritized. Data governance cannot be applied all at once, and thus there should be a prioritized sequence of implementations based on the importance of the issue, the return on investment, and the willingness of involved people to participate. Peyret and Goetz (2014) emphasize the importance of prioritization but link this to the implementation of a data governance solution. They state that organizations that buy data governance tooling usually start with one module of the entire data governance suite and later expand this with other functionalities once the organization requires these, implying that the implementation of a solution is to a large extent dependent on the organization at hand. As Petzold et al. (2020) suggest, data-governance archetypes should be used to decide on the level of sophistication needed for implementing data governance, and it should be checked that current efforts are aligned to these standards.

Although much research has been conducted on the topic of implementation, implementing change still poses a problem for a great deal of organizations (Peper, van Gunsteren, van Linge, & Robben, 2014). The knowledge about change implementation as introduced in implementation science is applied too little in practice, which leads to a gap between research and practice as described by Peper et al. (2014). This gap results in a need for guidelines that can be used for implementing change. One of the ways to look at the implementation of changes is the contingency approach (Peper et al., 2014). The contingency approach describes that there is not one best way to organize an enterprise. By applying the contingency approach to change implementation for organizations, it considers which implementation strategies are effective under what circumstances. The theories that are part of the contingency approach focus on the different aspects of organizing, such as organizational structure, strategic management, leadership style. There is currently a high focus in the contingency approach on complex and hybrid organizational forms (Van de Ven, Ganco, & Hinings, 2013). This makes the implementation of change even more difficult, as there is not one best way of implementing due to changes inside or outside the organization. Therefore, there needs to be a search for a fit between the implementation of a change and the organizational environment. This fit is a dynamic concept, as changes in the environment continuously require a different approach to managing them. This means that the contextual and situational factors of the organization are important to consider. Although evidence can be found of concepts from the contingency theory being applied to IT governance, little research has been done about factors influencing data governance (Otto, 2011). There is opportunity to apply this in the domain of data governance, as Abraham et al. (2019) describe that there is no clear overview of the antecedents and consequences in data governance research, while this is needed to understand the effect of data governance practices.

To summarize, existing frameworks often do not provide concrete guidance for how the implementation of data governance and related systems should proceed in practice. Moreover, current data governance frameworks frequently neglect the complex and dynamic environment in which data governance is implemented which requires an approach that is tailored and prioritized to the domain. Therefore, it is useful to design a framework which considers the contingency approach that considers contextual and situational factors of the organization and which draws on principles of change implementation. Even though some examples of situational models can be found (Cameron & Quinn, 2006; Van Linge, 2006), there is opportunity to look at situational models for change implementation by applying it in the context of the implementation of data governance platforms. Consequently, the research objective is to support an organization in implementing a data governance platform by designing a framework which considers the organization's situational factors. This leads to the main research question: *MRQ: "What framework can be designed that supports an organization in the implementation of a data governance platform by considering the situational factors of the organization?"* In order to specify what this research question intends to answer and what is not considered part of this research question, the scope is described in the next section.

1.3 Scope

Data governance covers many processes and touches many areas of the organization. This study specifically focused on data governance platforms. Therefore, the aim of this study was to discover how data governance should be set up in order to be able to successfully implement a data

governance platform. The purpose was to determine what makes the implementation of a data governance platform a success, not what makes data governance in general a success. For the scope of this study, it was chosen to look at the organizational implementation rather than the technical implementation. This was because most data governance programs that fail stem from organizational issues (Ladley, 2012). Thus, most benefits could be gained by looking into how these organizational issues can be avoided. Which technical changes have to be made on a developer level are not considered in this research. In order to ensure that the outcomes of the research were also technically feasible, advice from experts and industry-standards were also considered. Additionally, for the scope of this study it was assumed that the organizations have some knowledge of data governance before they start implementing a data governance platform. A data governance program builds on the policies, standards, and procedures that are already in place in the organization (Ladley, 2012). Data governance systems should support these principles, and they should not be a goal in itself (Russom, 2008). Without any foundations for data governance in place, a platform cannot be implemented. Thus, organizations that do not have any plans for this were left out of the scope. Moreover, the guidance for organizations that want to implement a data governance platform was in the form of critical success factors. Critical success factors are factors that an organization needs to have in order to achieve its goals, and in the context of data governance implementation this means the factors that are a prerequisite for a successful outcome (Mahanti, 2018). As Ladley (2012) indicates that readiness of an organization plays an important role in the success of a data governance program, it was useful to investigate which success factors an organization should have depending on its characteristics in order to ensure an effective implementation. Data governance platforms are often implemented as part of a digital transformation. However, it was chosen to focus on the implementation of a data governance platform regardless of the bigger transformation process.

This research was conducted during an internship at Avanade. Consequently, for the practical part of the research, the organizations that were looked into and the interviewees who were questioned all had some link to Avanade. Avanade was founded in 2000 as a joint venture between Accenture and Microsoft (Avanade, 2022). As Avanade delivers solutions and services using the Microsoft platform, there was a strong focus on Microsoft Azure Purview for the leading data governance platform during this research. Azure Purview is a cloud based service that helps to govern on-premise, multi-cloud and software-as-a-service (SaaS) data (Microsoft, 2022). The first version of Purview was made generally available in 2020 and thus it is quite a new platform. Therefore, there was much opportunity to examine how this product can best be used. For the interviews and the case study, it was chosen to look at organizations that have implemented a data governance platform or who are currently in the process of implementing it. Organizations that have not started this transformation were not researched. This was because organizations that have already started the implementation or that have made plans for this can provide more detailed information about what made the platform a success. A large part of these organizations were looking into Microsoft Purview for the implementation of a data governance platform, so the outcomes of the data governance platform implementation were mainly guided by Purview. Some other data governance platforms were also considered in these organizations, but these considerations played a smaller part in the practical part of the research. However, in order to avoid bias due to the high focus on Purview, some mitigation tactics were applied that are elaborated upon in section 2.3.

1.4 Contribution

For the relevance of this research, a distinction can be made between the scientific and the practical contribution. First of all, the scientific contribution is the introduction of data governance platforms to the current theory on the foundations of data governance. Few papers have focused on data governance platforms to support data governance activities. Additionally, Alhassan et al. (2016) state that limited research has been done on the implementation part of data governance, which leads to a bridge between theoretical data governance foundations and data governance practices in a real-life setting. As Alhassan, Sammon and Daly (2018) mention, a practitioner perspective might prove to be a valuable addition to scientific literature. This study aims to combine both scientific and evidence-based approaches, and attempts to bridge the gap between theory and practice. It therefore adds to the current knowledge base on data governance by connecting data governance foundations to the implementation of related platforms.

As for the practical contribution of the research, the outcome can help an organization in determining when they should implement a data governance platform. Not all organizations can successfully implement a data governance platform, as readiness is an important factor that determines the success of such an implementation (Ladley, 2012). However, in order for organizations to remain competitive, it is important that they think about long-term solutions for the governance of the growing and changing amount of data. As volumes of data are growing at such a fast rate, new solutions for managing data are needed. Reinsel et al. (2018) stress that this growing volume of data also comes with a cost; maintaining and protecting it is more difficult. The pace in which digital transformation is happening is so high that it exceeds the ability to all handle this manually and therefore automation of governance activities becomes more important (IBM, 2018). Resultantly, it is essential for organizations to start thinking about new approaches for governing data. The understanding of how data governance platforms can be implemented in the organization can guide future decisions for example on strategy, data landscapes, and decision-making, in order to make sure that these organizations are mature enough to implement such a system.

1.5 Outline

In the remainder of the report, the justification and execution of the research is further discussed. In chapter 2, an explanation of the research methods is given, which describes how the design cycle of Wieringa (2014) is followed. Next, in chapter 3, the results of the literature review are given, and data governance and implementation fundamentals are addressed. In chapter 4, the results of the subject matter expert (SME) interviews are given, which includes an overview of which success and situational factors were derived from the interviews. Thereafter, in chapter 5, the design of the framework is described. Moreover, in chapter 6, a description is given of the framework validation interviews, with lists of possible points of improvements and redesigned frameworks. In chapter 7 the discussion of the research is given, which includes an evaluation of the research validity and an overview of future work. Finally, in chapter 8 the conclusion is given. In Appendix K, a Gantt chart of the project is given, which shows the planning for each of the phases of the research.

2. Research approach

Based on the formulation of the research problem in the previous chapter, the research questions and research methods were specified. In section 2.1, a description and a justification of the research questions is provided. These are then linked to the research methods that were used to answer them. In section 2.2, the research methods are described in detail. Finally in section 2.3, the threats to the validity of the research are discussed.

2.1 Research questions

The research framework for this study is given in figure 1, which is a framework that shows the steps that are taken in order to achieve the research objective and the knowledge that is necessary for this (Verschuren, Doorewaard, & Mellion, 2010). This research framework was used to specify the research questions. Firstly, for the information gathering phase, preliminary research, data governance theory, and change implementation theory provided the basis for deciding which definitions and frameworks to focus on. Secondly, definitions and frameworks from theory were part of the conceptualization phase, where the preliminary research ideas were translated into common concepts for this study. The definitions were used to determine which frameworks were suitable to use for this research and which frameworks were not. The result of this phase was an overview of the general implementation success factors based on the definitions and frameworks. Thirdly, in the investigation phase the general implementation success factors were combined with experience from practice that was gained by conducting expert interviews, so that specific data governance platform implementation success factors could be derived. These implementation success factors were used as input for the interviews, since they were used as guidance throughout the interviews to check whether they also applied to data governance platforms. Fourthly, situational factors derived from practical experience were linked to the data governance platform implementation success factors in the framing phase. Fifthly, in the design phase, these factors were visualized in a preliminary data governance platform implementation framework. This framework was validated by practical experience in a case study and with expert opinions of data governance academics. Resultantly, the validation phase resulted in the final data governance platform implementation framework. Figure 1 connects each output of the phases with the research questions and research methods.

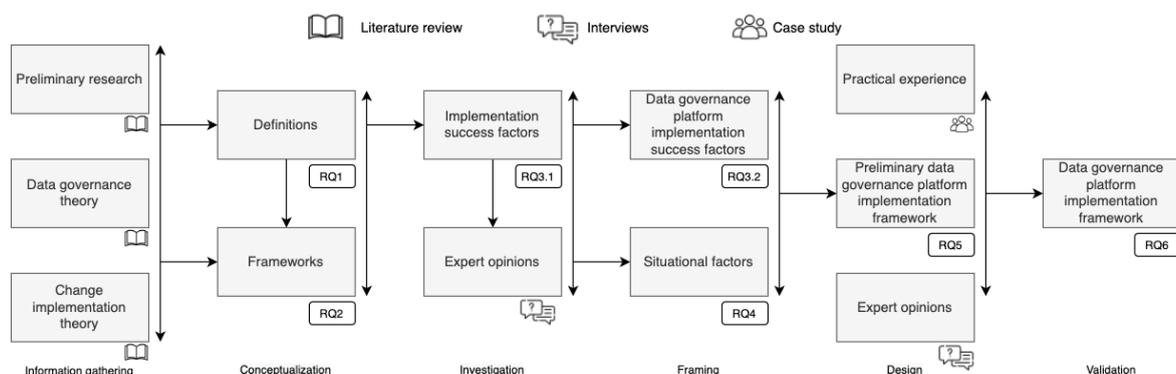


Figure 1. Research framework

Considering the research outcome as proposed by the research framework, the main research question of this study is as follows:

MRQ: "What framework can be designed that supports an organization in the implementation of a data governance platform by considering the situational factors of the organization?"

The purpose of the main research question (MRQ) of this study was to identify a data governance platform implementation framework that can help an organization in implementing a data governance platform. This framework was supposed to draw on models from contingency theories by considering situational factors of an organization. The resulting framework should be able to classify an organization in a category, and should provide a set of success factors that the organization should have in order to successfully implement a data governance platform.

In order to provide an answer to the central research question, the question was narrowed down to more concrete sub-questions. The aim of these sub-questions was to decide which knowledge needed to be gathered or which research activity was required in order to give an answer to the central research question.

RQ1: What does the implementation of data governance mean for an organization?

The first research question (RQ1) looked at the definition of data governance and its implementation. Due to the lack of understanding from leadership about data governance, it was important that the concept of data governance is well defined, including how it is applied in practice (Ladley, 2012). A common barrier to the success of data governance within an organization is the incorrect assignment of tasks to the data governance program due to unclarity of what they entail. Resultantly, a common and unified definition needed to be established. Moreover, phases of a data governance program, including its implementation and what processes are involved, depend on the organization at hand (Ladley, 2012). Consequently, it was important to determine what is considered part of the implementation in this study and what is not. This question was answered by performing a multivocal literature review, where the systematic literature review guidelines were followed as proposed by Kitchenham (2004) and extended with the multivocal literature review guidelines from Garousi, Felderer, & Mäntylä (2019). This was because both scientific and gray literature could be identified this way, so that the results were built on the theoretical as well as practical approaches. As a result of this question, the definitions that were used in this study were determined.

RQ2: What current frameworks that can be used for the implementation of data governance platforms exist?

RQ2.1: What are current data governance frameworks?

RQ2.2: What are current organizational change implementation frameworks?

There is scarce evidence in literature of how data governance platforms should be implemented and frameworks that support this. Therefore, the research question was split into two parts, in order to look up which frameworks exist for data governance (RQ2.1) and which frameworks exist for change implementation (RQ2.2). This provides a basis for a framework that draws on both data governance and change implementation principles. Most frameworks developed for data governance are used for building blocks of data governance and not necessarily its implementation thereof (Brous et al., 2016). Resultantly, it was useful to compare theoretical frameworks from literature with

practice-oriented frameworks that are more implementation focused (Al-Ruithe et al., 2019). Therefore, in order to answer this research question a multivocal literature review was used which considered both scientific literature and gray literature (Kitchenham, 2004; Garousi, Felderer, & Mäntylä, 2019). By comparing the similarities and differences of the framework, an overview was given of which elements from the frameworks an organization needs for data governance as well as the implementation of new technologies. This resulted in a list of general success factors that can be used for the implementation of data governance.

RQ3: What critical success factors in the organization are needed for the successful implementation of a data governance platform?

RQ3.1: What are general success factors needed for the successful implementation of software systems?

RQ3.2: What are success factors specific to the implementation of data governance platforms?

Petzold et al. (2020) mention that a successful implementation of data governance requires guidance on how an organization can ensure the successful implementation of a data governance program. If the existing infrastructure of the organization is not mature enough to support a data governance program, then there is opportunity to implement a new organizational infrastructure (Panian, 2010). Therefore, it was useful to look into which factors make the implementation of a governance platform successful in the form of critical success factors (RQ3). Critical success factors can be described as factors that are critical for the organization to have in place to ensure a successful outcome, meaning that without these factors the organization fails to meet its objectives (Mahanti, 2018). If the organization does not ensure these critical success factors are in place, then it needs a different approach before implementing the data governance platform. There is little evidence in literature of the implementation of data governance in practice (Brous et al., 2016). Likewise, there is scarce evidence of data governance platforms in literature, since suites of tools for data governance are still new and will evolve drastically in the future (Petzold et al., 2020). Consequently, as there is not enough evidence in literature to answer this research question, this question was split up in two parts. One question focuses on general implementation success factors as found in literature (RQ3.1), and the other question focuses on implementation success factors specific for data governance platforms supported by practice-based evidence (RQ3.2). The first research question (RQ3.1) was answered by conducting a multivocal literature review (Kitchenham, 2004; Garousi, Felderer, & Mäntylä, 2019). The second research question (RQ3.2) was examined by conducting expert interviews, based on Hove and Anda (2005). The experts for the interviews had extensive knowledge of the implementation of data governance platforms or related data technologies and had experience with the implementation of these platforms in organizations. The output of this research question was a set of success factors that are needed for the implementation of a data governance platform.

RQ4: What situational aspects of the organization have the biggest influence on how a data governance platform is implemented?

In order to gain insights into what success factors are most critical to have in place for a given organization, the fourth research question (RQ4) focuses on the areas of the organization that require immediate attention when implementing a data governance platform. As Petzold et al. (2020) mention, data governance should be prioritized to the domain and should be implemented through quick iterations in order for it to provide the most value as it cannot be implemented all at

once. Therefore, there is opportunity in looking at what characteristics of the organization have a considerable effect on the implementation efforts of a data governance program, so that an indication can be given of what areas of the organization to focus on first. This research question was answered by conducting expert interviews (Hove & Anda, 2005). The experts that participated in the interviews had first-hand experience with the implementation of a data governance platform in a practical setting, and as a result have gained insights into where problems occur during such a program. As a result of this research question, a list of situational characteristics was made that links the success factors to the characteristics of the organization.

RQ5: What framework can be developed that can support an organization in implementing a data governance platform?

In order to bring together the critical success factors into one overview and help the organization decide which success factors to focus on, a framework was developed for the fifth research question (RQ5). This framework should take the contingency approach into account which considers that there is not one solution (Peper et al., 2014). This framework should serve as a foundation to narrow the search for situational factors that affect the implementation of a data governance platform, as it is impossible to include every single factor when assessing the organization (Cameron & Quinn, 2006). The framework was developed by following the design science methodology cycle from Wieringa (2014). The output of this research question was a framework which can be used to classify an organization in a stage or phase based on some characteristics of the organization, which gives as a result a set of critical success factors that the organization should have in order to implement a data governance platform.

RQ6: How can the data governance platform implementation framework be validated?

The sixth research question focused on validating the framework, which was done with both practitioners and academics in iterations. Firstly, in order to test whether the data governance implementation framework could be used in practice, the framework was validated with organizations that had implemented a data governance platform or were thinking of implementing it. Factors that originated from literature could be tested with practitioners this way. For this, case study research was used which allows the study of contextual factors when testing an artifact (Haiz, 2008). To gather data for this, interviews were conducted with data governance practitioners from organizations that were part of the case study. Secondly, in order to test whether the practical experience from the interviews with the subject matter experts also applied to concepts from theory, expert interviews were conducted with academics (Hove & Anda, 2005). This way, there was a triangulation of research methods, which combines multiple methods or data sources to test the validity of the research through different sources in qualitative research (Carter, Bryant-Lukosius, DiCenso, Blythe & Neville, 2014). Data source triangulation, which collects data about different kinds of groups (i.e. organization vs researchers) was done to ensure both the practical as well as the academic perspective was covered. Moreover, investigator triangulation, which involves multiple participants, was applied to ensure that multiple perspectives within these groups were covered. The result of this research question was a discussion of the points of improvements of the framework. The steps for this research question were executed in iterations. Designing a framework in iterations can help identify the most appropriate framework and increase its utility (Wynn, Eckert, & Clarkson, 2007). During each iteration, an interview was conducted with a practitioner at an organization and

an academic. The results from both these interviews were combined and outcomes were applied to the framework. The new iteration started with an improved version of the framework.

In table 1, the research questions together with the approach to answer them are summarized. For each research question, the research method is specified. Besides that, the intended outcomes of the research questions are given, next to the section of the report where they are discussed.

Table 1. Research questions and methods

#	RQ	Research method	Outcome	Chapter
1	What does the implementation of data governance mean for an organization?	Multivocal literature review	Definition of data governance implementation	3
2.1	What are current data governance frameworks?	Multivocal literature review	List of success factors for data governance and change implementation	3
2.2	What are current organizational change implementation frameworks?			
3.1	What are general success factors needed for the successful implementation of software systems?	Multivocal literature review	List of implementation success factors	3
3.2	What are success factors specific to the implementation of data governance platforms?	Expert interviews	List of data governance platform implementation success factors	4
4	What situational aspects of the organization have the biggest influence on how a data governance platform is implemented?	Expert interviews	List of situational factors	4
5	What framework can be developed that can support an organization in implementing a data governance platform?	Design science methodology cycle	Data governance platform implementation framework	5
6	How can the data governance platform implementation framework be validated?	Case study, expert interviews	List of points of improvements	6

In the following section, the research methods as noted in table 1 are discussed according to the three phases of the design cycle approach from Wieringa (2014).

2.2 Research methods

This research follows the design cycle approach as proposed by Wieringa (2014). It has been chosen to follow this design cycle due to its focus on the investigation of an artifact in context. Since this research aims to bridge the gap between theory and practice by proposing a new framework based on practical experiences (see the research framework in figure 1), the context is an important factor in this research. This design cycle iterates over three tasks, which are the problem investigation, treatment design, and treatment validation. The artifact of this study is a data governance platform implementation framework. The treatment is the implementation of a data governance platform according to this framework. The design cycle is part of a larger cycle called the engineering cycle, which also includes treatment implementation. It is chosen not to implement the treatment because

the focus of this research is on the design of a framework, not the use thereof. However, future research could look into the implementation of this framework in a real-life setting. The phases according to the design cycle as used in this research are visualized in figure 2.

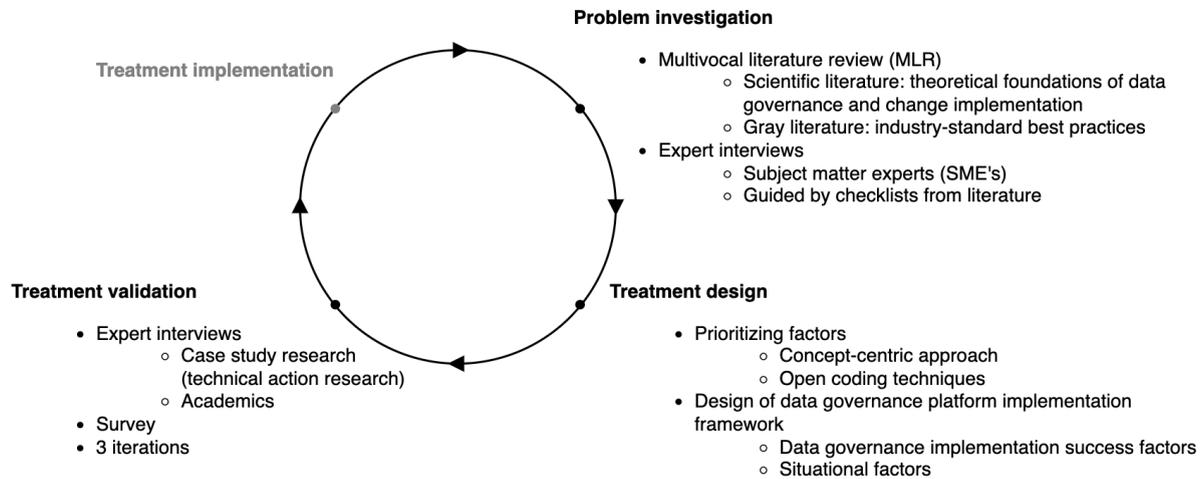


Figure 2. Phases and research methods of the research approach according to (Wieringa, 2014)

In the problem investigation, knowledge is gathered about how the problem can be treated (Wieringa, 2014). The problem of this research is that data governance platform implementations often fail due to the complexity of the change that is needed within the organization to implement it. A data governance platform can help make data governance more effective and efficient, but it is important the organization is ready to use it and knows how to use it (Ladley, 2012). In order to understand when and how a data governance platform should be implemented, more information was gathered about data governance, change implementation, and related data platforms. This was done by conducting a literature review. This literature review followed a multivocal literature review (MLR) to gather more information about the theoretical foundations of data governance and change implementation, as well as gather more about best practices from industry standards. Moreover, during the problem investigation practical knowledge was gathered about the implementation of data governance platforms by conducting expert interviews so that the theory could be connected to practice. During the treatment design phase, the artifact is designed which treats the problem as identified in the problem investigation (Wieringa, 2014). In this research, the to-be-designed artifact was a data governance platform implementation framework. In this phase, the concepts used for the framework were determined by means of the concept-centric approach and open coding techniques. With the help of these techniques, the success factors and situational factors were derived from the literature and interviews and the framework was designed. In the treatment validation phase, it is validated whether the design from the design phase actually treats the problem (Wieringa, 2014). In this phase, it needs to be justified that the treatment would contribute to the stakeholder goals if it were to be implemented. In this research, the treatment validation consisted of case study research and expert interviews with academics. The intention of the case study research was to validate with stakeholders of the framework if it could be used in a real-world setting (technical action research), and data were gathered for this by means of interviews with data governance stakeholders from organizations which could benefit from the framework. The purpose of the expert interviews with

academics was to test whether the framework was in line with concepts from theory. Points of improvement were applied to the framework for each iteration of the treatment validation.

2.2.1 Literature review approach

A multivocal literature review (MLR) was conducted in order to provide background information about data governance and its implementation. A MLR extends a systematic literature review (SLR) by also including gray literature in the search next to published scientific literature (Garousi et al., 2019). This means that in addition to the formal published literature on the topic, other forms of literature such as white papers and reports were also looked up which were more practice-oriented. A SLR allows a more complete search compared to a regular literature review since it uses a predefined search strategy (Kitchenham, 2004). By conducting a SLR, a wide range of studies are examined which gives more robust and reliable results, thereby reducing bias. It was chosen to extend the SLR by conducting a MLR so that literature published by industry standards could also be considered. Various industry standards can provide more guidance for the implementation of data governance instead of the focus on the conceptual building blocks that is frequently found in scientific literature (Al-Ruithe et al., 2019). Moreover, including MLR can reduce publication bias and can provide a different perspective to the regular SLR (Kitchenham, 2004). The procedures for a SLR from Kitchenham (2004) were used, and they were combined with additional procedures for conducting a MLR according to Garousi et al. (2019). This is because the process for conducting a MLR as described by Garousi et al. (2019) is based on the procedures from Kitchenham (2004), and thus the search process only needed to be extended instead of replaced when also including gray literature. The intention of the MLR was to identify what is already known about the topic, and what needs to be researched in the rest of the study.

The aim of the literature review was to answer the first and second research questions of this research. For the first research question, definitions were established so it is clear what is meant with the implementation of data governance in this research. Preliminary research showed there is little evidence of the implementation of a data governance platform since these platforms are just recently introduced in organizations (Petzold et al., 2020). Consequently, the literature review yielded no search results of complete data governance platform implementation frameworks. Thus, the search for frameworks for the organizational implementation of a data governance platform was split up into two main parts; data governance frameworks and change implementation frameworks. For the second research question, the definitions were used to decide which frameworks to use from data governance literature and which frameworks to use from change implementation literature. Additional literature was gathered from industry standards that connected both data governance and change implementation.

Review protocol

For a MLR, a review protocol is needed that specifies the methods that are used in order to avoid research bias. The SLR components for a review protocol as described by Kitchenham (2004) were used, and they were extended by guidelines from Garousi et al. (2019) for a MLR. The following guidelines were used to search for the formally published scientific literature:

- **Background:** The rationale for this SLR was to gain a better understanding of data governance and change implementation definitions and building blocks, so that this

knowledge could be used as input for a framework that covers the implementation of a data governance platform.

- **Research question:** As noted in table 1, the SLR aimed to answer:
 - RQ1: How can the implementation of data governance be defined?
 - RQ2.1: What are current data governance frameworks?
 - RQ2.2: What are current organizational change implementation frameworks?
 - RQ3.1: What are general success factors needed for the successful implementation of new systems?
- **Search strategy:** For the search strategy, a decision has been made which resources (i.e. search engines) to utilize, and which search terms to use.
 - **Resources:** The main search engine that was used for finding scientific sources was Google Scholar. Google Scholar is a scientific search engine that consistently shows higher retrieval than other related databases, and is therefore a suitable tool for SLR to ensure retrieval of all relevant articles (Cecchino, 2010). In order to reduce the bias of using just one search engine, additional use was made of Mendeley to find sources. By using this search engine, other sources than the ones found on Google Scholar could be included in the SLR, as Mendeley also takes into account the readers of the articles when providing search results (Naude, 2017).
 - **Search terms:** The keywords used for the search were derived from the research questions. First, broad keyword terms such as “data governance” and “change implementation” were used in preliminary research to have an idea of the concepts used in the field. These keywords were refined accordingly to find more sources relevant for this specific study. In order to refine these keywords, Boolean operators were used in the search query, where the “AND” boolean was used to join main terms, and the “OR” boolean was used to find synonyms (Xiao & Watson, 2019). From the initial search it became clear that multiple terms were used in literature to refer to data governance information systems, including tools, platforms, and technologies. Moreover, frameworks and models were sometimes used interchangeably in literature. To include all of these, the “OR” boolean was used. Also, in order to find sources that mentioned the systems separately from data governance or change implementation, the boolean “AND” was used. This was because the purpose was to find the requirements for the implementation of a new system, not an implementation system itself (i.e. a system that is used to support another system’s implementation). Resultantly, this led to the following selection of search terms: *“Data governance (framework OR model)”*, *“Data governance implementation”*, *“Data governance AND (tools OR platform OR system OR technologies)”*, *“Implementation (framework OR model)”*, *“Change implementation (framework OR model)”*, *“Change implementation AND (tool OR platform OR system OR technologies)”*.
 - **Snowballing:** Additional sources were found by making use of forward and backward snowballing, which means that the reference list and citations of one particular source were examined in order to identify further sources (Wohlin, 2014). The backward snowballing method was applied whenever a source needed more explanation or more in-depth information, for which the reference list of the source

was checked. Besides, the forward snowballing method was applied to sources that were recognized as highly relevant for the research questions in order to build on top of these theories. This was done by checking the citations of the source.

- **Study selection criteria and procedures:** The abstract was read in order to determine which sources were suitable for inclusion in this study. The following criteria were used to determine whether to include a source in the search or whether to exclude it:
 - The sources should be written in a language interpretable to the author. Search queries were performed in English, so mostly English sources were retrieved. However, when encountered, sources written in Dutch were also accepted.
 - The sources should be published after the year 2005. This was done so that some older articles were included (i.e. older than 10 years) in order to have an idea of the evolution of sources on data governance and change implementation, as well as newer sources (i.e. newer than 10 years) to be able to build on the most up-to-date literature.
 - The sources should be from the information system or the change implementation domain. Some searches yielded systems or platforms (e.g. ecological systems) that had nothing to do with these domains, and they were excluded from the SLR.
 - The sources should focus on the implementation of platforms in an organization, not platforms made to facilitate change in an organization. Sources that described systems used to support another system's implementation or other implementation in an organization were excluded.
 - The sources should answer one of the research questions. This means that sources which provided definitions, frameworks and success factors were accepted.
 - The sources should have as a unit of analysis an organization, and more specifically an enterprise. Studies that have looked at public sector institutions, non-profit organizations or tools for individuals were excluded.
- **Study quality assessment checklist and procedures:** The quality of the articles was assessed by skimming through the texts. The importance of each of the studies was kept track of by weighting the studies qualitatively (Xiao & Watson, 2019). This meant the studies were placed in a category ranging from 1-3, where 1 was most relevant and 3 was least relevant for this study. The following characteristics of the source were taken into account for determining the quality category:
 - The source's relevance to the research questions from this study. Sources who directly seemed to answer research questions scored higher.
 - What research methods were used to form a conclusion. Studies that conducted multiple research methods were seen as more reliable and thus scored higher.
 - The number of citations to this article. This was seen as less relevant for studies with a more recent publication date, as this naturally equals a lower number of citations.
- **Data extraction strategy:** The extraction of data from articles was done systematically by noting for each paper its characteristics and relevant information in an Excel spreadsheet. A preview of this spreadsheet can be found in Appendix A, together with the data extraction form. The following columns, as further described in the data extraction form, were kept track of for each retrieved article:

- Meta-data about the source: title, authors, year, main topic, type (e.g. article, conference paper, book, etc), publisher (in case the paper was published by an institution) and number of citations.
- Search engine information: resource (e.g. Google Scholar), hyperlink, keywords used for search, and snowballing (i.e. backward/forward snowballing and from which article).
- Content of the paper: relevance to the topic (on a score from 1 to 3 with 1 having the highest relevance), reason for inclusion or exclusion, and a summary of relevant information from the abstract.
- **Synthesis of extracted data:** The included studies were synthesized by means of summarizing the results in a descriptive (i.e. non-qualitative) way. Sources that were placed in a higher relevance category were used as foundational literature, while sources with a lower category were used as supplementary material in case more explanation was needed.
- **Project timetable:** The SLR is conducted at the beginning of the project before the interviews in order to have the theoretical foundation of the research in place. Theory can then be tested in a practical setting.

As mentioned by Garousi et al. (2019), the standard review protocol structure from a SLR according to Kitchenham (2004) can also be applied to a MLR. Therefore, for reviewing the gray literature, the SLR review protocol as described above was used. However, the following alterations to the SLR review protocol were considered for including gray literature:

- **Background:** The MLR was intended to provide a more practice-oriented perspective on the topic of data governance. The intention was to look for publications on data governance from industry standards, and look for publications on data governance platforms.
- **Search strategy:** The search strategy for the MLR was altered so that only targeted websites were searched for publications on data governance. These had the following consequences for the search strategy:
 - **Resources:** The general web search engine Google was used. In order to identify which sources were needed from Google, relevant industry standards on the topic of data governance were retrieved from the literature gathered in the SLR. Moreover, sources that were recommended by experts (i.e. people familiar with the domain of data governance) were also assessed.
 - **Search terms:** The following industry standards were searched on Google for publication of white papers or reports on data governance: “DAMA”, “ISACA”, “Data governance institute (DGI)”, “Data warehouse institute (TDWI)”, “ISO 900”, “IBM Data Governance Council”, “Informatica”, “Microsoft”, “International data corporation (IDC)”, “McKinsey”, “The Forrester Wave”
- **Data extraction strategy:** Each paper was kept track of by adding it to the spreadsheet (Appendix A) as described in the SLR review protocol. However, instead of the number of citations, “MLR” was noted in this column.

Results

In total, 62 sources were reviewed for the literature research. Of these 62 sources, 31 were written on the topic of data governance, 21 were written on the topic of change implementation, and 10 sources included gray literature on the topic of data governance platforms. In figure 3, an overview

can be seen of the number of articles reviewed for the literature research per year, with a distinction between the three kinds of keyword searches. As can be seen from the bar chart, most publications originated from the year 2018 and later. However, some publications before these years have also been considered.

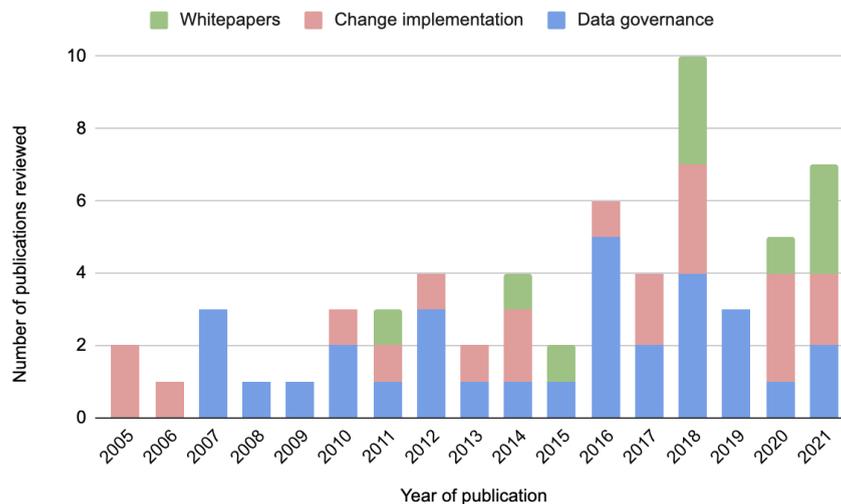


Figure 3. Number of sources reviewed for literature research per publication year

2.2.2 Expert interviews with SMEs

The literature review showed that little evidence can be found about data governance platforms as it is still a new area where lots of developments are being made (Petzold, 2020). In order to apply the outcomes of the literature review to data governance platforms, expert interviews were conducted with subject matter experts (SMEs) who had experience with data governance programs and/or who had extensive knowledge of data governance platforms. Expert interviews are useful in research that relies on qualitative information, and these interviews can be used in software engineering research to clarify data, obtain opinions, identify benefits and difficulties, and determine reasons behind projects (Hove & Anda, 2005). Therefore, the expert interviews provided a way to gather qualitative information on what makes a data governance platform a success at an organization.

Participants

In total, 11 SMEs participated in the interviews who had experience with data governance programs at organizations, and could tell more about the challenges and successes in such a program. The purpose of these interviews was to obtain more information about a data governance platform implementation in a practical setting, and therefore the interviewees were asked about an example case organization where a data governance platform was implemented. For all interviewees, the following information was gathered: role and department, experience with data governance, and if applicable the data governance programs or projects on which they worked. The interviewees were recruited by contacts from Avanade. Avanade provides advice for its clients on the data governance platform Microsoft Purview after it was introduced in 2020. Therefore, most of the interviewees had experience or extensive knowledge of this particular platform and resultantly, this platform was the main focus of the interviews. However, as the Purview platform is newly released, most interviewees also had extensive experience with other kinds of platforms such as Informatica and Collibra, and

during the interviews they were asked how Purview compared to these alternatives. As Avanade is a global organization, backgrounds of the interviews varied. The Avanade Netherlands section was the starting point, so most interviewees came from Europe, but there were some interviewees who came from the United States or Australia. Moreover, the interviews had varying roles within Avanade, from analytics architects to technology leadership. The aim of this variation was to gain different insights and perspectives on the implementation of data governance platforms.

Procedure

The experts were asked questions according to semi-structured interviews, where the interviewer aims to elicit information by asking predefined questions, but also offers the interviewer the chance to explore issues further by asking additional follow-up questions (Longhurst, 2003). Semi-structured interviews make the interview flow more like a natural conversation and allows the interviewer to delve into unforeseen topics (Adams, 2015). The semi-structured expert interviews were scheduled for half an hour. Most interviews were conducted online, since participants came from different geographical areas. Audio taping was used during the interviews, which is the suggested method for retrieving the most accurate data (Hove & Anda, 2005). The sets of open-ended questions that were prepared before the interviews took place can be found in appendix B.

The interview started with an introduction of the interview and a brief explanation of the purpose of the research, namely to determine the factors that lead to the successful implementation of a data governance platform in an organization. Next, the agenda of the interview was briefly introduced. Thereafter, it was explained how the results of the interview were going to be processed, and that no confidential information was going to be used for the report without consent. The interviewees were asked if they agreed with the interview being recorded and if they gave their consent to the processing of the results before continuing. Besides that, during the introduction phase of the interview, the interviewees were also asked about their role and their experience with data governance platforms. After that, the interview continued with the list of pre-prepared questions from appendix B. Besides the predefined set of questions, additional questions were asked based on the interviewee's answers if a topic required more explanation or clarification. Interview prompts were used to elicit more elaborate answers, such as asking "Why is that?" or "Could you expand on that?" (Adams, 2015). Due to the difference in role or data governance experience of the interviewees, sometimes the interview questions deviated from the pre-prepared questions in case the question was not applicable to the interviewee. In order to make sure that the interviews did not go overtime, questions were prioritized prior to the interview, so that some questions could be skipped in case the interview was taking too long.

In order to be able to compare the answers given by the interviewees, a grouping was made of possible constructs by categorizing answers and options before the interviews took place. Based on the literature review, concepts were identified that should be considered for the implementation of data governance. These were used as a form of a checklist to guide the interview answers. The list with factors derived from the literature review and used as input for the interviews can be found in appendix C. After the interviews were conducted, they were analyzed both qualitatively and quantitatively. The qualitative analysis comprised the identification of the most relevant answers from the interview. These were summarized in a description and reported in chapter 4 about the interview results (see Appendix D). In order to identify the success factors for the framework design, the interviews were also quantitatively analyzed by assigning codes to the interview answers. This

way, it could be counted how many times the same category of answers occurred in all interviews (see Appendix E). More about the coding technique is discussed in the next section about framework design, where it is described how the dimensions for the framework were prioritized.

2.2.3 Framework design

The design of the framework was done in different steps. First, the factors that should be included in the framework were determined. This was done by (1) deriving the general implementation success factors for data governance from literature, (2) deriving the implementation success factors for data governance platforms from the expert interviews, and (3) deriving the situational factors of the organization that affect the implementation of a data governance platform from the expert interviews.

Different techniques were used for prioritizing the most important factors for inclusion in the framework. Firstly, the concept-centric approach from Webster and Watson (2002) was used to synthesize the information from the literature review and determine which data governance implementation success factors should be taken into account. For this, the key concepts from the literature studies were identified and for each of these concepts, it was noted in which articles they occurred. The concepts that were identified in most studies were deemed as more relevant. Secondly, the data governance platform implementation concepts and the situational factors derived from the expert interviews were identified by using open coding analysis techniques (Corbin & Strauss, 2008). Answers were grouped by creating “codes” to label them. For every code, it was noted in which interviews they occurred. This way, it could be counted how many times the codes occurred in total. The concepts that were mentioned by most experts gained a higher importance score. They were compared with the success factors that were derived qualitatively, to decide which success factors were most important.

In order to visualize the concepts in a framework, the situational factors with the highest relevance were used to determine the dimensions of the framework. The success factors as identified in the results were then grouped according to this dimension.

2.2.4. Framework validation expert interviews

The framework was validated in two ways. Firstly, it was validated by means of a case study in which interviews were conducted with stakeholders from the organizations that participated in the case study. Secondly, expert interviews were conducted with academics who were knowledgeable in the field of data governance. This resulted in the use of triangulation, which means that multiple results or data sources were used in order to increase the validity of the research (Carter et al., 2014). The framework was validated in iterations, where each iteration consisted of an interview with a case study organization and an academic. This resulted in data source triangulation for each iteration, as both insights from organizations and researchers were used. Moreover, by executing multiple iterations, investigator triangulation was applied so that there were different perspectives within the same group of data sources.

Case study

A case study was chosen as one of the methods for validating the framework. This was because contextual factors of the organization played an important role in the development of the

framework, and a case study allows the studying of artifacts in their context (Hafiz, 2008). Contrary to other types of research, cases are studied one by one in case study research (Wieringa, 2014). The unit of analysis of the cases in this research were organizations; more specifically enterprises that were undertaking a data governance program and were implementing (or considering the implementation) of a data governance platform.

Data for this case study were gathered by means of interviews with stakeholders from the participating organizations. During the interview, the data governance background of the organization was discussed so more context could be discovered about the organization. The case study took the form of technical action research, which in the context of the design cycle means that a newly designed artifact is validated in the field by using it with the goal of helping a client (Wieringa, 2014). Technical action research is frequently used in combination with the design cycle research where an artifact is to be designed, and therefore it is a suitable method to test the framework that had been developed in the design phase. With technical action research, organizational problems are created that can be solved with the designed artifact (Wieringa & Morali, 2012). In the case study interviews, this was applied by asking the interviewee to classify the organization in the framework in order to give a set of critical success factors that should be in place.

The aim was to include organizations from different sectors, as to cover different types of organizations so that this could be taken into account for the results. These organizations were reached by asking the SMEs from the expert interviews about whether their example organization as the focus of the interviews could be contacted. Besides, personal contacts with organizations that were thinking about data governance platforms were explored. In total, three organizations were reached. These organizations were all at different stages with their implementation of their data governance platform.

Expert interviews with academics

Next to how the framework can be applied in practice, the framework was also validated on its connection to theory. The expert opinions of practitioners were tested by involving academics in the framework validation by conducting expert interviews. It was chosen to do this in order to take triangulation of the research into account (Carter et al., 2014).

Since data governance platforms are an emerging topic where lots of developments are being made, there is a lack of evidence of publications on this topic. Thus, instead of only considering academics with published papers about data governance platforms, academics were also considered who had affinity with data management, data security, or data privacy. Moreover, practitioners with an academic background were also considered as these experts could provide opinions specific to data governance platforms. The academics were personally contacted by asking whether they were open to an interview.

Procedure

The semi-structured interviews with the interview prompts as described in section 2.2.2 were used as a way to conduct the interviews (Hove & Anda, 2005; Adams, 2015). Similarly to the previous interview round, questions were prioritized in case there was not enough time to ask all of them. The interviews consisted of three parts: (A) an introduction part to gather background information about the interviewee and/or the organization, (B) a validation part to test the critical success and situational factors, and (C) a validation part to test the visualization of the framework. In Appendix B,

the interview questions for the semi-structured interviews can be found. The open questions during the first part (A) differed for the organizations and the academics. The second part (B), the validation of the success factors, was the same both for the practitioners as well as the academics. The third part (C), the validation of the data governance platform implementation framework, was only done with the participants who were experts on data governance platforms, as participants with limited background herein could not accurately review this.

All questions that were asked during the interviews are given in Appendix G. Firstly, for the introduction part (A), open questions were drawn up. After that, the validation of the success and situational factors took place (B). The participants were asked to rate these factors on a 5-point Likert scale based on how much they thought these factors were relevant for the implementation of a data governance platform. A 5-point Likert scale provides a suitable way to measure the opinions and perceptions of an individual (Joshi, Kale, Chandel, & Pal, 2015). Additionally to the 5-point Likert scale, there was an option for “Not sure” in case the participant had not enough knowledge to answer the question properly. Next, the framework was validated (C). The purpose of the designed framework was explained and the use of the framework was demonstrated. A list of statements was drawn up based on evaluation criteria, and after the demonstration of the framework, the participants were again asked to rate these statements on a 5-point Likert scale, with an option for “Not sure”. The evaluation criteria used for evaluating the framework, shown in table 2, was based on the criteria for implementation evaluation and criteria for assessing quality of theories.

Table 2. Evaluation criteria for the framework

Criteria	Definition	Source
Clarity of concepts	To what extent the guidelines and success factors are clear	(Davis et al., 2015)
Clarity of relationships	To what extent the relations between concepts are clear	(Davis et al., 2015)
Accuracy of classifications	To what extent the guidelines / success factors are accurate for the chosen classification	(Davis et al., 2015)
Understandability	To what extent the framework is understandable	(Hong & Kim, 2004)
Applicability	To what extent it can be applied to the organization	(Birken et al., 2018)
Usability	To what extent it uses relevant constructs	(Birken et al., 2018)
Robustness	To what extent it handles exceptional situations	(Hong & Kim, 2004)
Generalizability	To what extent it can be generalized to other organizations	(Davis et al., 2015)

During the interviews, the participants were presented with a survey made in Google forms to fill in for the validation part (B and C), and they were asked to give any additional comments of why they filled in a certain score. The explanations provided by the interviews were combined with the results from the survey, which led to a mixed-method approach. A mixed-method approach combines both qualitative and quantitative results in order to improve the scope of the analysis so that better conclusions can be drawn from it (Sandelowski, 2000).

Iterations

This interview round was conducted in iterations, where after each iteration the feedback from the previous interviews were applied. This way, the most appropriate framework could be designed as developing a framework in iterations can help increase its utility (Wynn et al., 2007). Each iteration consisted of an interview with an expert from an organization (i.e. practical validation) and an expert on data governance or a related area (i.e. theoretical validation). In total, three iterations of interviews took place with each iteration having two interview participants, so there were six participants in total.

2.2.5 Research overview

In figure 4, an overview is given of the research approach. The gray boxes indicate the phases of the design science cycle from Wieringa (2014). The white boxes illustrate the research methods. The white boxes with rounded corners portray the resources used for the research methods.

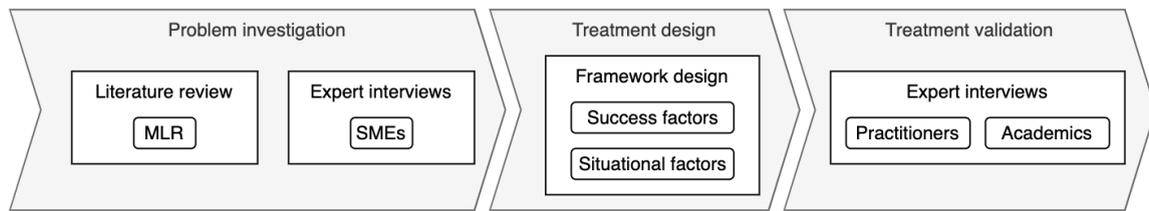


Figure 4. Overview of research approach

In table 3, an overview is given of the different interview rounds. The first round focuses on SMEs who have been involved with the implementation of data governance platforms at organizations. Interview round two is split up in two parts; experts from organizations which have implemented a data governance platform or are thinking about it, and academics proficient in the area of data governance / management or a topic related to this.

Table 3. Overview of interviews

Interview round	1	2	
		1	2
Stakeholders	Subject matter experts (SMEs)	Organizations	Academics
Goal	Elicit success and situational factors	Validate framework in practical setting	Validating framework in theoretical setting
Chapter results	4	6	6

2.3 Validity threats

The risks of this research should be considered so that they can be minimized accordingly. Wohlin et al. (2012) specified four threats to validity; conclusion validity, internal validity, construct validity and external validity. For each of these threats to validity, table 4 gives an explanation of what they entail, how they could impact this research (i.e. the risk), and how they were mitigated.

Table 4. Threats to the validity and mitigation tactics

Threat	Definition	Risk	Mitigation tactic
Conclusion validity	The extent to which correct conclusions are drawn from the treatment	The research cannot be repeated in the same way (reliability)	Protocols and guidelines are used for the research methods (see chapter 2.2)
		Sample sizes are too small to draw correct inclusions	Various interviewees are questioned for the expert interviews so that conclusions are supported by multiple experts
		Incorrect categorization of possible interview answers	Various sources are gathered and compared for the literature review so that categories can be derived that are supported by multiple sources
Internal validity	The extent to which the treatment causes the outcome	Interviewees provide incorrect or biased information	The interviewees have a variety in backgrounds so that as many as possible perspectives are covered
		Incorrect assumptions are made from the interviews	Conclusions are drawn from topics that are supported by multiple experts
Construct validity	The extent to which the research actually reflects the construct under study; whether the researcher measured what was intended to be measured	Questions from the interview do not accurately cover all implementation factors	Questions and answer categories are made based on the results of the literature review so that validated factors are considered in the interviews
		Incorrect dimensions are considered for the design of the framework	The framework is validated with experts interviews so that any incorrect assumptions can be corrected
External validity	The extent to which the results of this research are generalizable	Interviewees / organizations are selected that are not representable stakeholders	Interviewees / organizations from various industries are selected who have experience with different stages of the implementation so that the outcomes are based on multiple cases
		Interviewees / organizations originate from one organization (Avanade) which can cause bias	The research approach covers multiple research methods (literature review, expert interview, case study) so that outcomes are not based on opinions of interviewees alone
		High focus on Purview in interviews	Experts were asked about their experience with other data governance platforms and how it differed from Purview

		Framework draws is too practical or too theoretical	The success factors were based on both literature review + expert interviews, validation interviews were conducted with both practitioners + academics (triangulation)
--	--	---	--

2.4 Conclusion

The research was executed by following the steps from the design science lifecycle as proposed by Wieringa (2014), which consists of the problem identification (literature review and expert interviews with SMEs), the treatment design (design of the framework), and the treatment validation (expert interviews for case study and academics). These three phases led to the design of the final data governance platform implementation framework. The next chapter starts with the results of the multivocal literature review as part of the problem identification.

3. Results related work

In this chapter, the results of the multivocal literature review (MLR) are given, where both formally published literature and gray literature are examined. From the previous section it becomes clear that both literature was gathered about data governance and implementation approaches in the MLR, as there is not yet one best approach for the implementation of data governance platforms. The related works chapter provides the theoretical background needed for the rest of the research.

First of all, more background information about data governance is given in section 3.1 with a description of its origin. In chapter 3.2, data governance fundamentals are discussed by giving its definition, explaining the building blocks, elaborating on related disciplines, and by giving an overview of data governance frameworks from theory. Next, implementation fundamentals are discussed in chapter 3.4, and an overview is given about definitions and approaches, strategies, and frameworks from literature that can be applied to the implementation of new systems. After that in chapter 3.3, data governance platforms are described in more detail. Next, chapter 3.4 presents the success factors for data governance and for the implementation of new systems that can be derived from literature. Finally, chapter 3.5 gives a brief overview on the importance of frameworks for implementation.

3.1 Origin of data governance

Data governance has shifted from something that is nice to have to something that is of critical importance in an organization (Abraham et al., 2019). Originally, data was seen as a product of transactional systems and it was not considered as something valuable beyond the transaction and the corresponding application (Al-Ruithe et al., 2019). However, the value of data has increased enormously since the early 1990s.

One of the key trends that caused a growing popularity in data governance is the growing pressure on organizations from regulatory requirements (Abraham et al., 2019). The need for data governance started with the Sarbanes-Oxley (SOX) act that was enacted in 2002 in the US (Cheong & Chang, 2007). This was a result of the Enron scandal in 2001, when it came to light that Enron was involved in accounting and corporate fraud. The SOX act required organizations to have better controls and audits in place, as executives were considered responsible for the credibility of their financial reporting in this act. When the General Data Protection Regulation (GDPR) was introduced in 2018, there was also more pressure for organizations in Europe to have strict control over where data are stored and how these data are being used. Thus, data governance aims to ensure regulatory compliance and mitigate data security and privacy risks (Mahanti, 2018). This avoids lawsuits, regulatory fines, penalties and brand damage associated with data breaches.

Another trend is the growing amount of data volumes coming from diverse sources (Abraham et al., 2019). These growing data volumes are due to the high emphasis on digital transformations, the introduction of the Internet of Things (IoT), fast advances in technologies, and decreasing costs for the storage of data (Mahanti, 2018). Moreover, the shift to the cloud has led to large volumes of data coming from multiple sources (Al-Ruithe, Benkhelifa, & Hameed, 2016). This resulted in the rise of Big Data, meaning large volumes of complex data that grow at a high rate, which requires organizations to impose new ways of governing data (Al-Badi, Tarhini, & Khan, 2018).

Abraham et al (2019) emphasize that these larger data volumes make data inconsistencies more likely to happen, which can result in decisions made based on incorrect data.

Fragmented enterprise architectures and legacy systems contribute to the problem of inaccurate and incomplete data (Abraham et al., 2019). In the past, IT systems were more centralized which made it easier to keep track of that data (DeStefano, Tao & Gai, 2016). Data is now stored in a multitude of data repositories, which leads to a clear need for a common understanding of data within organizations. This is especially relevant now that self-service reporting and analytics is more frequently being used in organizations. Thus, if these problems in the organization are not properly addressed by data governance, data quality issues and inconsistent data models are more likely to happen (Mahanti, 2018). Having high quality data in turn ensures better informed strategic decision making.

In the past, data intensive initiatives included amongst others data warehouses, data integration, data migration, and Enterprise Resource Planning (ERP) systems (Mahanti, 2018). However, due to the increasing volumes and the higher complexity of data, traditional data management solutions are no longer able to handle this data, as they are becoming too costly and cannot process the large data volumes (Al-Ruithe et al., 2019). Master data management is one of the solutions for this, which provides a single point of reference for all the data stored in the different systems (Al-Ruithe et al., 2019). However, Mahanti (2018) describes that in order for data initiatives such as data migrations to be successful, some form of data governance should be in place. Resultantly, according to Al-Rutihe et al. (2019), in order for organizations to be able to properly manage data demands in the long-run, data governance should be effectively implemented.

Currently most organizations are aware of the criticality of data governance, but not all organizations know how to execute it properly (Ladley, 2012). Survey research by Mahanti (2018) into the state of data governance undertakings at organizations showed that currently most organizations have some type of formal data quality processes and procedures in place, but that it is still a work in progress. Many organizations use a software tool for tracking data issues, but there is also a big group who manually document this (Mahanti, 2018). As for data governance platforms, there is a large group of organizations which use in-house developed tools. As volumes of data increase at a high speed every year, new solutions for managing it are needed (Reinsel et al., 2018). Thus, because of the increasing need for data governance, it is essential to know more about its fundamentals, which are discussed in the next section.

3.2 Data governance fundamentals

In this section, background information is given about data governance. The section starts with the definition of data governance, and introduces its building blocks. After this, related approaches are compared with data governance to position the topic. Next, an overview is given of data governance frameworks. Lastly, a description is given about the perspective of industry standards on data governance.

3.2.1 Data governance definitions

The term “governance” can be defined as the way an organization ensures that strategies are set, monitored and achieved (Al-Ruithe et al., 2019). However, there is no evidence of one uniform

definition of the term “data governance” in scientific literature and practitioner publications (Abraham et al., 2019). Therefore, the most important definitions and elements of data governance are discussed in this section.

The Data Governance Institute, a leading practice-oriented and vendor-neutral institution bringing out best practices and guidelines for data governance, defined data governance as follows: “Data Governance is a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods” (The Data Governance Institute, 2022). In this definition, decision-making and authority are important elements. The Data Governance Institute (2022) describes that data governance can refer to organizational bodies, rules (e.g. policies), decision rights, accountabilities and enforcement methods for data-related matters.

Begg and Caira (2012) describe that earlier definitions of data governance in theory mainly focused on procedures and policies, but that there is a shift in emphasis on roles and business structures. From a more theoretical approach, by studying the common characteristics in definitions found in their literature research, Abraham et al. (2019) propose the following definition of data governance: “Data governance specifies a cross-functional framework for managing data as a strategic enterprise asset. In doing so, data governance specifies the decision rights and accountabilities for an organization’s decision-making about its data. Furthermore, data governance formalizes data policies, standards, and procedures and monitors compliance.” Whereas decision-making and authority are key elements in the practitioner's perspective from The Data Governance Institute (2022), this is just one part of the definition from theory. To be more specific, the following six parts of the definition from Abraham et al. (2019) can be found in definitions of other publications and are fundamental to data governance: cross-functional effort; framework; data as a strategic enterprise asset; decision rights and accountabilities; data policies, standards, and procedures; and monitoring compliance. Each of these fundamentals of data governance are described in the next section.

3.2.2 Data governance building blocks

As indicated by Abraham et al. (2019), each data governance program touches upon the following topics:

- *Cross-functional effort.* Mahanti (2018) describes that the purpose of data governance is to align all business units within an organization. If there is no formal data governance structure in place, individual business units or departments can start developing their own rules and standards, resulting in a lack of alignment between units.
- *Framework.* A data governance framework can be defined as the logical structure for classifying, organizing, and communicating activities that revolve around the decision-making about data (Mahanti, 2018). This framework is specific to an organization, and resultantly depends on the organization’s size, culture, structure and practices. However, there are three basic components that should be considered in every framework, and they include the policies, processes, and roles / responsibilities for data governance. Here, policies are a documented set of guidelines which specify how data should be managed. The processes comprise the definitions of the activities and their deliverables needed to manage data. The

roles and responsibilities are to be defined also, which include a formal structure of the functions in the organization that are responsible for the data.

- *Strategic enterprise asset.* Mahanti (2018) highlights the importance of treating data as a strategic enterprise asset for data governance. Khatri and Brown (2010) describe that organizations should start thinking about data as an asset from which potential value can be gained. Instead of seeing it as a result of systems, it should be used to gain an advantage. Taking this into account, data governance should comprise all practices at an organization related to ensuring high quality of data assets throughout their entire lifecycle (Mahanti, 2018).
- *Decision rights and accountabilities.* This component refers to the organizational bodies that oversee the data assets (Mahanti, 2018). Khatri and Brown (2010) highlight that an essential component of data governance is to decide who is responsible for a certain data asset, and who holds accountabilities for the decision-making about this asset. Assigning responsibilities to data is therefore one of the key activities in data governance. The different data roles are specified further in table 5.
- *Data policies, standards, and procedures.* Data principles are used for setting requirements for the intended use of an organization’s data (Khatri & Brown, 2010). These data principles come in the form of policies, standards, or procedures. They explain the desired behavior for business or IT users of the data. They help to ensure the consistent and appropriate use of data across domains.
- *Monitoring compliance.* One of the main goals of data governance is ensuring compliance with regulatory requirements (Panian, 2010). Many data governance tasks revolve around improving the accountability and auditability of data.

Table 5. Data governance core roles, adapted from (Mahanti, 2018)

Role	Level	Description
Steering committee	Executive level	Responsible for sponsoring and approving the data governance plan
Data governance council	Strategic level	Responsible for approving the data governance policies, roles, frameworks, tools, etc and for promoting the data governance practices
Data steward council	Tactical level	A group of data stewards and coordinators, responsible for making data-related decisions
Data stewards	Operational level	Responsible for the data quality
Data producer	Support level	Responsible for creating the data
Data owner	Support level	Responsible for the quality of a dataset
Data consumer	Support level	Responsible for determining what the requirements should be for the data
Data custodian	Support level	Responsible for ensuring the data in IT systems is consistent with the business requirement

It should be noted that data governance is a broad topic and Mahanti (2018) describes that data governance programs can vary widely according to their focus (e.g. compliance, data integration or improving data quality). It therefore depends on the organization which of the building blocks is a focus of a data governance program. Data governance is not a project that has a definite end but rather a long-term initiative that needs to be sustained in order for it to work. There is no one-size-fits all solution that can be applied to every organization, and therefore every data governance undertaking should be tailored to the organization at hand. In the next section, related approaches to data governance are given to show that data governance is dependent on other disciplines as well.

3.2.3 Related approaches to data governance

In this section, related disciplines to data governance are discussed in order to position the topic. What should be noted is that data governance should not be seen as something separate from these disciplines, but that they influence each other and thus should be practiced together.

Data governance vs data management

Data governance is closely related to data management, as data governance is about the control and authority over the management of data (Mahanti, 2018). Whereas data governance is about the decisions that must be made about data and who makes the decisions, data management is about executing these decisions on a daily basis (Alhassan et al., 2016). Resultantly, there is a strong connection between data governance and data management.

Data governance vs IT governance

IT has become a part of every organization, and therefore IT governance is part of every organization's corporate governance (Al-Ruithe et al., 2019). Al-Ruithe et al. (2019) describe IT governance as a way to ensure the organization's IT sustains and extends the organization's strategies by providing mechanisms for leadership. Whereas data governance is about the organization's data in different technology environments, IT governance is about the IT equipment in the organization (Al-Ruithe, 2019). To be more specific, the focus of IT governance is around the organization's hardware and applications, but not the data. Khatri and Brown (2010) note that there is a clear distinction here between IT assets, which are the technologies that help support the organization's processes, and data, which are facts that have potential value. However, Begg and Caira (2012) describe that principles from IT governance can influence the way that data governance should be handled. Rather than these disciplines to be seen as separate, they should be aligned with each other (Russom, 2008).

Data quality

Data governance also overlaps with data quality management, as data governance aims to improve the quality of data. Data quality is about leveraging accurate, timely, relevant, complete, understandable and trustable data (Cheong & Chang, 2007). One of the issues with data quality results from the many systems that cause the data being spread across the organization. Moreover, the same data often is used by multiple levels within the organization. Next to that, many systems are developed without taking data quality assurance in mind. According to Cheong and Chang (2007),

these issues can be addressed by having an effective data governance program in place. Brous et al. (2016) mention that data quality is often used for measuring the effectiveness of data governance within an organization. There are several data quality indicators that can be used for measuring the quality of data, and these include accuracy, completeness, consistency, timeliness, validity and uniqueness of data (Dai, Wardlaw, Cui, Mehdi, Li & Long, 2016). As part of a data governance program, a goal can be to improve those data quality dimensions.

Enterprise data modeling for data governance

Data governance can only be applied successfully if the data that need to be governed is properly understood (Brous et al., 2016). DeStefano et al. (2016) describe that in order for a data governance program to be properly executed, it is important to first understand what data there is, as *“you cannot govern what you do not know”*. Therefore, it is important to first understand the organization’s data landscape. Brous et al. (2016) mention that an enterprise data model (EDM) can help avoid misinterpretation and aid the implementation of a data governance program.

3.2.4 Data governance frameworks

In order to demonstrate different perspectives on what data governance looks like, this section gives an overview of some of the data governance frameworks. In this section of the report, a data governance framework is a conceptual visualization used for structuring and classifying the concepts related to data governance and showing the relationship between these concepts (Abraham et al., 2019). In table 6, an overview is given of data governance frameworks identified in literature, together with their scope or main focus, and their core elements.

Table 6. Overview of data governance frameworks

Author	Year	Scope / main focus	Core elements
Abraham et al.	2019	Building blocks	Six dimensions: governance mechanisms, organizational scope, data scope, domain scope, antecedents, consequences
Al-Badi et al.	2018	Big Data	Eight components: identify organizations structure, stakeholder selection, Big Data scope determination, policies and standards setting, optimize and compute, measure and monitor quality, data storage, communication and data management
Al-Ruithe et al.	2016	Cloud	Five processes: data governance structure, data governance assessment, data governance function, negotiation, data governance level agreement
Al-Ruithe et al.	2019	Cloud	Eight dimensions: data governance structure, cloud deployment model, service delivery model, organizational, technical, environmental, cloud actors, SLA
Alhassan et al.	2016	Activities	Eight areas: data roles & responsibilities, data policies, data processes & procedures, data standards, data strategy, data technology, data guidelines, data requirements
Khatri & Brown	2010	Decision domains	Five domains: data principles, data quality, metadata, data access, data lifecycle

Otto	2011	Organization (corporate)	Two dimensions: goals (formal goals, functional goals), data governance structure (locus of control, organizational form, roles and committees)
Petzold et al.	2020	Archetypes	Two dimensions: level of data related regulation in the industry, level of data complexity

What becomes clear from the core elements of the different frameworks is that many of the frameworks focus on the foundations of data governance; what is generally needed to have good data governance in place. Most frameworks mention that this should include roles and responsibilities, policies, a governance structure, and scope determination. Moreover, as can be seen from the core elements of the framework, data governance can be looked at in different ways, for example in dimensions, domains, areas, components and processes. Since this research focuses on data governance platforms by taking the characteristics of the organizations into account, one framework that recognizes this is shown as an example framework and is elaborated upon underneath.

Petzold et al. (2020) developed a framework with the goal of informing on the level of sophistication needed for a data governance program, which can be seen in figure 5. They divided the framework into two dimensions: level of regulation and level of data complexity, which leads to four quadrants. Each quadrant has different requirements for what level of data governance needs to be applied. A highly regulated industry requires more sophistication, as there should be additional controls, reports, and documentation. Additionally, a higher level of data complexity needs more centralized support for standards, minimum controls, and effective management. The level of data complexity increases with a broad scope of business operations (i.e. many divisions, e.g. business lines, geographies), high evolution and speed of critical data, and a low level of data automation. The right level of data governance sophistication can also be adjusted to a specific data set instead of the whole organization.

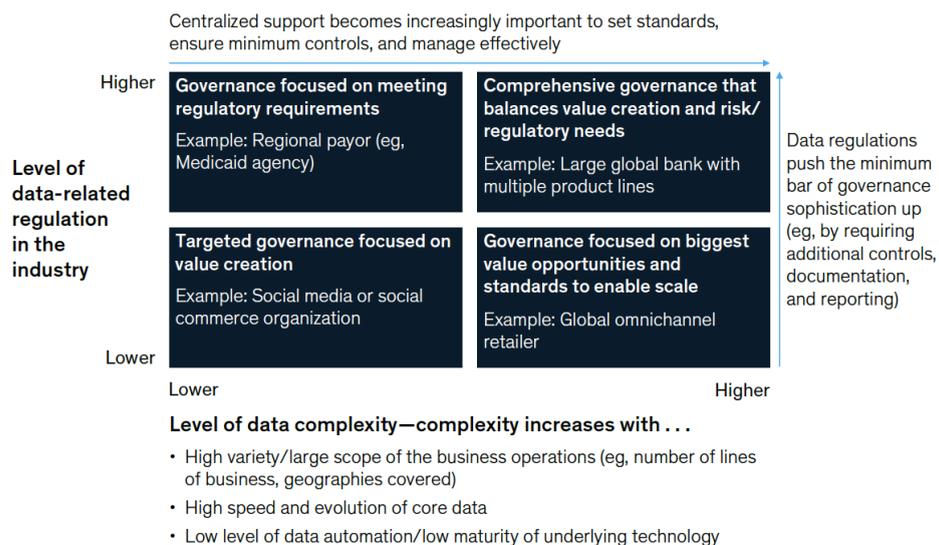


Figure 5. Data governance archetypes framework (Petzold et al., 2020)

3.2.5 Industry standards on data governance

DAMA is a professional organization of data management professionals and they consider data governance to be one of the core functions of data management (Mahanti, 2018). They proposed a framework that gives guidance for how an organization can implement data management, and this guide is called the DAMA-DMBOK Guide (DAMA, 2014). Herein, 11 data management knowledge areas are specified, and data governance is at the core of these areas, which can be seen in figure 6. This shows that data governance touches upon many different subjects that are also covered by data management.

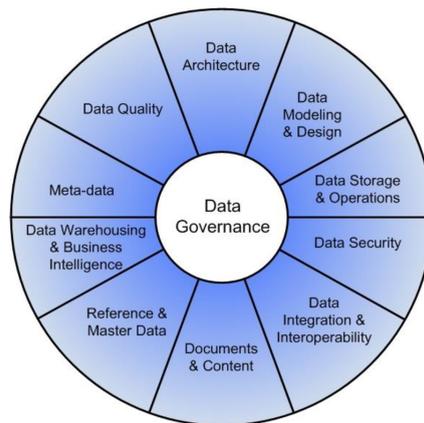


Figure 6. Data management knowledge areas according to (DAMA, 2014)

IBM, a multinational technology enterprise, formed the IBM Data Governance Council together with partner organizations as a way to establish best practices for data governance (Begg & Caira, 2012). They designed a framework called the Data Governance Maturity Model, with the aim of assessing the data governance maturity level of an organization. As part of this framework, they have developed a tool in which an organization can measure its data governance maturity by answering questions on different data-related areas. According to Al-Ruithe et al. (2016), IBM stresses the importance of software support for data governance in their framework, as IBM is a software provider and thus aims to also sell products.

Another advisory body on data governance, the Data Governance Institute (DGI), is a vendor-neutral institution which develops best practices for data governance (Begg & Caira, 2012). They aim to provide a framework which helps organizations to set up their data governance program. This is a single framework, because they believe that any data governance program follows the same universal objectives. Their framework is divided in three components, which are rules, people, and processes (Al-Ruithe et al., 2016).

3.3 Implementation fundamentals

This section addresses the fundamentals of the implementation of new technologies. First, definition is given of implementation and related disciplines. Next, the contingency and configuration approach are introduced. Thereafter, some implementation strategies are addressed. The section closes with two examples of change implementation frameworks.

3.3.1 Implementation definitions

According to Van Linge (2006), implementation can be defined as the actions and events that lead to the use of an innovation. This should be distinguished from adoption, which is the decision to actually use the innovation. For a successful implementation, more is needed than a high adoption (Van Linge, 2006). Innovations have a subjective element and are dependent on the involved people. When people perceive an idea, model, system, method or tool as new, it can be considered an innovation. Organization systems should have the structure, culture, and human resource practices that support this implementation. Any implementation of a new system requires organizational changes (Atkins et al., 2017). Resultantly, organizations are ought to adapt their processes, culture, strategies and structure in such a way that the change can be accommodated (Imran, Rehman, Aslam, & Bilal, 2016).

The way an organization should handle the change can be called change management, which is defined as the continuous process of renewing an organization's direction, structure, and capabilities in order to satisfy the constantly changing needs of stakeholders (AlManej, Salonitis, & Tsinopoulos, 2018). The goal is to guide the stakeholders in the organization during the change from the initial state to the intended future state by managing the challenges associated with the change and consequently minimizing the resistance. Due to continuous changing environments, change is necessary for organizations to survive (Hussain et al., 2018). The success of change implementation can help an organization in achieving a competitive advantage, but it can also lead to disastrous consequences when it fails (Sonenshein & Dholakia, 2012).

The implementation of data governance can be defined as the process of ensuring that the capabilities required for data governance value are made operational (Ladley, 2012). It is about going from a non-governed state to a governed state. To put this in the context of systems, a system can be called implemented as it is made available for use. A successful implementation of a system is the case when the system is implemented on time with reasonable costs with an acceptable risk, and is embedded in the organizational processes and routines (Arvidsson, Holmström, & Lyytinen, 2014). However, the majority of information implementation system projects end in failure (Ćirić & Raković, 2010). Moreover, in many cases successful completions of a project exceeded the project time and budget. Therefore, it is important to consider the way implementations can take place and the factors that affect the success of an implementation.

3.3.2 Contingency & configuration approach

If an innovation has been proven effective in a certain organization, one cannot be certain that this same innovation is also effective in another organization. Van Linge (2006) describes that the effectiveness is dependent on the situation in the organization. Situational factors that increase this complexity include the variability in use of the innovation, the uncertainty about the effectiveness, the stability of the environment, the dependency on other information, and the shareability of its use. The contingency approach takes into account that the circumstances under which innovation takes place are different for each situation, and that these circumstances have an effect on the outcome of an innovation (Van Linge, 2006). By taking this contingency approach into consideration, different models can be distinguished that illustrate the relations between situational factors and the implementation success of an innovation.

A continuation of the contingency approach is the configuration approach (Van de Ven et al., 2013). Instead of single elements it focuses on combinations or clusters of elements. A configuration is a system (e.g. an organization) with a consistent set of characteristics. It aims to find an ideal configuration between the demands of the external environment and the design of the internal structure of the organization. The configuration approach suggests that a structure can be designed in such a way that it has an optimal fit with its environment and thus maximizes the performance of an organization.

3.3.3 Implementation strategies

Different classifications of implementation strategies can be found in literature. A strategy that is often determining for the success of an implementation is whether the change is handled top-down or bottom-up (Tenkasi & Chesmore, 2003). A top-down approach to the implementation of a change means that the change is driven by the highest level of the organization, whereas a bottom-up approach means that the change started from the lower levels of the organization. An advantage of a top-down approach is that adoption is at a higher speed with a predictable outcome as it is more structured and systematic. Regarding a bottom up approach, it is more easily accepted by the employees as it also originates from them. However, it can take more time to be introduced and adopted in the rest of the organization (AlManei et al., 2018).

Van Linge (2006) addresses other ways to look at the implementation of a new change, and describes that the classification of implementation strategies is often made in pairs. Examples of this are (Van Linge, 2006):

- Planned and emergent strategies; about the preparation beforehand.
- Transition and transformation strategies; based on the intended outcome (AlManei et al., 2018).
- Incremental and radical strategies; scale and time of the change (AlManei et al., 2018).
- Reactive and proactive strategies; whether the organization is the initiator of the change or the external environment (Hussain et al., 2018).
- Coercive and cooperative strategies;

Other important categorizations of implementation strategies that can be considered are based on (Van Linge, 2006):

- The dynamics of the change (e.g. unfreezing, moving, freezing).
- The interventions in the strategy (e.g. human-resources, structure, politics, culture, communication and network, language, learning and knowledge, environment).
- Focus of the interventions (e.g. individual, group, organization, network, environment, cross-level interventions).
- Timing and relation between interventions (e.g. sequential or parallel relation).
- Depth of interventions (i.e. core process, peripheral).

Van Linge (2006) describes that every implementation strategy should accommodate the fit between innovation, organization, people, and environment, and that the best strategy therefore depends on the situation.

3.3.4 Change implementation frameworks

Cameron and Quinn (2006) describe that a wide variety of dimensions have been proposed for organizing and describing the characteristics of an organization, more specifically the organizational culture. They give as an explanation for this variety in dimensions that organizational culture is an extremely complex subject, and it is impossible to include every single characteristic for assessing the organizational culture. Consequently, it is important to use an underlying framework that helps to determine the most relevant set of dimensions, as this theoretical foundation can help narrow the search for characteristics to focus on. Due to the complexity of dimensions, Cameron and Quinn (2006) argue that there is no “right or wrong” framework, but that the choice for an appropriate framework should be based on (1) empirical evidence, (2) an accurate depiction of the reality described, and (3) its ability to capture and integrate most of the proposed dimensions. Resultantly, Cameron and Quinn (2006) proposed the Competing Values Framework, which can be seen in figure 7. In this framework there are two dimensions, which are flexibility vs stability and internal vs external. These dimensions form four quadrants, each quadrant representing a distinct set of indicators that determine the organizational effectiveness.

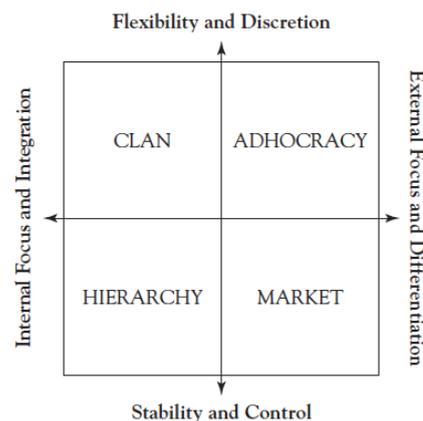


Figure 7. The Competing Values Framework (Cameron & Quinn, 2006).

In the proposed framework from Cameron and Quinn (2006), the quadrants are labeled based on organizational forms that fit the characteristics of the quadrants. These labels are as follows, together with the characteristics of the culture type:

- *Hierarchy*: formalization, procedures, rules, stability, predictability, efficiency, specialized jobs, centralized decisions.
- *Market*: market niches, exchanges, sales, contracts, profitability, secure customer bases, competitiveness, productivity.
- *Clan*: teamwork, employee development, customers as partners, participation, commitment, loyalty.
- *Adhocracy*: adaptability, flexibility, creativity, uncertainty, temporary.

Each quadrant has its own suggested leadership roles, effectiveness criteria, and core management theories associated with them. This can help an organization in choosing their best approach to managing a cultural change. Jones, Jimmieson and Griffiths (2005) describe that the organizational culture can influence the way that employees within an organization are ready for change as a result

of a new implementation. Employees in a clan or adhocracy culture are often more likely to have a positive attitude towards the change.

Building on top of this Competing Values Framework, Van Linge (2006) introduced a new model by applying the model from Cameron and Quinn (2006) in the context of implementation science, which is the study that promotes the translation of research findings and evidence-based practice into routine practice, mostly in healthcare settings (Cook & Odom, 2013). This model is called the innovation-contingency model and it is shown in figure 8. Whereas the model of Cameron and Quinn is focused on culture, the model of Van Linge is focused on innovation, with a high focus on implementation. The intention of this model is to choose the best implementation strategy in practice, by taking into account the characteristics of the change itself, the involved people, the organization, the environment, and any other pending changes.

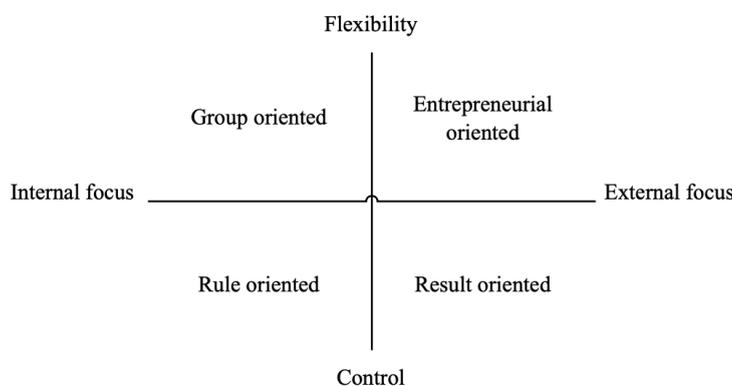


Figure 8. Innovation-contingency model, adapted from (Van Linge, 2006).

In the model of Van Linge, there are two dimensions; a relation dimension (control versus flexibility), and a direction dimension (intern versus external focus). When combining the two dimensions, four configurations can be identified, each with their own set of interventions that are most effective for this specific configuration:

- **Rule-oriented:** detailed implementation plan, education about technical competencies, allocation of tasks, standardization, supervision on compliance, procedural learning, coercion.
- **Group-oriented:** global implementation plan, action learning and reflection, team development, coaching, shared responsibility for compliance, communication, team learning and sharing of experiences, shared decision making.
- **Result-oriented:** implementation plan with focus on KPI's, education about goal-oriented thinking, auditing, feedback, control of goal realization, operationalizing and measuring of outcomes, formalizing result responsibility, rational convincing, ICT support, working with reminders.
- **Entrepreneurial-oriented:** interactive learning, forming of temporary teams, incentives for creativity, education focused on creativity, handling complex situations, participation in networks with externals, promoting knowledge sharing.

Regarding the implementation of innovations, every configuration has its favorite strategy for the way it handles implementation. It can be possible that an organization shows characteristics from multiple configurations. These multiple configurations then form a so-called fitness landscape. Ćirić

and Raković (2010) discuss the kind of configuration that is most likely to be successful for an implementation effort, which is the case for a task-based organization. This is because they are highly focused on disciplines including planning, supervision, and group responsibilities. In the model of Van Linge (2006), this corresponds to characteristics from the group-oriented and result-oriented configuration.

A successful implementation can be defined as the realization of the requirements and processes needed for the innovation and the extent to which the innovation is used as intended (Van Linge, 2006). In order for a successful implementation, there should be a good fit between the characteristics of the environment and the characteristics of the innovation. A good fit between innovation and environment results in the implementation phase in acceptance of people with the innovation and the willingness to share resources (e.g. money, information). If the implementation is seen as successful, it is still not finished; the actual adoption of it requires continuous monitoring and improvement.

3.3 Data governance automation

The automation of data governance activities is important as it can make sure that data governance can be executed more accurately, faster, and it provides more options for scaling (Russom, 2008). Platforms that are fully dedicated to data governance were rare in the past (Russom, 2008). However, due to the growing volumes of data, tools solely for data governance are rapidly being developed (Peyret & Goetz, 2014). Besides dedicated data governance platforms, there are also functionalities in other tools that can provide opportunities for the automation of data governance, such as metadata management or master data management functionalities (Russom, 2008). The platforms and related functionalities are discussed in this section.

3.3.1 Data governance platforms

A data governance platform is a suite of tools that provide functionalities for the automation of data governance tasks (Russom, 2008). Benefits include a higher consistency and better auditability of data. Moreover, governance tasks can easily be scaled to more systems or data. It can also help with the communication of information. Mahanti (2018) mentions that data governance platforms are not a replacement of the organization's approach to data governance, but they can help support the tasks and processes for data stewardship. Functionalities that these platforms support include the creation of data policies, managing workflows, and monitoring of data use and policy compliance. Data governance platforms play a supporting role, and should be an addition to a well-established data governance framework (Russom, 2008). The last step that should be considered in a data governance program is the technology (Ladley, 2012).

Important functionalities which are included in a data governance platform are a data catalog and a data glossary, as these functionalities provide a way to organize the metadata (Ladley, 2012). Metadata can be defined as "data about data" and it can be used for interpreting and understanding the meaning of data (sometimes referred to as semantic data) (Khatri & Brown, 2010). Examples of metadata are the creator or user of data, or audit information about data. In a tool, some functionality for metadata is useful as this provides a way to search for the data that you need (Ladley, 2012). Moreover, data lineage is a functionality included in many data governance platforms

(Wróbel, Komnata & Rudek, 2017). Data lineage gives insights into the data journey from the origin of data in source systems, to the way it moves through multiple systems, and finally to consumption in applications or reports. Gaining insights in the lineage is essential for evaluating the completeness and correctness of data, as it can be traced back to where possible conflicts or issues occur or it can be seen where data flows to. With data lineage it can be checked whether the data sources are reliable. Additionally, traceability of data is needed for compliance with regulations that require organizations to justify the accuracy of data.

Survey research by Mahanti (2018) was done on the functionalities that organizations desire in a data governance platform. The most frequently reported functionality was a business glossary for establishing commonly used data definitions and metadata management. Another highly desired feature was the tracking of data lineage. Additionally, many organizations indicated that they would like the platform to be integrated with other data management products. Other desired functionalities revolved around workflow management, data standards, assignment of accountability for data, data governance metrics linked to business outcomes, policies, and collaboration options. Some organizations also indicated an assessment of the data quality to be a highly desired functionality (Mahanti, 2018).

3.3.2. Related tools

Many tools include some functionalities for the management of data quality, options for data integrations, and metadata or master data management (Russom, 2008). Instead of having a whole platform dedicated to this, it can be functionalities that are included in tools that do not have a primary purpose to enforce data governance, for example an enterprise resource planning (ERP) tool. In some cases, data governance can even be managed in traditional technologies such as SharePoint, Word, and Excel (Ladley, 2012). Ladley (2012) made a categorisation for the different tools which offer data governance functionalities, and these include tools for discovery, administration (e.g. workflow), data efficacy (i.e. quality), data provisioning (e.g. access), lifecycle management, metadata, data storage, and provenance (i.e. data lineage). There is not always a clear distinction between data governance tools or tools for areas such as data quality (Ladley, 2012).

Metadata and master data management are highly desired functionalities in the area of data governance. As tested in a survey in the research of Russom (2008), executives in the area of data management believed the management of semantic data to be the most important for data governance. Semantic data gives meaning to the data which helps to build an inventory for it. Whereas master data is about providing context about one single version of data, metadata helps to identify attributes and describe other data. Functionalities for this are often built in other tools for data integration or data quality (Russom, 2008). Metadata and master data management is not effective without some kind of data governance (Ladley, 2012).

In order for metadata and master data capabilities to work, there needs to be some kind of central repository with data that needs to be analyzed (Russom, 2008). A single repository would be beneficial as then there is one place of data which needs to be considered, but in most cases organizations have multiple repositories in which their data resides. Here, data warehousing and data lakes are important tools to consider. The different repositories in the organizations need to be synchronized in order for the metadata and master data to be accurate. This is where a data governance tool can help.

There are also data quality technologies which provide options for the cleansing, validation, and monitoring of data, which can directly contribute to the execution of data governance (Russom, 2008). Monitoring functions in data quality tools can for example also be used to ensure compliance. Other functionalities that can be part of data quality technologies include the identification of data improvement opportunities, defining rules for data transformations, and reviewing reports. However, according to Russom (2008), a limitation of these tools is that they often scan one database at a time. With the distributed data environment of today, tools should be used that are able to perform cross-system analysis.

Data integration tools can also provide data governance functionalities (Russom, 2008). They help with the moving and transformation of data, and can bring data together into one unified overview. However, even though integrating data into one central database makes it easy to assess the security and privacy of data, it is often not feasible to integrate and synchronize the large volumes of data in unstructured forms.

A valuable type of tool that can contribute to data governance is data profiling tools (Dai et al., 2016). It can improve data quality by discovering patterns in datasets and it can offer functionality such as data cleansing, data integration, and data analysis. A data profiling tool works by determining the metadata about a dataset. They are usually part of a packaged data governance suite. Thus, instead of implementing a complete data governance suite (i.e. platform) in one go, it is also possible to implement parts of it.

3.3.2 Vendors

There are multiple vendors in the market that have introduced data governance platforms. The selection for a suitable platform should be based on the extent to which the capabilities of the platform can support the business requirements and related use cases (Mahanti, 2018). Leading vendors that provide these platforms for example include Collibra, Informatica, IBM, SAP, Oracle, and Datum. Wróbel et al. (2017) describe that most governance software have some functionalities in common, which at least include a glossary and an option for generating data lineage. Additional capabilities depend on the vendor, and the most commonly used platforms are discussed below.

IBM offers the IBM InfoSphere Information Governance Catalog (IGC) (Wróbel et al., 2017). This solution is specifically a straightforward choice for organizations who have IBM tools already in place, as their IGC is fully compatible with the IBM environment. Moreover, IBM's solution offers workflow management and administration of data governance processes, which makes it a good option for highly regulated industries.

Collibra offers the platform called Data Governance Center (DGC) (Dai et al., 2016). It has a Data Stewardship Manager (DSM) tool that functions as a data profiling tool which provides data quality reporting and trend analysis. Besides that, included in the platform is a Business Semantic Glossary (BSG) and a Reference Data Accelerator (RDA) module.

3.5 Factors affecting the implementation of data governance

In this section, the factors that influence the outcome of a data governance program are discussed. Firstly, the success factors for the implementation of data governance are given. Secondly, an overview is given of the situational factors that affect the implementation of data governance.

3.5.1 Success factors for the implementation of data governance

An overview of all success factors, together with the publications in which they were covered, is given in table 7. A more extensive description of each of the factors is given below.

Mahanti (2018) conducted a study to determine critical success factors associated with data governance implementation by looking at results from literature and surveys with information management professionals with experience in data governance. In this study, it was revealed that the most important factor for ensuring a successful data governance implementation concerns commitment from leadership. As data governance requires a significant change in the way of working and the attitude of employees, it is important that there are strong drivers and dedicated leadership to build up commitment for the initiative. In order to avoid resistance from employees, there needs to be clear communication and education to inform the employees of the change, where leadership should set the example (Figuroa-Flores, Acosta-Gonzaga & Ruiz-Ledesma, 2020).

Another important factor is an active executive sponsor (Mahanti, 2018). If there is not enough financial support for the initiative, then the data governance program is stopped quickly. If there is an active sponsor, data governance can gain attention from managers and the rest of the organization. In order to ensure commitment from leadership and sponsors, it is important that the business value is clearly demonstrated (Al-Ruithe et al., 2019). This can be in the form of a business case, which includes an overview of the vision, the costs, the benefits and the risks. By looking at how the data governance program can add value for the business at every stage of the process, requirements from IT are aligned with the business.

A robust data governance strategy that aligns with the organizational strategy is determining for the success of a data governance program according to Mahanti (2018). This is the foundation for the implementation plan, and it should describe the vision, a roadmap and realistic milestones. An essential part of the strategy for the implementation of a new system is the formulation of the objectives and vision, as this provides the basis for what the organization wants to have in the system (Figuroa-Flores et al., 2020). Ćirić and Raković (2010) confirm this by mentioning that it is important for the implementation of an information system to have a clear plan and objective, preferably documented. A failed information systems (IS) strategy implementation is often due to the misalignment between strategic intent and the capabilities of the system (Arvidsson et al., 2014). Therefore, when deciding which capabilities to include in the system, it is important to consider the IS strategy. Figuroa-Flores et al. (2020) also mention that in order for the implementation of a new system to be successful, the company's mission and vision need to be carefully considered.

It is also important to have an organizational structure in place that supports data governance (Russom, 2008). In many cases this is a central data governance board or committee which decides the direction of the data governance program. There need to be diverse people included in this board from different departments, in order to ensure that there will be cross-functional collaboration between stakeholders, and it is important that senior management is included (Petzold et al., 2020). However, Russom (2008) mentions that there are also other organizational structures suitable for this. A structure that is cross-functional and collaborative can also work, for example data enterprise architecture groups or BI and data warehousing teams. Brous et al. (2016) suggest that this organizational structure should be made formal, as it needs to be clear who makes the decisions.

Additionally, a practical and enforceable data policy adds to the success of a data governance initiative (Mahanti, 2018). This policy should describe the guidelines and policies in place at an organization that ensure the appropriate use of an organization's data. A good policy is regularly reviewed and updated to make sure it reflects current stakeholder needs. Another factor that is critical for the success of a data governance program is a structure with clearly defined roles and responsibilities (Mahanti, 2018). The core roles of a data governance program are described in section X. In order to enforce accountability, it is important that it is documented who has responsibilities for the data and who makes the decisions. This way, people can be responsible over data and ensure that the data quality is effectively managed (Cheong & Chang, 2007). Additionally, a common vocabulary needs to be established in the organization, so that the organization represents the same data in a consistent way (DeStefano et al., 2016).

According to DeStefano et al. (2016), in order for a data governance program to be executed properly, there needs to be some kind of understanding of the data landscape in the organization. In order to know what data needs to be governed and how it should be governed, it is important that there is awareness in the organization of the data assets that they have. Panian (2010) mentions that before a data governance program can be executed, the existing landscape should be assessed.

A critical factor that is successful for the implementation of data governance, as well as any other implementation within an organization, is change management (Jones et al., 2015). Change management helps with the cultural changes required to transition into a new way of working. It helps overcome resistance from employees by making sure that employees have reached a certain level of readiness before the change is implemented. Employees need to be educated on the current problems together with how data governance is going to solve it (Mahanti, 2018). Communication and education should be aligned to the different stakeholder groups affected by the change. Training programs are important for making sure the employees know how to use the new system, and these should be required (Jami Pour & Hosseinzadeh, 2021). Besides, ideally, a communication plan is developed that communicates the reason for data governance and how it is going to be implemented, stressing the benefits that it will bring to the stakeholders (Imran et al., 2016). Sonenshein and Dholakia (2012) stress the importance of addressing why a certain change needs to be made in the communication, so that it can help with the understanding of employees.

Moreover, data governance should be a cross-functional initiative as it brings different business units and roles together and these should all be aligned (Mahanti, 2018). Data governance covers all parts of the organization that touch data, from entering it to the storage and analysis of it. This makes it important that there is cross-functional collaboration. As data governance is a cross-departmental and cross-functional undertaking, it is intertwined with many other initiatives and IT projects within an organization (Russom, 2008). One way to introduce data governance within an organization and to maximize its potential to be successful is to link it to another IT initiative (Ladley, 2012). Russom (2008) confirms this by stating that the scope of data governance is determined by initiatives and implementations that are already introduced in the organization. These initiatives are already at the top of the priority list, and require immediate attention. Petzold et al. (2020) describe a data governance initiative which finally became successful after it was introduced as part of an enterprise-wide analytics transformation.

It is also important that the organization takes an incremental approach to the implementation of data governance according to Mahanti (2018). A governance program that aims to implement too many changes in one go, often results in resistance from employees, confusion,

and discussions. By implementing data governance in phases, employees are also more inclined to consider it as a core business practice rather than a standalone project. One key business initiative should be chosen to focus on first, preferably one which faces most data challenges, before scaling the program to the rest of the organization (Panian, 2010). Besides prioritizing use cases and domains, it is also important to prioritize on a day-to-day basis, by reviewing data quality issues daily (Petzold et al., 2020).

Furthermore, as data governance is a long-term initiative, Mahanti (2018) says it is important to measure its progress. This can be in terms of costs, for example money loss due to poor quality data and associated business errors, or the amount of time spent on finding data (Petzold et al., 2020). Quantifying the outcome of the data governance program can help control the progress. There should be a well-defined plan for capturing and reporting on metrics. Ćirić and Raković (2010) highlight that the project should be continuously monitored.

What is crucial for the rest of the implementation project is to conduct concept tests in order to select the best advisory body or product (Figueroa-Flores et al., 2020). A service or product should be selected that best fits the project’s goals. Panian (2010) describes that a pilot project should first be started before implementing data governance, as that increases the chance of success. Besides, instead of implementing data governance all at once, it is better to first scope it to a specific data domain to ensure that the most important data is covered (Petzold et al., 2020). These domains should be prioritized based on the effort for transformation, regulatory pressure and any other important input for the organization, and preferably two or three domains should be picked to focus on first. Additionally, data can also be prioritized based on their level of criticality within each domain, as often only 10 to 20 percent of data is critical within an organization (Petzold et al., 2020).

Lastly, another factor that can help make a data governance program a success is the introduction of a data governance platform (Mahanti, 2018). The techniques offered by this platform offer capabilities that can help increase the effectiveness of a data governance program by atomizing the activities around data governance and by providing techniques for monitoring the data governance performance. However, as found in the study Mahanti (2018), the implementation of data governance tools had the lowest priority for the success of data governance as indicated by information management professionals. This was because the professionals indicated that data governance can be executed to a certain extent without the help of tools.

Table 7. Data governance success factors from literature

Concepts	Description	References
Commitment from leadership	Dedicated leaders in the organization that are committed to do the work necessary for establishing data governance	(Mahanti, 2018), (Al-Ruithe et al., 2019), (Figueroa-Flores et al., 2020), (Petzold et al., 2020)
Executive sponsorship	Having a sponsor in the organization that gives enough financial support to the initiative	(Mahanti, 2018), (Al-Ruithe et al., 2019), (Russom, 2008), (Ladley, 2012), (Panian, 2010), (Petzold et al., 2020)
Demonstration of business value	Showing the data governance program adds business value to every step	(Al-Ruithe et al., 2019), (Ladley, 2012), (Petzold et al., 2020)
Alignment between stakeholders	Stakeholders affected by the program agree with each other about the changes that have to be made	(Mahanti, 2018), (Russom, 2008)

Robust data governance strategy	The vision, the roadmap, and the milestones for the data governance program, aligned with the organizational strategy	(Mahanti, 2018), (Ćirić & Raković, 2010), (Figueroa-Flores et al., 2020), (DeStefano et al., 2016)
Alignment with other strategies	The data governance initiative is implemented in such a way that it supports the strategies of the organization	(Arvidsson et al., 2014), (Figueroa-Flores et al., 2020), (Cheong & Chang, 2007)
Data governance organizational structure in place	Have an organizational structure in place that supports data governance	(Russom, 2008), (Brous et al., 2016), (Cheong & Chang, 2007), (Khatri & Brown, 2010), (Otto, 2011), (Panian, 2010), (Petzold et al., 2020)
Well-defined data policies	The guidelines and policies that ensure the appropriate use of an organization's data	(Mahanti, 2018), (Al-Ruithe et al., 2019), (Cheong & Chang, 2007), (Khatri & Brown, 2010), (Panian, 2010)
Clear defined roles with responsibilities	Documentation of who has responsibilities for the data and who makes the decisions	(Mahanti, 2018), (Al-Ruithe et al., 2019), (Begg & Cairra, 2019), (Cheong & Chang, 2007), (Khatri & Brown, 2010), (Panian, 2010)
Common vocabulary	A consistent set of definitions of data used in the organization	(DeStefano et al., 2016)
High understanding of data landscape	Have an idea of where data in the organization is and how it is connected	(Brous et al., 2016), (DeStefano et al., 2016), (Panian, 2010)
Change management	Techniques for dealing with the cultural changes required to transition into a new way of working	(Mahanti, 2018), (Al-Ruithe et al., 2019), (Russom, 2008), (Jones et al., 2005), (Figueroa-Flores et al., 2020), (Ladley, 2012), (Petzold et al., 2020)
Trainings and education	Teaching employees how they have to carry out activities related to the new way of working	(Mahanti, 2018), (Al-Ruithe et al., 2019), (Cheong & Chang, 2007), (Jami Pour & Hosseinzadeh, 2021), (Figueroa-Flores et al., 2020), (Ladley, 2012), (Petzold et al., 2020)
Effective communication	Communicating the reason for data governance and how it is going to be implemented, together with the benefits that it will bring to the stakeholders	(Mahanti, 2018), (Al-Ruithe et al., 2019), (Sonenshein & Dholakia, 2012), (Imran et al., 2016), (Figueroa-Flores et al., 2020)
Cross-functional collaboration	Collaboration between different business units and roles together and aligning these	(Mahanti, 2018), (Cheong & Chang, 2007), (Russom, 2008), (Ladley, 2012)
Part of another transformation project	Introducing data governance within an organization by linking it to another IT initiative	(Russom, 2008), (Petzold et al., 2020)
Incremental approach to implementing data governance	Implementing data governance in smaller phases	(Mahanti, 2018), (Ladley, 2012), (Panian, 2010), (Petzold et al., 2020)
Measurements for tracking progress	Quantifying the outcome of the data governance program by measuring it, e.g. savings in costs	(Mahanti, 2018), (Al-Ruithe et al., 2019), (Cheong & Chang, 2007), (Ćirić & Raković, 2010), (Figueroa-Flores et al., 2020), (Alhassan et al., 2016), (Ladley, 2012), (Panian, 2010), (Petzold et al.,

		2020)
Concept tests or pilots	Testing out concepts, ideas, and products before actually implementing something	(Figuerola-Flores et al., 2020), (Ladley, 2012), (Panian, 2010), (Petzold et al., 2020)
Scope and prioritize data governance to a data domain	First implement data governance for a specific data domain instead of implementing it all at once	(Petzold et al., 2020)
Use of data governance tools	Make use of techniques offered by a data governance platform that automates activities related to data governance	(Mahanti, 2018), (Al-Ruithe et al., 2019), (Russom, 2008), (Ladley, 2012)
Establish master data management	Master data management ensures the quality of master data	(Petzold et al., 2020)

3.5.2 Situational factors for implementation of software systems

The situational factors of the implementation of software systems are summarized in table 8. An explanation of each factor is given below.

The Data Governance Institute (DGI) recommended that organizations should go from an informal to a formal data governance structure, when organizations grow so big that they can no longer manage their data by traditional practices, when they need to have a cross-functional approach rather than a siloed approach, and when regulatory rules require organizations to have data governance practices in place (Mahanti, 2018). Therefore, the extent to which organizations should have a data governance structure in place is dependent on the maturity of the organization. A mature organization is one that requires a comprehensive data governance solution, with a governance council, a high degree of automation, and a broader domain scope (Petzold et al., 2020). On the other hand, a less mature organization might have only a few periodic meetings with senior executives, data tracking in Excel, and a narrow domain scope with limited use cases. Regarding the implementation of a data governance platform, Huth & Vietor (2021) say that the maturity of the organization is important to consider when implementing technology. The first step is to determine whether the organization is ready to automate data governance activities, and thus maturity plays a crucial role (Ladley, 2012).

Imran et al. (2016) mention that change depends on the sector that the organization is part of. For example, high-tech industries such as insurance, banking, or telecom frequently change, and the organization needs to take this into account when implementing a change. Ladley (2012) also elaborated on this, by saying that organizations which find themselves in a highly regulated industry require a different approach to data governance. In highly regulated industries, there is a higher need for data discovery and tracking data lineage in order to ensure compliance (Ladley, 2012).

Besides that, Ronnenberg, Graham and Mahmoodi (2011) describe that the way a new system (in their study, an environmental management system) is implemented depends on the reason or goal of the implementation. For example, there might be strategic reasons (e.g. increasing market share), economic reasons (e.g. increasing revenue), or ethical reasons (e.g. improving environmental sustainability). Ćirić and Raković (2010) add to this that an organization can implement an information system for business survival, to achieve a competitive advantage, or due

to pressure from external factors. Moreover, when an important goal is to minimize the risk, organizations are more likely to take a different approach on the implementation of a new system, as there is a higher chance that changes are resisted (Arvidsson et al., 2014). Russom (2008) put this in the context of data governance, by mentioning that the starting point of a data governance program, for example focussing on data quality or on compliance, affects the way it is implemented.

When looking at contingency approaches, some theories focus on the implementation stages of new systems, where different stages require different interventions (Van Linge, 2006). Imran et al. (2016) support this by emphasizing that every stage in the implementation requires a different type of knowledge in order to properly complete the stage. The distinction made by Van Linge (2006) is an early, middle, and late phase of the implementation. In an early phase, structural and procedural characteristics dominate, and interventions should focus on structure, politics, and culture. There should be visionary leadership with a strong influence so that employees can be convinced and informed of the new changes. In a late implementation phase, characteristics of people have the upperhand, and interventions should be about human resources. Leadership ought to be more transactional and participative, where leadership should listen to wishes from employees and employees are given more responsibility. Examples of interventions at this stage are trainings, incentives, and feedback.

Another factor that can influence the acceptance of an innovation is the radicality of the change (Van Linge, 2006). Radical innovations require a great deal of changes in the existing practices in the organization, while incremental innovations have less impact on this. Different learning methods are needed to handle this change, where radical change requires new ways of thinking instead of extending existing knowledge. Moreover, in situations with radical changes, there needs to be more guidance and convincing from leadership.

The nature of the innovation can also affect the rate of success (Van Linge, 2006). Innovations that affect primary processes in the organization require a high degree of professionalism from employees, decentral decision making, and less standardization (i.e. rules and regulations). Innovation that affects support and administrative processes have a lower need for professionalism, but a high need for central decision making and more standardization.

The organizational culture is essential for the introduction of innovations in organizations (Van Linge, 2006). A distinction can be made between two types of environments in an organization: a participational and a controlling environment. A participational climate is characterized by a strong influence from employees on decision making. Here, the rate of implementation success is highest if the tactic is to accommodate, meaning adjustments are made based on employees and their way of working. A controlling environment can be distinguished by the strict supervision on performance and any deviations from the norm. An implementation tactic that focuses on negotiation and convincing has a higher contribution to the success. In both cases, giving incentives is an effective tactic. Moreover, the Competing Values Framework from Cameron and Quinn (2006) and the innovation contingency model from Van Linge (2006) (see section 3.3.4) suggest that the culture in an organization has a big influence on the way a change is accepted. Organizations where there is a flexible environment and a high focus on communication and employee involvement often prove to be more ready for the change of a new implementation (Jones et al., 2005).

Specifically for the implementation of a data governance platform, Begg and Cairn (2012) mention that it is not about the size of an organization, but rather about the nature or volume of data used by this organization. Therefore, the complexity of the data environment should be taken

into account when determining how a data governance platform can best be implemented (Petzold et al., 2020). Ladley (2012) describes that the type of data highly affects a data governance program. Structured data that comes in during a fixed period requires a different kind of attention than unstructured data which has to be kept for a long time and may be subject to certain privacy regulations. Besides that, data that is used in an explorative environment may not need as much rigorous data governance than data that is used in interactions with customers (Petzold et al., 2020). Data for which privacy is important are more likely to require non-disclosure agreements (NDA's), data masking, and fully available metadata.

Panian (2011) says that the existing data technology infrastructure has an influence on when and how a data governance program should be executed. In case the existing technology infrastructure is not able to support the data governance practices, a different approach should be taken. Ladley (2012) describes that the choice for a tool for a large extent depends on what the organization already owns in terms of technology. Staying with a known vendor from which several products are already used in the organization is more efficient than switching to a new vendor.

Moreover, Begg and Caira (2012) note that a contingency factor such as organizational structure influences the way a data governance program is set up. Jami Pour and Hosseinzade (2021) elaborated on this in the context of implementing a Customer Relationship Management (CRM) system, for which it was important that a flexible organizational structure was necessary for a successful implementation of the system. There is evidence of research into the dimension of centrality versus decentrality of an organization and how it affects information systems, as introducing roles often requires a hierarchy to be in place (Otto, 2011). Russom (2008) mentions that for data governance, there needs to be some kind of level of bureaucracy in order to ensure effective implementation. However, too much bureaucracy which requires a lot of coordination tends to be counterproductive. The extent to which an organization is federated with different divisions highly affects the scope of a data governance program (Ladley, 2012). Some kind of decision needs to be made about whether data should be managed centrally, regionally, or locally. Organizations that have divisions in many different geographies also have more complex needs for their data governance (Petzold et al., 2020).

Hussain et al. (2018) say that leadership style affects the way a change is implemented. Leaders of the organization have an effect on the outcome of a change implementation because of how they motivate employees, communicate the changes, and control the formation of alliances. This is confirmed by Begg and Caira (2012), who describe that the outcome of a data governance program can be influenced by the decision-making style of leaders. There are transactional leaders who are mainly concerned with rewards and punishments, but there are also transformational leaders who are charismatic and inspirational (Hussain et al., 2018). Khatri and Brown (2010) describe how decision-making on a lower level (i.e. decisions about the use of data) might be influential, as some domains require a more centralized approach to decision-making than others.

Moreover, the implementation history of other implementation efforts within the organization might affect the outcome (Jami Pour & Hosseinzadeh, 2021). Previous experience with similar technologies might help guide the implementation and it might also decline organizational resistance.

Table 8. Concept matrix for implementation situational factors from literature

Concepts	Description	References
Data governance maturity level	The extent to which the organization already has data governance in place	(Mahanti, 2018), (Schneider te al., 2021), (Begg & Caira, 2012), (Ladley, 2012), (Petzold et al., 2020)
Sector	Area in which organizations share the same activities or products / services	(Imran et al., 2016), (Ladley, 2012), (Petzold et al., 2020)
Goal	The main objective of why the system is implemented	(Ronnenberg et al., 2011), (Arvidsson et al., 2014), (Ćirić & Raković), (Russom, 2008), (Ladley, 2012), (Otto, 2011)
Implementation stage	The moment in time when an innovation is implemented; the maturity of the innovation	(Van Linge, 2006), (Imran et al., 2016), (AlManei et al., 2018)
Radicality of change	The extent to which the organization has to change in order for the innovation to be implemented	(Van Linge, 2006), (AlManei et al., 2018), (Hussain et al., 2018)
Nature of innovation	Whether the innovation affects primary or support processes	(Van Linge, 2006)
Organizational culture	The values, beliefs, and assumptions of people in the organization	(Van Linge, 2006), (Cameron & Quinn, 2006), (Jones et al, 2005), (AlManei et al., 2018), (Hussain et al., 2018), (Ronnenberg et al., 2011), (Jami Pour & Hosseinzadeh, 2021)
Data environment (type, nature, volume)	The nature and volume of data used in the organization	(Begg & Caira, 2012), (Ladley, 2012), (Petzold et al., 2020)
System infrastructure	What the current tool stack looks like within the organization	(Ladley, 2012), (Panian, 2010)
Organizational structure	The hierarchy, roles, and responsibilities within an organizations	(Begg & Caira, 2012), (Jami Pour & Hosseinzadeh, 2021), (Russom, 2008), (Ladley, 2012), (Otto, 2011), (Panian, 2010), (Petzold et al., 2020)
Decision-making style	The way decisions are made within the organization	(Begg & Caira, 2012), (AlManei et al., 2018), (Hussain et al., 2018), (Ronnenberg et al., 2011), (Jami Pour & Hosseinzadeh, 2021), (Russom, 2008), (Khatri & Brown, 2010)
Implementation history	The organization's previous experience with similar implementation efforts	(Jami Pour & Hosseinzadeh, 2021)

3.5 Framework design

According to Moullin et al. (2020), the effective use of frameworks can be very valuable for an implementation process. Frameworks play an important role in the generalization of implementation efforts across diverse situations in order to have a cumulative overview of evidence (Birken et al.,

2018). Birken et al. (2018) mention that frameworks and models can be considered as checklists of factors that apply to different aspects of an implementation. Whereas frameworks often focus on the identification of determinants that affect an implementation, models have a higher focus on translating findings into practice. The purpose of frameworks is to explain or organize information and the relationship between concepts. They consist of a narrative or graphical representation of concepts or factors that make up a certain phenomenon (Moullin et al., 2020). This phenomenon can be an implementation effort. Moullin et al. (2020) describe that a framework for an implementation effort can depict an implementation process (frequently representing a series of phases), factors that influence the implementation (also called determinants, barriers, or enablers), and implementation strategies used for guiding an implementation effort. Strifler, Barnsley, Hillmer, & Straus (2020) add to this that implementation frameworks can depict the underlying mechanisms or theories that explain the success of an implementation. The depth of the framework depends on the scope chosen, the aim, and the intended target for change. It might also be beneficial to combine multiple frameworks so that all factors can be considered in an implementation effort (Moullin et al., 2020).

Regarding their use in research, frameworks can help guide the design of a research, add to the theoretical thinking from researchers, and add to the interpretation of results (Moullin et al., 2020). From a practitioner's perspective, frameworks provide a common use of concepts so that it can be used as a consistent explanation of implementation-related factors, which leads to shared understanding between stakeholders (Birken et al., 2018). With a shared understanding about the implementation, it can help pick an implementation strategy and guide the implementation process. However, suboptimal use of a framework can affect the success of an implementation effort (Moullin et al., 2020). This happens when a framework has incorrect concepts or when it is not incorporated in the phases of an implementation. Therefore, it is important that a framework is designed and used effectively.

When selecting a framework, Moullin et al. (2020) describe that it is important that the following criteria are considered: (1) the purpose of the framework (e.g. describing, analyzing, evaluating), (2) the levels included (e.g. organization, individual), (3) the depth of analysis (e.g. process, determinants, strategies), and (4) the orientation which includes the intention for which the original framework was designed (e.g. a specific intervention, guidelines). For effective use of the framework, it is important that the project's purpose, scope, and intended target are described (Strifler et al., 2020). Moreover, Strifler et al. (2020) mention that a framework should be easy to use and accessible.

Some frameworks are more general than others which are more focused on a specific context. Frameworks can both expand or limit the factors or processes considered during an implementation effort (Moullin et al., 2020). On the one hand, it can expand the consideration by suggesting factors that would have not been thought of otherwise. On the other hand, it can limit the consideration by ignoring some of the factors that might have also been important during the implementation. One way to solve this is to make use of multiple frameworks that complement each other. Moreover, it is always important to know the context specific issues that should be solved, which can be achieved by involving stakeholders with a deep understanding of the situation. It is not necessary or possible for a framework to list all possible factors, so therefore the most important factors should be prioritized.

3.6 Conclusion

To conclude, data governance is essential for ensuring compliance, unlocking the value of data, and increasing data quality. This is especially relevant in current times as the volumes of data continue to grow at an exponential rate every year, and traditional solutions for data management can no longer handle these growing volumes. However, not all organizations succeed in realizing the benefits that a data governance program can provide. Therefore, it is important to understand what data governance means and what can be affected by it. Data governance is made up of multiple components: cross-functional effort; framework; data as a strategic enterprise asset; decision rights and accountabilities; data policies, standards, and procedures; and monitoring compliance. When these components are in place, a data governance tool or platform can be used to improve data governance further. The challenge is that there is no one-size-fits-all solution for this, and every organization needs a different approach for a successful implementation of such a tool. The theory that describes that an implementation depends on the situation of the organization is called the contingency approach. By looking at factors for ensuring a successful data governance program and by drawing from factors from the contingency approach, an overview can be gained of what needs to be addressed during a data governance implementation program. Coming back to the main research question *“What framework can be designed that supports an organization in the implementation of a data governance platform by considering the situational factors of the organization?”*, the success and situational factors identified in this chapter can be used as a foundation for the data governance implementation framework. In order to establish which data governance implementation success factors are actually relevant when a data governance platform is implemented, subject matter experts (SME) were consulted about their experience with data governance platforms. The results are discussed in the next chapter.

4. Results expert interviews

In the previous section, definitions and frameworks for data governance and the implementation of new technologies were identified, and success factors from literature were derived from this. In order to identify success factors specific for the implementation of data governance platforms, expert interviews were conducted with subject matter experts (SMEs) who had experience with data governance programs. During the interviews, the participants were asked to answer questions about one specific example they experienced of an organization where they worked on implementing a data governance platform. The general success factors for data governance and the implementation of new systems from literature were used as guidance throughout the interviews by checking whether these also applied to the implementation of data governance platforms (see Appendix C). After the interviews were conducted, the interviews were first summarized for the qualitative analysis (described in Appendix D). After that, the answers were analyzed and grouped into categories (see Appendix E), and for each category it was noted in which interview they occurred so a quantitative analysis could be made. Some categories were the same as the ones discovered in literature (from Appendix C) but new categories were added based on the answers that were provided by the SMEs (results are described in Appendix E). For each answer category, an explanation was noted according to the answers given by the SMEs during the interviews (see Appendix F). This chapter presents the results of the interviews.

In section 4.1, an overview is given of all the interviewees that participated in the expert interviews, together with an overview of the example organizations that were focused on during the interviews. Section 4.2 provides a summary of the main findings of each individual interview. In section 4.3 the analysis of the interview findings is provided.

4.1 Interviews overview

In total, 11 interviews were conducted with subject matter experts (SMEs) who had experience with the implementation of data governance platforms at organizations. All SMEs were contacted through the Analytics department from Avanade and all had many years of experience with the implementation of data platforms at clients. Some interviewees also had multiple years of experience with data governance platform implementation, while some had recently started on data governance platform implementation programs. Table 9 shows an overview of all the interviewees. In the table, information is given about the date when the interview was conducted, the interviewee's function, the interviewee's experience with data governance, and the organization that was taken as an example to focus on during the interview. The interviewee's and the organization's names are anonymized for privacy reasons, and they are given a number instead.

Table 9. Summary of interview participants

Inter- viewee #	Date of interview	Interviewee function	Data governance experience	Organization of focus
I1	09-03-2022	Analytics consultant, data architect	Multiple years of experience with data platforms and data governance platforms at clients, background as an enterprise data architect	O1

I2	09-03-2022	Data platform consultant	Multiple years of experience with data platforms, recently a data governance program	O2
I3	10-03-2022	Solutions architect	Multiple years of experience with implementation of data platforms and data governance platforms at clients, also familiar with the theory behind data governance	O3
I4	11-03-2022	Data platform consultant	Multiple years of experience with implementation of data platforms and data governance platforms at clients	O4.1, O4.2, O4.3
I5	11-03-2022	Technical advisory	Worked on multiple data governance implementation projects at clients, from an overall strategic level instead of technology	O5
I6	15-03-2022	Business technology consultant	Multiple years of experience with data governance projects at clients and a data governance research background	O6
I7	15-03-2022	Solutions architect	Two years of experience with data governance projects at clients	O7
I8	16-03-2022	Data platform consultant	Multiple years of experience with data platform implementation projects, limited experience with data governance	O8
I9	17-03-2022	Analytics consultant	Three years of experience due to a data platform implementation project, before that a great deal of experience with systems such as data warehouses	O3
I10	18-03-2022	Solutions architect	Multiple years of experience with data governance platforms at clients, first with data governance strategy implementation but also with data governance platforms	O9
I11	24-03-2022	Solutions architect	Multiple years of experience with delivery of data products at client from an architectural standpoint, hands-on experience with data governance platforms from a client project	O10

In table 10, the organizations which were focused on during the interviews are summarized. Information in this table is given about the sector the organization was active in, the location where the data governance platform implementation took place, the size or reach of the company, the organizational structure, the level of data maturity in the organization prior to the program, and the product that was implemented. In the cases where cells are left empty in the table, this specific information was not obtained during the interviews.

Table 10. Summary of organizations discussed in the interviews

Organization #	Sector	Project implementation location	Size	Organizational structure	Data maturity organization	Implemented product
O1	Utility	Australia	Small company	Start-up which grew organically, different business	Each business line managed its own systems and data, and	Collibra

				lines for each utility product	no centrally managed program	
O2	Financial	America	Big company		Distributed data, thought about data governance but no central data overview	Microsoft Purview
O3	Financial	Europe	Operating worldwide	Hierarchical for decisions, flat for communication	Established data governance board, but data governance activities in Excel	Azure Data Catalog
O4.1	Health	Europe	Operating worldwide	Member states with each their own structure	No centrally managed program around data, very basic data systems (mostly Excel)	Microsoft Purview
O4.2	Consumer goods	Europe			Many different systems without one centrally managed program	Microsoft Purview
O4.3	Manufacturing	Europe	Operating worldwide	Many brands with each their own structure	No central management or governance structure around data	Microsoft Purview
O5	Financial	America	Midsize company, operating worldwide	Central office which did reporting, multiple subsidiaries	Had experience with data governance and established roles and policies, but no data management platforms	Informatica Data Governance & Catalog
O6	Tourism	America	Operating worldwide	Lack of formal organizational structure regarding data management	Just completed a transformation to the Cloud, but no governance over this data	Microsoft Purview
O7	Financial	America	Midsize company	Relatively flat organization	Foundational data stores & integrations, established policies but no data ownership	Microsoft Purview
O8	Energy	Europe	Big company		Many different systems without one centrally managed program	Microsoft Purview
O9	Retail	Europe	Operating in multiple European countries		Just completed a transformation to the Cloud, but many data issues due to lack of governance	Microsoft Purview
O10	Energy	Europe	Operating in multiple European countries	Central programs are managed from headquarters, but local business units (BU) with local markets	Already on a Cloud journey, different BUs with different levels of maturity	Microsoft Purview

As Avanade is a Microsoft oriented company, most of the interviews focused on Microsoft’s solution for data governance, which is Microsoft Purview. However, as Microsoft Purview is a new product, all SMEs were familiar with data governance platforms from other vendors, and were asked during the interviews why a particular solution was chosen over solutions from these other vendors. There was one interview that focused on an initiative where Collibra was implemented, as Purview was not yet released during the time of the data governance platform implementation. Besides, Azure Data Catalog was the focus of one interview, which was a predecessor of Purview. Moreover, there was one interview that focused on Informatica Data Governance & Catalog, as the interviewee had not yet experienced a full implementation of Purview and therefore chose to focus on an organization that implemented Informatica’s solution.

4.2 Interview analysis

The interviews were analyzed both quantitatively and qualitatively. Firstly, summaries of the interviews were made (see Appendix D) for the qualitative analysis. Based on these summaries, situational and success factors that occurred in the interviews were identified. Secondly, the answers from the interview were classified into categories by using open coding techniques (see section 2.2 for an explanation of the open coding techniques). Checklists of success factors were made from literature to guide the interviews (see Appendix C), and with the open coding techniques it was tested whether these checklists also applied to the interviews. For each interview question, the answers of the experts were categorized by labeling the answers. An overview of this is given in Appendix E. For each interview question, it was calculated which answer category occurred the most throughout the eleven interviews for the quantitative analysis. The categories with the highest number were considered as most important. For each interview question (from the interview question list from Appendix B) and answer category (from the overview in Appendix E) an explanation from the answer was given as to why this answer category was, and these are given in Appendix F. These answer rationales were noted as accurately as possible as how the interviewee provided the answer, citing where possible (some interviews were translated to English and sometimes answers were shortened or better sentence structures were made).

4.2.1 Qualitative analysis

In Appendix D, the interviews are summarized. Based on this, in table 11 it was noted for each interview and example case, what platform and functionalities were implemented, what the situational factors of the organization were, and what the success factors were to make the implementation successful.

Table 11. Summary of situational and success factors from the SME interviews

#	Platform	Situational factors	Success factors
I1	Collibra (data catalog, glossary)	<i>Sector:</i> utility <i>Age:</i> start-up <i>Division:</i> different business lines operating autonomously <i>Data structure:</i> each business line	<ul style="list-style-type: none"> - Have high support from top-level management to guide the change (top-down) - Set up a data governance committee that manages the changes from a central point - Implement pilot version / proof of concept first to

		with its own system and data <i>Systems:</i> many point-to-point integrations between systems <i>Data governance maturity:</i> no enterprise wide strategy <i>Goal:</i> improve data quality	demonstrate the platform's use for the business
12	Purview (data catalog, glossary)	<i>Sector:</i> financial <i>Size:</i> big <i>Type of data:</i> lots of PI data <i>Data structure:</i> distributed data in different systems <i>Data governance maturity:</i> domains structured according to strategy <i>Goal:</i> use data to drive business analysis	<ul style="list-style-type: none"> - Have visionary leadership that clearly communicates the data strategy, vision, and intended outcome across all departments - Create a central and unified set of data definitions - Create data architecture where data is treated as a product around a data domain (i.e. data mesh approach)
13	Azure Data Catalog (data catalog, glossary)	<i>Sector:</i> financial <i>Hierarchy:</i> formal regarding decisions, informal regarding communication <i>Data governance maturity:</i> central data governance board <i>Goal:</i> one central overview of data	<ul style="list-style-type: none"> - Have visionary leadership that clearly communicates the data strategy, vision, and intended outcome across all departments - Assign roles and responsibilities to data - Make sure there is a direct integration between the data source and the data governance platform so data sources can easily be imported in the system
14	Purview	General <i>Sector</i> <i>Data governance maturity</i> <i>Data structure</i>	<ul style="list-style-type: none"> - Have high support from top-level management to guide the change (top-down) - Assign roles and responsibilities to data - Set up policies that ensure the right data access and use - Have a high understanding of the origin of the data - Make sure there is a direct integration between the data source and the data governance platform so data sources can easily be imported in the system - Create data architecture where data is treated as a product around a data domain (i.e. data mesh approach)
14.1	Purview (data catalog, glossary)	<i>Sector:</i> health <i>Division:</i> member states in different countries <i>Data structure:</i> distributed across member states <i>Systems:</i> mostly excel sheets <i>Goal:</i> increase data quality	<ul style="list-style-type: none"> - Create a central and unified set of data definitions - Connect the system to a central data platform that first standardizes data instead of the original source
14.2	Purview (data catalog, glossary)	<i>Sector:</i> consumer goods <i>Goal:</i> get one central overview of data	<ul style="list-style-type: none"> - Combine platform with data quality management system or functionality to manage the quality of data
14.3	Purview (data catalog, glossary)	<i>Sector:</i> manufacturing <i>Division:</i> many different brands within same organization <i>Data structure:</i> each brand owns and controls their data	<ul style="list-style-type: none"> - Set up policies that ensure the right data access and use
15	Informatica (data catalog, data classifications, workflow)	<i>Sector:</i> financial <i>Size:</i> midsize <i>Division:</i> many subsidiaries in different countries	<ul style="list-style-type: none"> - Have support from a sponsor that is willing to invest enough resources until the outcome is reached - Have high support from top-level management to

	management, notifications)	<i>Decision-making style:</i> subsidiaries report to central office <i>Data governance maturity:</i> roles, ownership, and policies in place <i>Goal:</i> comply with new regulation	guide the change (top-down) - Assign roles and responsibilities to data - Connect the system to a central data platform that first standardizes data instead of the original data source
16	Purview (data catalog, glossary)	<i>Sector:</i> tourism <i>Data governance maturity:</i> no data governance structure <i>Goal:</i> stay competitive with industry	- Create personas and use cases to determine what should be included in the system - Create a central and unified set of data definitions - Have a high understanding of the origin of the data - Get the data catalog and glossary in the system first before adding other functionalities - Classify a few sources successfully first before scaling to more sources - Combine platform with data quality management system or functionality to manage the quality of data
17	Purview (data catalog, glossary, lineage)	<i>Sector:</i> financial <i>Age:</i> 25 years <i>Size:</i> midsize <i>Hierarchy:</i> flat organization <i>Data governance maturity:</i> data stores and data integrations in place <i>Goal:</i> improve current (data analytics) processes	- Have support from a sponsor that is willing to invest enough resources until the outcome is reached - Have visionary leadership that clearly communicates the data strategy, vision, and intended outcome across all departments - Assign roles and responsibilities to data - Create a central and unified set of data definitions - Combine platform with data quality management system or functionality to manage the quality of data
18	Purview (data catalog, lineage, policy rules)	<i>Sector:</i> energy <i>Size:</i> big <i>Data structure:</i> distributed data <i>Systems:</i> mostly specialized on-premise systems <i>Data governance maturity:</i> no clear strategy <i>Goal:</i> find data easily	- Create personas and use cases to determine what should be included in the system - Set up policies that ensure the right data access and use - Make sure there is a direct integration between the data source and the data governance platform so data sources can easily be imported in the system - Combine platform with data quality management system or functionality to manage the quality of data
19	Purview (data catalog, glossary)	<i>Sector:</i> financial <i>Size:</i> big <i>Data governance maturity:</i> Azure Data Catalog was implemented <i>Goal:</i> to stay competitive with data governance platform offerings	- Have support from a sponsor that is willing to invest enough resources until the outcome is reached - Make sure there is a direct integration between the data source and the data governance platform so data sources can easily be imported in the system - Connect the system to a central data platform that first standardizes data instead of the original data source
110	Purview (data catalog, glossary, lineage)	<i>Sector:</i> retail <i>Size:</i> big (worldwide) <i>Market:</i> offline and online product offerings <i>Type of data:</i> sensitive data <i>Data governance maturity:</i> central data platform but no governance <i>Goal:</i> higher data understanding to ensure compliance and improve efficiency	- Set up a data governance committee that manages the changes from a central point - Assign roles and responsibilities to data - Get the data catalog and glossary in the system first before adding other functionalities - Include as many data sources as possible in order to cover the whole data landscape
111	Purview (data	<i>Sector:</i> energy	- Assign roles and responsibilities to data

catalog, glossary, lineage, classification)	<p><i>Division:</i> business units (BU) upcoast in different countries</p> <p><i>Market:</i> BU with own local market and branding</p> <p><i>Decision-making style:</i> centrally managed from headquarters</p> <p><i>Data structure:</i> varying levels of data maturity between BU's</p> <p><i>Data governance maturity:</i> central security and governance department (but not sure whether BUs complied with that)</p> <p><i>Goal:</i> standardize data</p>	<ul style="list-style-type: none"> - Set up policies that ensure the right data access and use - Create a central and unified set of data definitions - Connect the system to a central data platform that first standardizes data instead of the original data source - Implement pilot version / proof of concept first to demonstrate the platform's use for the business - Classify a few sources successfully first before scaling to more sources. - Create an open culture which encourages embracing change
---	--	---

Taking into account all interviews, the following success factors were derived that are given in table 12. For every success factor, the literature review was consulted to confirm whether these factors were confirmed by literature. The factors that were not confirmed by literature might be due to the lack of evidence about data governance platforms in literature. These are taking into account for the discussion in chapter 7.

Table 12. Success factors derived from interviews

Success factor	Source interview	Confirmed in literature
Assign roles and responsibilities to data	I3, I4, I5, I7, I10, I11	(Mahanti, 2018), (Al-Ruithe et al., 2019), (Begg & Cairn, 2019), (Cheong & Chang, 2007), (Khatri & Brown, 2010), (Panian, 2010), (Petzold et al., 2020)
Create a central and unified set of data definitions	I2, I4.1, I6, I7, I11	(DeStefano et al., 2016)
Set up policies that ensure the right data access and use	I4, I4.3, I8, I11	(Mahanti, 2018), (Al-Ruithe et al., 2019), (Cheong & Chang, 2007), (Khatri & Brown, 2010), (Panian, 2010)
Have a high understanding of the origin of the data	I4, I6	(Brous et al., 2016), (DeStefano et al., 2016), (Panian, 2010)
Get the data catalog and glossary in the system first before adding other functionalities	I6, I10	(Mahanti, 2018), (Ladley, 2012), (Panian, 2010)
Connect the system to a central data platform that first standardizes data instead of the original data source	I4.1, I5, I9, I11	-
Set up a data governance committee that manages the implementation from a central point	I1, I10	(Brous et al., 2016), (Cheong & Chang, 2007), (Khatri & Brown, 2010), (Otto, 2011), (Panian, 2010), (Petzold et al., 2020)
Have high support from top-level management to guide the implementation (top-down)	I1, I4, I5	(Mahanti, 2018), (Al-Ruithe et al., 2019), (Figueroa-Flores et al., 2020), (Petzold et al., 2020)
Implement a pilot version / proof of concept first to demonstrate the platform's use for the business	I1, I11	(Figueroa-Flores et al., 2020), (Ladley, 2012), (Panian, 2010)

Combine platform with data quality management system or functionality to manage the quality of data	I4.2, I6, I7, I8	(Petzold et al., 2020)
Have visionary leadership that clearly communicates the data strategy, vision, and intended outcome across all departments	I2, I3, I7	(Mahanti, 2018), (Al-Ruithe et al., 2019), (Figuroa-Flores et al., 2020), (Petzold et al., 2020)
Create an open culture which encourages embracing change	I11	(Mahanti, 2018), (Al-Ruithe et al., 2019), (Russom, 2008), (Jones et al., 2005), (Figuroa-Flores et al., 2020), (Ladley, 2012), (Petzold et al., 2020)
Have support from a sponsor that is willing to invest enough resources until the outcome is reached	I5, I7, I9	(Mahanti, 2018), (Al-Ruithe et al., 2019), (Russom, 2008), (Ladley, 2012), (Panian, 2010), (Petzold et al., 2020)
Create personas and use cases to determine what should be included in the system	I6, I8	(Petzold et al., 2020)
Create data architecture where data is treated as a product around a data domain (i.e. data mesh approach)	I2, I4	-
Make sure there is a direct integration between the data source and the data governance platform so data sources can easily be imported in the system	I3, I4, I8, I9	(Ladley, 2012)
Classify a few sources successfully first before scaling to more data sources	I6, I11	(Mahanti, 2018), (Ladley, 2012), (Panian, 2010), (Petzold et al., 2020)
Include as many sources as possible in order to cover the whole data landscape	I10	-

The following situational factors from the organizations were mentioned in the interviews as a characteristic which might have affected the implementation of the data governance platform or the way the data governance program was handled. These are shown in table 13. Again, they are checked for whether they also occurred in literature. The situational factors that were not confirmed by literature were not taken into account for the rest.

Table 13. Situational factors derived from interviews

Situational factor	Source interview	Examples	Confirmed in literature
Sector	I1, I2, I3, I4, I4.1, I4.2, I4.3, I5, I6, I7, I8, I9, I10, I11	Utility, financial, energy, health, manufacturing	(Imran et al., 2016), (Ladley, 2012), (Petzold et al., 2020)
Age	I1, I7	Start-up, 25 years	-
Size	I2, I5, I7, I8, I9, I10	Midsized, big	-
Hierarchy	I3, I7	Flat, formal regarding decisions	(Begg & Cairns, 2012), (Otto, 2011)
Division	I1, I4.1, I4.3, I5, I11	Autonomous business lines, BU's in different countries	(Jami Pour & Hosseinzadeh, 2021), (Ladley, 2012), (Panian, 2010), (Petzold et al., 2020)

Decision-making style	I5, I11	Subsidiaries report to central office, centrally managed from headquarters	(Russom, 2008), (Otto, 2011)
Market	I10, I11	Offline and online products, BU with own local market and branding	-
Type of data	I2, I10	PI, sensitive	(Begg & Caira, 2012), (Ladley, 2012)
Data structure	I1, I2, I4, I4.1, I4.3, I8, I11	Each business line with its own system and data, distributed data	(Begg & Caira, 2012), (Ladley, 2012)
Systems	I1, I4.1, I8	Many point-to-point integrations, specialized on-premise systems, mostly Excel sheets	(Ladley, 2012), (Panian, 2010)
Data governance maturity	I1, I2, I3, I4, I5, I6, I7, I8, I9, I10, I11	No enterprise-wide strategy, central data platform, central data office	(Mahanti, 2018), (Schneider et al., 2021), (Begg & Caira, 2012), (Ladley, 2012), (Petzold et al., 2020)
Goal	I1, I2, I3, I4.1, I4.2, I5, I6, I7, I8, I9, I10, I11	Improve data quality, use data to drive business analysis, one central overview, comply with new regulation, standardize data	(Ronnenberg et al., 2011), (Arvidsson et al., 2014), (Ćirić & Raković), (Russom, 2008), (Ladley, 2012), (Otto, 2011)

4.2.2 Quantitative analysis

In this section, the quantitative results are discussed. The full list of categories and their rankings can be found in Appendix E. After every success factor or situational factor, the score of the quantitative analysis was noted and these were used to decide on the priority of the factors. Whether a success factor or a situational factor was discussed in a question was based on the qualitative analysis (table 12 and table 13).

Success factors

The following success factors were derived, ranked from most important to least important:

1. *Make sure there is a direct integration between the data source and the data governance platform so data sources can easily be imported in the system (18)*. In order for the platform to be implemented with the least amount of effort, it is important that the platform can easily be connected to the data sources that should be scanned (Q23: I1, I2, I4). The product choice for almost all data governance platforms was based on how easily it integrated with the existing infrastructure (Q12: I2, I3, I4, I5, I6, I7, I8, I10, I11)
2. *Assign roles and responsibilities to data (10)*. In many cases, the organizations had to assign ownership to data before the data governance platform implementation took place, so that someone was responsible for the dataset (Q14: I3, I5, I6, I9, I11).
3. *Get the data catalog and glossary in the system first before adding other functionalities (10)*. Regarding the implementation of the platform, first the data catalog and glossary need to be established as the metadata is needed to use the platform (Q23: QI6, I9). In almost all case organizations, the organizations established both the data catalog and glossary as a first step

- (Q11: I1, I2, I3, I4, I5, I6, I7, I9, I10, I11). After the glossary is established, it can be extended to adding policies, access rules, data lineage, etc (Q23: I6, I11).
4. *Create a central and unified set of data definitions (9)*. At the start of the data governance platform implementation program, most organizations worked on establishing a unified set of terms for their glossary, so that data would be used the same within a domain (Q14: I2, I4, I5, I6, I11).
 5. *Set up policies that ensure the right data access and use (9)*. Changes that needed to be made to ensure the data governance platform implementation could go well included setting processes in place that revolved around who had access to data and how the data would be used (Q14: I1, I2, I4, I5, I6, I10, I11).
 6. *Classify a few sources successfully first before scaling to more data sources (9)*. Once a few sources are successfully implemented, the platform can be scaled to more sources (Q23: I11). Many interviewees indicated that the platform was still a work in progress as it had to be scaled (Q19: I1, I2, I3, I4, I5, I7, I8, I9, I10).
 7. *Have high support from top-level management to guide the implementation (top-down) (8)*. One of the challenges included the reluctance from employees to change their traditional way of working or unclarity about what they needed to change (Q16: I2, I3, I5, I10, I11). In the interviews it was indicated that it was important there was a clear mandate from top-level management so that employees had to change (Q22: I1, I2, I3, I5).
 8. *Create personas and use cases to determine what should be included in the system (8)*. In many organizations, there was a lack of a formal plan for what data should be managed and how it should be managed (Q7: I1, I4, I6, I9, I11) To decide which glossary terms and classifications should be included in the system, personas and use cases need to be established for the foundational work (Q23: I6).
 9. *Have visionary leadership that clearly communicates the data strategy, vision, and intended outcome across all departments (8)*. Leadership was indicated as an important factor as they could communicate to the rest of the organizations what needed to happen. The better the communication, the more likely the employees were willing to change their way of working (Q22: I1, I2, I3, I5).
 10. *Have a high understanding of the origin of the data (7)*. Challenges during the data governance platform implementation occurred frequently because organizations did not have a clear idea where their data originated from (Q16: I4, I6, I8).
 11. *Create an open culture which encourages embracing change (7)*. Data governance is an ongoing effort, and in order to ensure that the platform will be used correctly after its implementation, there should be strong principles (Q23: I1). Commitment to change from stakeholders was important for this (Q22: I1, I7, I9, I11).
 12. *Have support from a sponsor that is willing to invest enough resources until the outcome is reached (6)*. In many interviews, it was indicated that the funding and commitment from the sponsor was important for the continuance of the project (Q22: I5, I7, I9).
 13. *Connect the system to a central data platform that first standardizes data instead of the original data source (6)*. In some cases, the data governance platform could not be directly connected to the data source, either because the platform could not be connected to the source or the data would come in a form that was too raw. Therefore, in multiple cases the data was first brought into a central data platform (Q14: I4, I9, I11).

14. *Create data architecture where data is treated as a product around a data domain (i.e. data mesh approach (6)).* Data should be ordered around a domain to get most value from it, so that there can be a domain expert and a product owner responsible for it (Q23: I4). Multiple organizations created such a data architecture before they implemented the system (Q14: I2, I3, I4).
15. *Combine platform with data quality management system or functionality to manage the quality of data (5).* Ensuring that there is also a solution for data quality management makes sure that most value can be gained from the platform (Q23: I1, I4, I6, I8). This can be a solution in the platform itself or another complementary solution such as a master data management platform. However, it should be noted that this is not necessary in some situations (Q23: I8) and that there is no prerequisite for implementing it before or after the data governance implementation (Q23: I6).
16. *Implement a pilot version / proof of concept to demonstrate the platform's use for the business (3).* In order to demonstrate the use of the platform and ensure it is used in the right way, proof of concepts (PoC's) are required (Q23: I1). In some cases, the organizations had only implemented the PoC and did not do a full implementation (Q19: I6, I8, I10).
17. *Include as many sources as possible in order to cover the whole data landscape (3).* In some cases it was mentioned that the more systems and data were covered by the data governance platform, the better the overview of the data so that important decisions could be made from it (Q22: I4, I10)
18. *Set up a data governance committee that manages the implementation from a central point (1).* Multiple organizations had a central data governance office or roles in place before the implementation started (Q7: I3, I7, I11), so this is an indicator that there should be some form of central function in place at the start of the program.

Situational factors

The following situational factors were confirmed in the interviews, ranked from most important to least important:

1. *Data governance maturity level (11).* Multiple interviewees indicated that the data governance maturity level is an important indicator for the way a data governance platform should be implemented (Q18: I3, I6, I7). Firstly, before anything happens, a data governance framework should be there, which means that there should be a specification of the roles, policies and processes (Q23: I1, I4). The implementation of tools should only take place after the foundations are in place (Q23).
2. *Sector (11).* The sector can affect the way a data governance platform should be implemented (Q23). For example, the financial sector has to deal with a great deal of regulations and the public sector has many protected silos on premise which makes it difficult to scan data (Q23: I4).
3. *Goal (11).* The goal of the data governance platform implementation is determinant for how the system should be implemented (Q9). For example, a data governance program that is intended to comply with a specific regulation should be handled differently than a program that is intended to discover data for analytics.
4. *Division (8).* There are multiple example cases in which a data governance platform was needed because the organization had many subsidiaries / divisions / business units that were

all handling data in their own way (Q9: I1, I2, I3, I4, I5, I8, I10, I11). It should be noted that the implementation of a data governance platform is not about the size or the hierarchy of the organization (Q23: I4). It is about the need to create a central overview when an organization is very divided.

5. *Systems infrastructure (6)*. The systems used to produce data (e.g. on-premise, Cloud) affect the choice for a data governance platform, as the systems need to be connected to the platform (Q23: I1, I2, I8). Data originated in the Cloud could be most easily be connected (Q15: I2, I6, I7, I8, I9, I10, I11) but in some cases, a separate data management platform had to be used because the data governance platform could not immediately be connected to the data source (Q15: I4, I5, I7, I9, I10, I11).
6. *Type of data (5)*. Multiple interviewees indicated that the sources to include in the system were based on the type of data (Q18). Regarding the type of data, sensitive or personal data needs a different approach because it needs to be handled more carefully regarding regulations (Q23: I4). Moreover, transactional data from which no business value can be gained should also be handled differently (Q23: I2).
7. *Complexity of data sources (5)*. In order for a data governance platform to bring benefits for an organization, there should be multiple complex data sources or a complicated data lake included in the system (Q23: I2, I4, I8, I11). Most frequently, changes that needed to be made before the implementation took place were changes related to the data architecture (Q14: I2, I3, I4, I9, I11).
8. *Number of data sources (5)*. In order for a data governance platform to be very beneficial for an organization, there should be multiple data sources included (Q23: I2, I4, I8, I11). The idea of a data governance platform is that you want to have an overview of all data in the organization (Q23: I2).
9. *Culture (2)*. In some cases, the commitment and willingness of the organization to try out the data governance platform was determinative for the adoption of the system which was a result of the culture in that organization (Q18: I7, I11).

4.3 Conclusion

In summary, 11 subject matter experts (SMEs) were consulted in order to discover what a data governance platform implementation project looks like and to what extent success factors derived from literature for the implementation of data governance were still applicable to a data governance platform implementation in practice. During the interviews, situational factors were also discovered that affected the implementation. The factors were prioritized based on the quantitative analysis, and the most important success and situational factors were taken into account for the design of the framework, which is explained in the next chapter. In line with the goal of the research, this framework intends to guide the organization in the implementation of a data governance platform by giving a set of recommended success factors to focus on first.

5. Framework design

A framework can consist of both a narrative and graphical representation (Moullin et al., 2020), and this chapter specifies both representations. Firstly, the success factors and situational factors were listed in a narrative form. These were based on both the literature review and the interview results; the data governance implementation factors from literature were used as a guide throughout the interviews and design of the framework to check whether these applied in practice to data governance platforms as well. The ranking of the factors was then based on the interview results. Secondly, the most important dimensions were used to visualize the factors in a graphical representation. To decide on what the visualization would look like, results and frameworks from literature were used. In order to link factors to the framework, the interview results were used.

5.1 Success factors

The success factors are the enablers for the implementation of the data governance platform, meaning that they help make a successful implementation possible (Moullin et al., 2020). They can be used by an organization as a checklist for factors that should be in place during the implementation. From the results chapters, it became clear that the most important success factors are as follows (ranked on importance):

1. Make sure there is a direct integration between the data source and the data governance platform so data sources can easily be imported in the system.
2. Assign roles and responsibilities to data.
3. Get the data catalog and glossary in the system first before adding other functionalities.
4. Create a central and unified set of data definitions.
5. Set up policies that ensure the right data access and use.
6. Classify a few sources successfully first before scaling to more data sources.
7. Have high support from top-level management to guide the implementation (top-down).
8. Create personas and use cases to determine what should be included in the system.
9. Have visionary leadership that clearly communicates the data strategy, vision, and intended outcome across all departments.
10. Have a high understanding of the origin of the data.
11. Create an open culture which encourages embracing change.
12. Have support from a sponsor that is willing to invest enough resources until the outcome is reached.
13. Connect the system to a central data platform that first standardizes data instead of the original data source.
14. Create data architecture where data is treated as a product around a data domain (i.e. data mesh approach).
15. Combine platform with data quality management system or functionality to manage the quality of data.
16. Implement a pilot version / proof of concept to demonstrate the platform's use for the business.
17. Include as many sources as possible in order to cover the whole data landscape.
18. Set up a data governance committee that manages the implementation from a central point

5.2 Situational factors

The situational factors are the determinants of the framework, meaning that they are factors that influence the implementation (Moullin et al., 2020). During an implementation, these factors should be considered for deciding what the best way is to implement a data governance platform. The results showed that the most important dimensions for a data governance program are as follows (ranked on importance):

1. Data governance maturity level.
2. Sector.
3. Goal.
4. Division.
5. Systems infrastructure.
6. Type of data.
7. Complexity of data sources.
8. Number of data sources.
9. Culture.

5.3 Visualization

From the literature on implementation frameworks, it became clear that it might be beneficial to use multiple frameworks that complement each other (Moullin et al., 2020). Resultantly, it was decided to use two frameworks; one that showed the implementation process in a series of phases, and one that focused more on implementation strategies based on some characteristic of the organizations. The design and application of the frameworks are discussed below.

5.3.1 Framework A

The first visual framework was intended as a way to guide the implementation process by showing what an organization needs to do to go to the next step. The dimension that was focused on for this stepwise representation was related to the level of data governance maturity. Ladley (2012) indicates that the most important factor for the success of a data governance platform implementation is the readiness of the organization to automate data governance activities. During the interviews, it became clear that different actions or factors are needed at different points in time. This was confirmed by Petzold et al. (2020) who describes that there are varying levels of data governance models which each require a different set of guidelines. Taking this into account with the data governance maturity level as the highest ranking situational factor, it was decided for the dimension of the graphical representation to focus on the situational factor *data governance maturity level*. A set of success factors were grouped together and linked to a data governance platform maturity level. The resulting framework can be seen in figure 9.

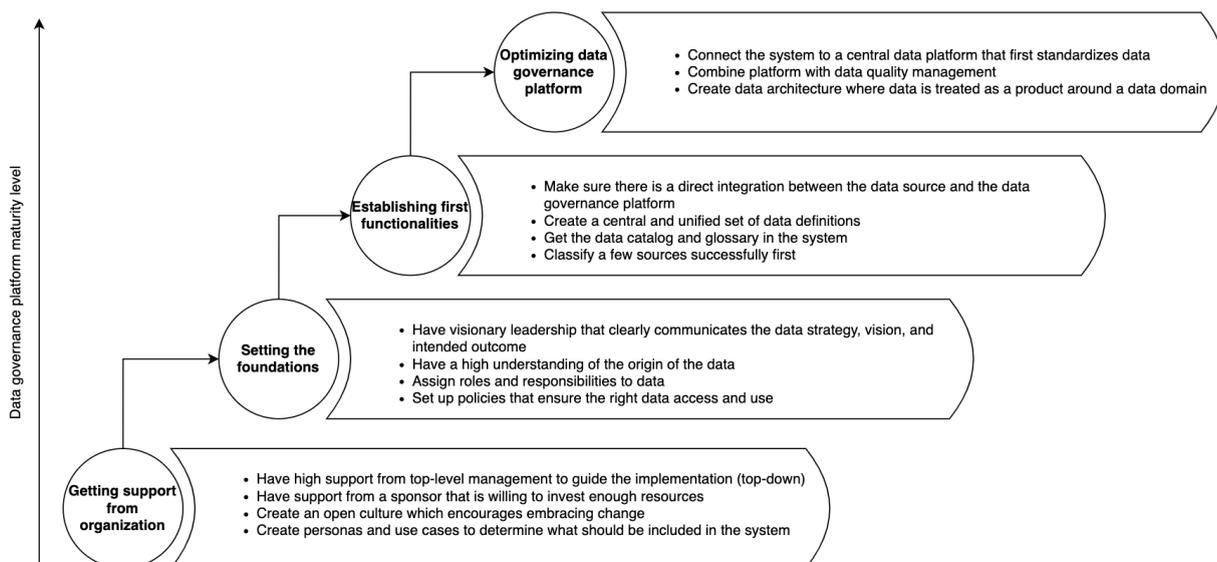


Figure 9. Framework A - data governance platform implementation based on data governance maturity level

The goal of the framework is to guide an organization in their data governance platform implementation process by making sure they have every factor of one stage in place before they move to the next one. This helps the organization to prioritize what factors they have to focus on first, so that they do not try to implement data governance all at once. An organization can apply this framework by checking for a stage whether they have all the factors associated with this stage. This starts at the bottom with the lowest level of data governance platform maturity and once they have all factors in place of one stage, they can move to the next. If factors are missing for a certain stage, the organization has to implement it before moving a stage higher.

5.3.2 Framework B

The second visual framework was intended as a way to help decide what level of automation is needed based on characteristics in the organization. Guiding frameworks from literature that considered contingency approaches were used for the design of the visualization (see section 3.2.4 and 3.3.4) (Cameron & Quinn, 2006; Petzold et al., 2020, Van Linge, 2006). For the dimensions it was chosen to focus on the data environment, as Petzold et al. (2020) mention that the level of data governance must be adjusted to the data set that it is applied to. Having rigorous data standards can ensure quality and limited access which minimizes risk, but it can also minimize innovation. By adjusting the level of data governance to a specific data set, the risks and opportunities can be balanced. As *type of data* was ranked as the most important data-related situational factors, this was included as one of the dimensions for the data environment. On the one hand, this could be transactional data (i.e. all data resulting from business events), and on the other hand this could be business critical data (i.e. data that is used in decision-making). Besides that, a comprehensive model requires a high degree of automation including a data catalog, data lineage and a broader scope, while a targeted model might have limited lineage tracking and few use cases (Petzold et al., 2020). To determine whether a model was considered comprehensive, the situational factor *number of data sources* was used. This dimension ranged from a few specific sources to sources in the entire enterprise. As it was assumed that the organizational foundations were already in place for the

platform implementation (as depicted in framework A), only the success factors which specifically focus on the data governance platform were taken into account for this visualization. The resulting framework is shown in figure 10.

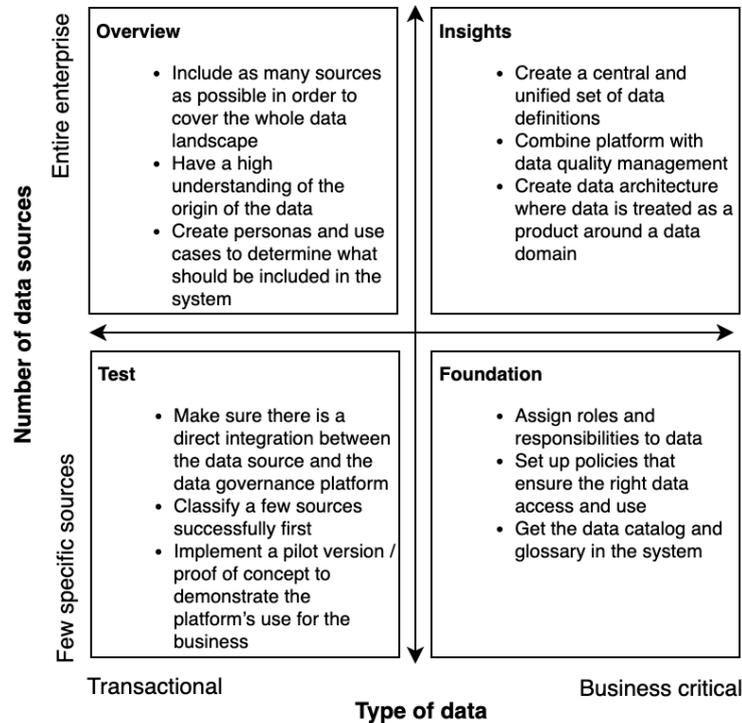


Figure 10: Framework B: data governance platform implementation based on data environment

The goal of this framework is to help organizations decide what they should implement in the data governance platform. Based on the dimensions, the organization can select a strategy for the implementation. The framework can be applied by first considering the dimensions of the data that they want to include in the system. Firstly, for the type of data, the organization should consider whether they want to include transactional or business critical data. Secondly, they should consider whether they want to cover only a few specific sources, or as many sources as possible in the enterprise. When taking both dimensions together, a quadrant can be selected which determines the implementation strategy. Each quadrant contains a set of factors that the organization should focus on for inclusion in the data governance platform. This is intended as guidelines; if the organization wants to include data which can be placed in multiple quadrants, then the organization can consider factors from both quadrants.

5.4 Conclusion

From the literature and interview results, 18 success factors and 9 situational factors were identified. When factors occurred more frequently or were specifically emphasized in interviews, they gained a higher rank of importance (see chapter 4 on the qualitative and quantitative results). For the success factors, the top factors were related to implementing the first functionalities in the platform; making sure there is a direct integration between source and platform, implementing the catalog and

glossary, and classifying a few sources. They were also related to setting up a data governance framework for data; assigning data ownership and setting up data policies. Besides that, the highest ranking situational factors were related to the organizational side of data governance; the maturity level, the goal, and the sector. These factors were then visualized into two complementary frameworks; one focusing on the process of implementation (A), and one focusing on picking a strategy of focus for the implementation (B). The guiding dimension for framework A was the data governance platform maturity level, whereas the guiding dimensions for framework B were the type of data and the number of sources to include in the platform. Whether the success factors were grouped accurately along these dimensions was validated by means of expert interviews, which is discussed in the next chapter.

6. Framework validation

The frameworks, both the narrative form (i.e. the list of factors) and the visual form, were validated by means of expert interviews. On one hand, this took the form of a case study. Organizations that completed a data governance platform implementation or were thinking about it were reached and data about them were gathered by means of an interview with a practitioner of that organization. On the other hand, researchers with an academic background in data governance were approached. The framework was validated in iterations, so in-between the iterations the visual framework could be updated to maximize its utility (Wynn et al., 2007). Firstly in this chapter an overview is given of the iterations. After that, for each iteration a description is given about the interviewee's background and view on data governance, a list of points of improvements for the framework, and a revised framework.

6.1 Validation interviews overview

Three iterations took place, and an overview of them can be seen in table 14. Every first interviewee (I... .1) of an iteration was a practitioner and every second interviewee (I... .2) of an iteration had an academic background. The points of improvements from both interviews were synthesized during each iteration.

Table 14. Overview of framework validation interviews

Iteration	Interviewee	Date of interview	Function interviewee	Organization	Background interviewee
1	I1.1	02-05-2022	Global data manager	O3	Risk reporting, data management, data governance implementation
	I1.2	19-05-2022	Lead in digital ethics	Forrester research	Governance (data & IT), risk and compliance, PhD in philosophy (digital trust & risk)
2	I2.1	08-06-2022	Data governance consultant	O10	Data (governance) solutions, product owner Purview
	I2.2	08-06-2022	Professor	Open Universiteit	IT artifact design, knowledge management, data privacy and security, intellectual property management, data governance and analytics
3	I3.1	17-06-2022	Enterprise architect	O11	Information management, architecture, workflow management
	I3.2	09-06-2022	Professor	TU Delft	Open data theory, system infrastructure

The validation interviews consisted of three parts. Part A focused on open questions to know more about the background of the interviewee and the organization in case of practitioners. These are summarized below. Part B consisted of a survey in which the interviewees could indicate on a 5-point Likert scale how much they thought the success or situational factor influenced the implementation of the data governance platform. Part C was the validation of the graphical framework and the

interviewees were asked to rate the framework on a 5-point Likert scale based on how much the given evaluation criteria applied (see section 2.2.4). The validation part of the visual framework was only conducted with the practitioners and the researcher who was part of Forrester Research, as they had experience with data governance in a practical environment. The other two academics had experience with data governance and information systems, but not data governance platforms. The full survey results for each participant are shown in Appendix H. In Appendix I, these are visualized in bar charts so it can be seen which factors or criteria scored highest. Moreover, Appendix J gives the comments from the interviewees about why they gave a factor or framework evaluation criteria a certain score, which includes the suggestions for improvement.

6.2 Iteration 1

Interview 1.1

During interview I1.1, the global data governance manager explained a bit more about the Purview implementation at organization O3. The interviewee was responsible for the most important reports of the organization in the financial domain. In order to ensure data privacy, to anonymize personal data, and to help with data retention, Purview was introduced. They chose to implement Purview at this moment in time because of their Cloud migration. The data governance program already started a couple of years ago for laying the foundations. The most challenging part was that Purview was an immature product at the time of implementation, and so the manual of the tool does not address how you use it in certain situations. In order to implement it as intended, it was trial and error. Besides, filling the glossary was easy, but to actually put that in use for hundreds of employees posed difficulties. For many employees, it is just another tool, so a big part was also to convince them of its benefits and to help them get used to the system. Furthermore, the interviewee noted that data governance platforms from different vendors offer more or less the same, and thus the choice mainly depends on how easy the integration is and the costs. At the time of the interview, Purview was not put in use yet, but the glossary was almost implemented, so the interviewee classified the organization between stage 2 and 3 of framework A. As for framework B, the interviewee said that the organization intended to focus on business critical data on a few specific sources to make it a master data and analytics platform.

Interview 1.2

The interviewee had a strong background in digital ethics and had experience working with Forrester research on governance, risk and compliance, and rules and policies in organizations. The interviewee was also working on getting a PhD in philosophy, with a high focus on digital trust & risk. Besides, the interviewee had hands-on experience with building a governance structure (e.g. building policies, processes) but not necessarily with the implementation of related technology. In the interview, it was said that theory on data governance with one ideal scenario often does not take into account the practicalities of data being extremely complicated within an organization. Most companies have many different systems in which data resides, and you cannot apply the same theory to each of those. There is a large variety in data and each department or function uses it in different ways. Normalizing or rationalizing this is challenging. Establishing good data governance also depends on the history of the organization, as most of the time organizations do not build a governance structure from scratch but they are working backwards to piece together what needs to

change. One way to address these problems is identifying where it makes most sense for the business to share data from one function to the next by reviewing the commonalities and mutually beneficial opportunities. There is value in systems if you can define use cases and personas of people who access data and how they access it. Moreover, platforms can bring a common set of truth and give people access to what they need. However, it cannot bring value to every single business situation. Regarding the frameworks; the second framework can be staged in a different way to use the two frameworks in conjunction.

Improvement opportunities

The following opportunities could be identified during the interviews:

- *Suggestion for a success factor:* involve the business with the platform implementation from the beginning of the project (I1.1). This was because Purview was leaning on the technical side with technical terminology, which makes it difficult for business users.
- *Suggestion for situational factors:* sector should score very high, as this decides regulatory requirements (I1.1). This should come back in the visual framework.
- *Suggestion for situational factors:* the organization's history should be taken into account (I1.2).
- *Suggestion for framework A:* adjust the level of detail of the framework (I1.1, I1.2). There should be more steps and shorter steps, with a smaller scope. Then the value after every step also becomes more clear. Some organizations are more commercially driven and have short iterations, so they want small steps.
- *Suggestion for framework A:* add information to the framework (I1.1). Include information about the duration or outcome of every step. What a data governance board immediately wants to know is how long it takes and what value it brings.
- *Suggestion for framework B:* give definitions of what is meant by the dimensions (I1.1). The term "transactional" is used differently between employees within the organization, so you have to specify what it means.
- *Suggestion for framework B:* add an explanation of what is meant to be placed in a quadrant (I1.2). It is not clear whether an organization should try to go to another quadrant.
- *Suggestion for framework B:* include the dimension sector in the framework (I1.1). The choice for a platform depended on the rules and regulations determined by the sector, so this should come back.
- *Suggestion for framework B:* adjust the title of the quadrant (I1.1, I1.2). A title such as "Improvement" might be misleading, because organizations that work with transactional data might also try to improve.

Redesign of framework

Taking these suggestions for improvement into account, some adjustments were made to the frameworks. The redesigned version of framework A is shown in figure 11. For framework A, the main feedback was that more steps were needed with a smaller scope. Instead of four stages, the redesigned framework has seven stages, each with fewer success factors. These steps for example also provided more guidance for when to assign labels to data and when to look into data lineage. When one stage has to be re-executed for a different dataset, the organization can go back to a small step with fewer factors instead of going over all factors in a big stage. Moreover, as feedback was given that there

should be a higher focus on involvement with the business at the beginning, this was taken into account as well for the success factors at the first stage.

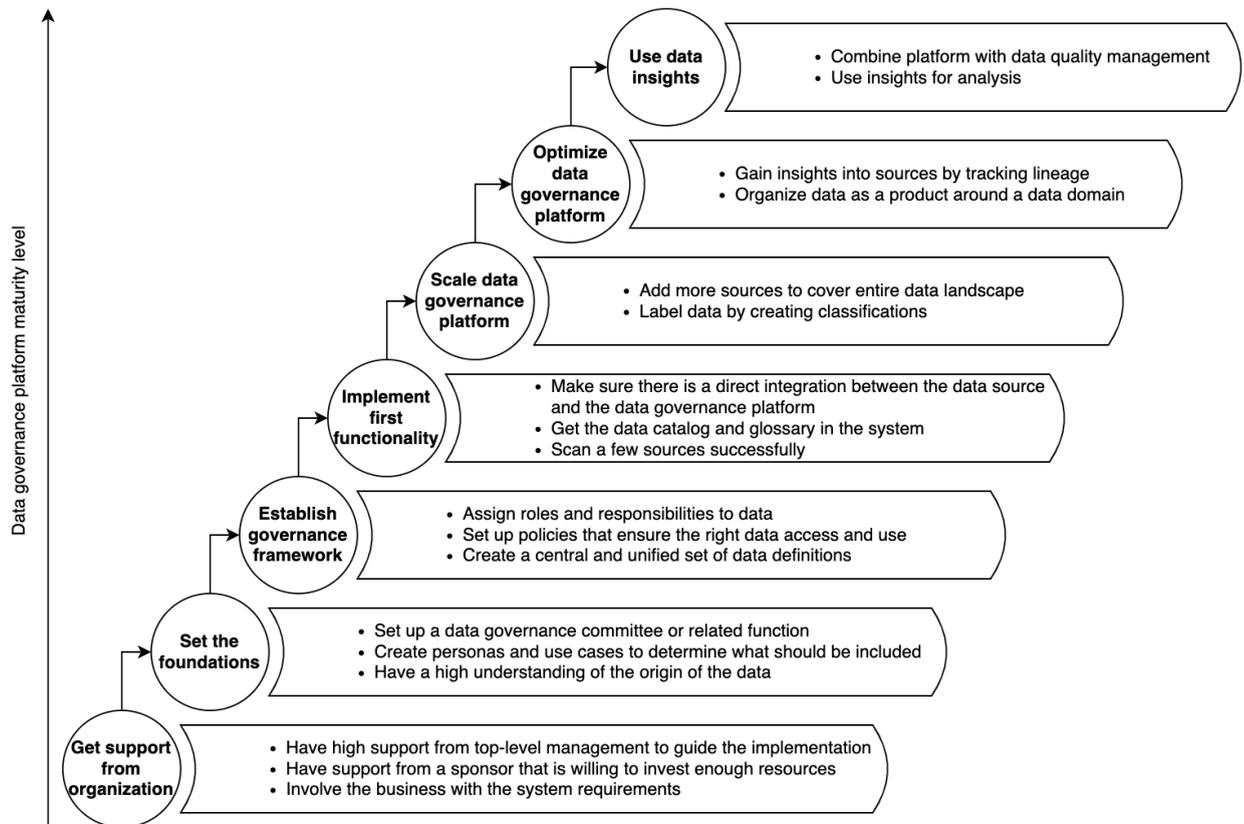


Figure 11. Redesign of framework A during iteration 1

In figure 12, the redesign of framework B can be seen. The main point of feedback for framework B was that the dimensions of number of data sources and type of data were not the most suitable combination. One suggestion for a different dimension was the sector, and this is applied in the redesign by changing one dimension to the *level of regulation in the sector*, ranging from few regulations (e.g. in the energy sector) to many regulations (e.g. in the financial sector). Besides that, the dimension for the *number of data sources* was changed to *complexity of data*. In the interviews it was indicated that it was not necessary about how many sources to include, but more about the content of the sources. For the redesigned framework, it was assumed that few regulations provide more opportunities for value creation (as less rigorous standards can limit innovation; see chapter 5), and that many regulations lead to a high focus on compliance. Combining this with the complexity of data, highly unstructured sources provide more opportunities to discover valuable data as the nature of these data is not yet known. The names of each quadrant were also changed for a more accurate description of its content, according to the feedback.

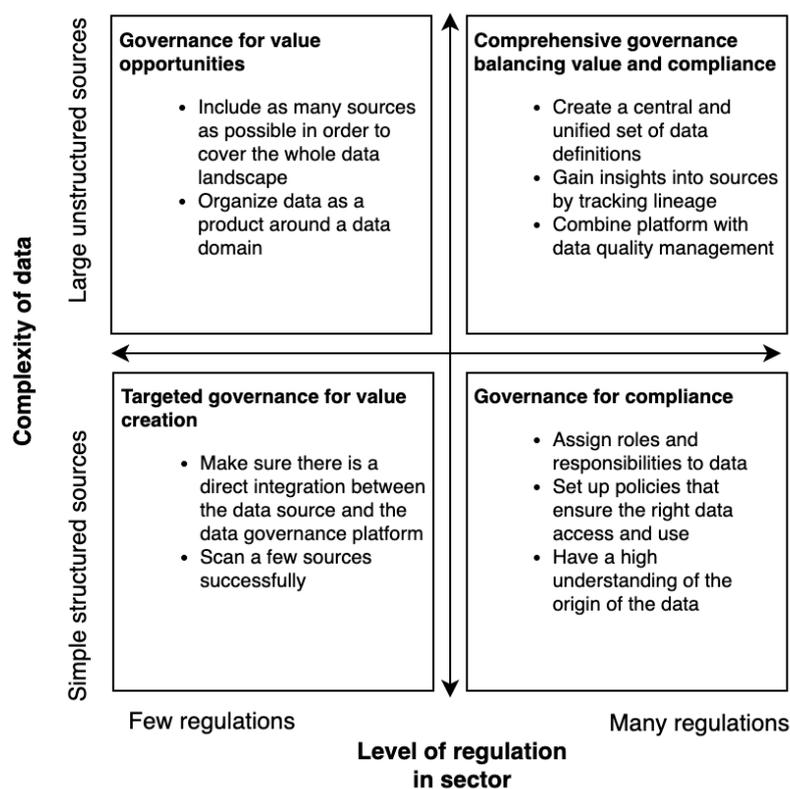


Figure 12. Redesign of framework B during iteration 1

6.3 Iteration 2

Interview 2.1

The interviewee from interview 2.1 worked as freelancer in the area of data (governance) implementations and was tasked to oversee the implementation of Purview at organization O10. The interviewee highlighted that the most important aspect of the implementation at organization O10 was not the tool but the business problem surrounding it, as the request for a tool was a result of a business need. The initiative was part of a bigger program of a data platform and warehouse implementation. Purview was intended to discover and find information easily in the data catalog and make it more accessible. Since Purview was a limited product, functionalities for data quality, business definitions, and more extensive data lineage were not possible to include in the tool. They developed their own tools for these functionalities as a solution. Challenges surrounded the immaturity of the product, as the pricing was not clear, some scans did not work and there were system failures at unexpected moments. From a business perspective, use cases could be carried out well, but adoption posed another challenge as employees rather used their old manual solutions (in MS Teams). At the time of the interview, Purview was running but its use was limited; it was mainly used by the development team and a limited group of information analysts. The next steps for the platform surrounded increasing data quality and using data lineage for impact analysis. The interviewee noted with the current maturity of data governance platforms, not a lot is possible, but in the future it should be Google for data; enter something in the platform and the tool searches for what you want.

Interview 2.2

The interviewee from interview 2.2 was an academic with a PhD and had an academic background in the area of knowledge management, IT artifact design (specifically algorithmic design), data privacy and security, information security management and intellectual property management. Besides, the interviewee had experience with teaching data governance at Master's level, and also oversaw case study work of students who developed data governance instruments at organizations. The interviewee explained that although organizations are aware that data is their biggest asset, it still poses a challenge to look at data like an asset that has a life and needs to be managed. Frameworks such as DAMA are not that well executed in organizations, and data quality is often a problem. Organizations need more guidance, but data governance is a very complex journey. Additionally, the problem with tools is that organizations invest in it, but they do not think enough beforehand how to use it and how benefits are gained. They adjust their way of working to the tool, instead of using the tool as a facilitator for support. These problems can be addressed by thinking carefully about business processes and how data is necessary for it. Data comes in many forms (e.g. structured, half-structured, master data, transaction data), and it needs to be decided how to use the data and how to get the quality data from this for the best value. An organization can start with a business model and mapping processes, and linking it to the data flows. Regarding data governance platforms, it is just another tool that brings together data governance practices. However, it can help to do data governance better if the organization already has a good idea of how to manage data in the first place. Then, it can be considered how to get the best platform that supports the business model. Data governance is dynamic and it keeps evolving, but organizations have to start somewhere. What is important for using models for implementation is to approach it in an agile way and implement in increments.

Improvement opportunities

The improvement opportunities derived from this iteration are as follows:

- *Suggestion for success factors:* make the terminology less Purview based (I2.1). Purview is a source-based system, so terminology highly focusses on things such as sources and connectors, but the business side talks more about things such as entities and assets.
- *Suggestion for success factors:* some factors should not be applied to all data, as the level of governance required depends on what you want to do with the data and how valuable it is (I2.1, I2.2).
- *Suggestion for a success factor:* start with a business problem and work from there (I2.1). Introducing a system does not help if you do not have a business problem that needs to be solved.
- *Suggestion for a success factor:* construct a business model which maps the business processes and links data flow to it (I2.2). A platform should be based on how it best supports the business model, so this needs to be established before implementing it.
- *Suggestion for framework A:* there should be iterations so that an organization can go through the cycle multiple times (I2.1, I2.2). It should be made clear that you can take it up piece by piece instead of implementing it all at once.
- *Suggestion for framework A:* it should be made clear that the steps are not needed for all kinds of data (I2.1). For some data (e.g. data that is not used for making decisions), the steps are not necessary as you do not want to use it for analytical insights.

- *Suggestion for framework A:* in the first stage, the scope of the program needs to be defined, which needs to be related to the business challenge (I2.1). This is important because it determines how the rest of the platform should be implemented, as not all data needs to be treated the same.
- *Suggestion for framework B:* there should be a distinction made between data that is relevant and data that is not relevant in the dimensions (I2.1). Governance is not relevant for some data (e.g. it is not needed for euro signs), so besides important it in the data governance platform it is not important to do something with it.
- *Suggestion for frameworks:* the organization should use it as guidance, and then make their own version from it (I2.1). It is important to look at what the problem is (framework B can be used for this), how you can solve it (framework A can be used for this), and then make a situation specific roadmap.

Redesign of the framework

The main points of feedback from this iteration were that not all data should be treated the same, as it is not necessary for some data to have governance. This was applied in framework A by adding in the first stage that the scope should be determined. The redesigned framework A can be seen in figure 13. After deciding the scope, the framework can be executed in iterations. This is because an organization can start with one data domain, and after that decide to implement the system in more than one data domain, so it should be possible for the organization at every stage to go back to deciding on a scope. However, this can only be done after there is a sponsor and support from leadership, otherwise the scope cannot even be decided.

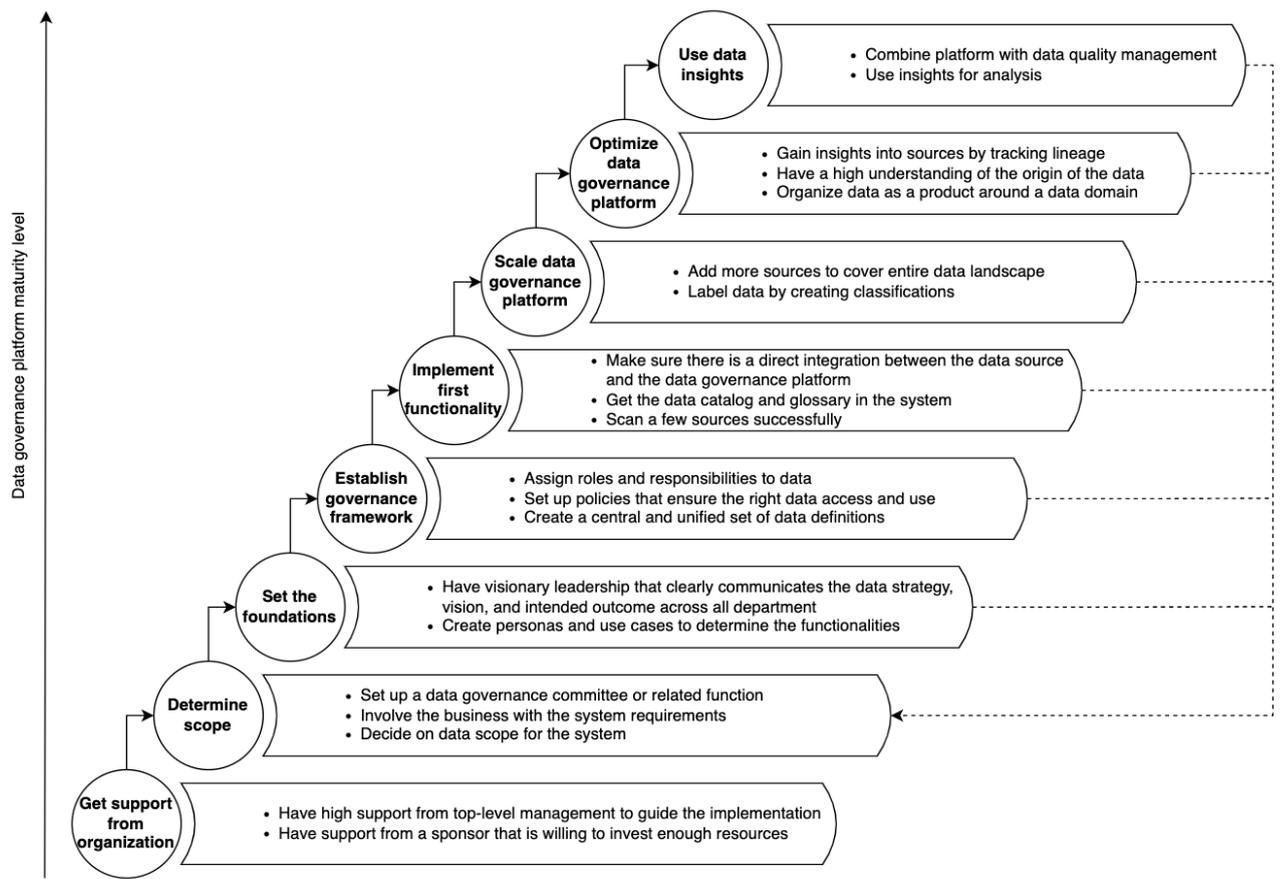


Figure 13. Redesign of framework A during iteration 2

The redesign of framework B is shown in figure 14. Instead of letting the dimension *complexity of data* range from simple structured sources to large unstructured sources, it was decided to let it range from transactional structured data to business-critical unstructured data. It was assumed that structured data could also be discovered by something like a data warehouse, so that would reduce the complexity and the necessity for a platform. Additionally, since transactional data (i.e. all data resulting from business events) is not a necessity to use for business analysis and therefore does not require as much governance, it was assumed that transactional data would also reduce the complexity. On the other hand, business critical data that is unstructured is a good potential for a data governance platform, as the level of automation from a data governance platform is needed to make sense of this data. On top of that, with many regulations, this is data that should be carefully tracked and improved. Framework B can be used in the *determine scope* stage of framework A, to decide on what data to focus.

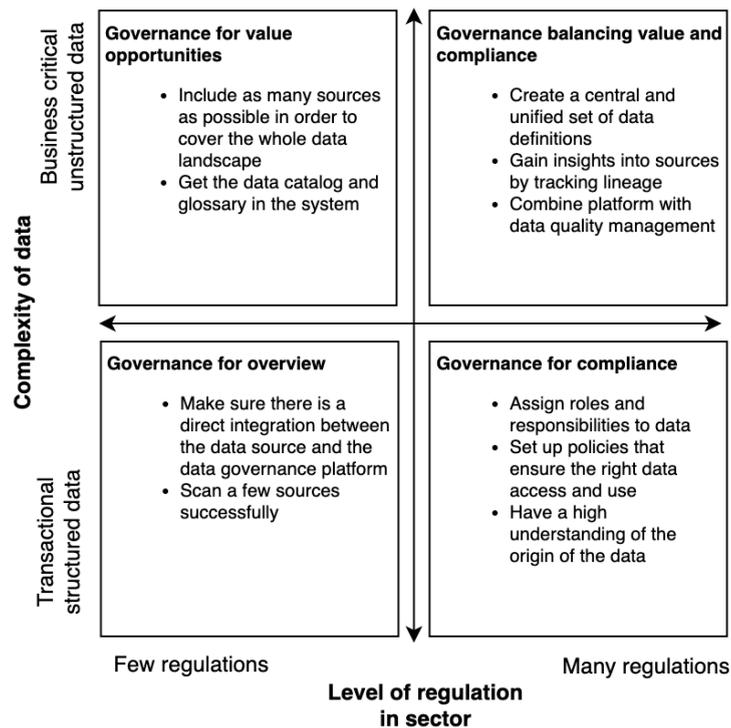


Figure 14. Redesign of framework B during iteration 2

6.4 Iteration 3

Interview 3.1

The interviewee from interview 3.1 worked as an enterprise architect within the financial organization O11. The interviewee had been involved with multiple initiatives at organization O11 that touched upon the area of data governance, which included projects focused on determining the data maturity and establishing a central data catalog with definitions at a high enterprise level (regardless of the sources). Not only was this about the technical side of systems, but also about areas such as culture, adoption and use. Organization O11 did not yet implement a data governance platform, but they were investigating the possibilities. They needed to control and validate their data quality in order to ensure compliance with the banking regulations. They also needed this for their reports for control, management and commerce. At the time of the interview, they had a central platform to which a data warehouse was connected and they had tooling for master data management to manage data quality. They also established central data definitions for all the business lines (although not linked to actual data), and they assigned data stewards and data ownership. Moreover, they defined a data architecture for every business line. What they missed was a data catalog to relate their data definitions to the data. This needed to reunite the data of the different business lines. They were looking into the IBM governance suite, as they already had affinity with IBM products and thus this product would be easiest to integrate.

Interview 3.2

The interviewee from interview 3.2 was a PhD candidate and a Professor at University, with a research background which focused on open data. This was from a socio-technical perspective, which

meant not only looking at the techniques, platforms, and infrastructure for enabling open data, but also at how this has to be institutionalized and making sure that people are willing to share data. The interviewee gave a course at the University on governance specialization in the area of information and communication. Regarding the practical experience with data governance, the interviewee carried out case studies at a governmental institution surrounding data governance. The interviewee also had experience with tools in the area of data governance, more specifically tools about how to follow meta-data standards and tools for data visualization and analysis. The main problems with data governance that the interviewee researched during the case study were related to the lack of central overview and standards of data. Many people in the organization requested access to data that was already made accessible by other projects, but this was not known. Additionally, data was saved in a form that was usable for one specific project, but not for others. In order to solve this, it is important to make an inventory of what datasets there are. Moreover, meta-data standards can be introduced within a domain (although these standards often differ per discipline or domain within an organization). Regarding the role of systems, there cannot be good data governance without tools. They are needed to create an overview of data and to standardize data. However, this can already be something simple like Excel in which everyone reports their data. On the other hand, it can be a more complex tool which specializes in processing data standards.

Improvement opportunities

From the interviews during iteration three, the following improvement opportunities were identified:

- *Suggestion for success factors*: you do not necessarily need central guidance as data governance can also work on a decentral level (I3.1, I3.2). When an organization is very decentral, it conflicts with things such as one central data strategy or one central data committee.
- *Suggestion for success factor*: it is not about standardizing or creating unified definitions across domains, but within domains (I3.2). Different domains handle different definitions of data, so within an organization data should be standardized for each domain and not for all data.
- *Suggestion for success factors*: it is not about covering all data, but about whether it is usable (I3.1, I3.2). For some data it is difficult to apply data governance.
- *Suggestions for framework A*: explain the terms used in the framework or make them more self-explanatory (I3.1). Some terminology was hard to understand when not having a specific platform background.
- *Suggestion for framework A*: it should be used complementary to another model that determines the maturity of the organization (I3.1). Which areas have to be improved (e.g. architecture, culture) influence the rest of the growth model.
- *Suggestion for framework A*: actions should be given that specify how you should go from one stage to another stage (I3.1).
- *Suggestion for framework B*: the data complexity dimension should be focused either on transactional vs business critical or structured vs unstructured, but not both (I3.1). Transactional data can also be unstructured data, so this makes the dimension difficult to understand.
- *Suggestion for framework B*: the *unstructured* classification should not consider all unstructured data like emails, social media, and videos, as it is difficult to set up governance

for this (I3.1). The success factor about covering the whole data landscape for this classification should therefore be adjusted or explained differently.

- *Suggestion for framework B*: add an explanation for its use, examples, and a reason for why this framework should be considered, as the framework is not self-explanatory (I3.1).

Redesign of the framework

As this was the last iteration of this research, the framework was not again redesigned. The points of improvement from this iteration can be taken further into account for a possible future study (see also section 7.3 about future research).

6.5 Conclusion

The framework designed in chapter 5 was validated by means of interviews with practitioners at case study organizations and academics who had a relation to the field of data governance. Both the framework in narrative and in visual form were rated on a 5-point Likert scale. This was done in order to check whether the success factors were grouped and classified accordingly in the framework. During the first iteration, the main point of feedback included that there should be more steps with a smaller scope for framework A, and that dimensions should be picked for framework B that involve the sector and complexity of data sources. For the second iteration, the main feedback was that the steps and guidance from the success factors should not be about all data, as some data is irrelevant and does not require rigorous data governance. For iteration three, suggestions for improvement surrounded the factors regarding centrality, as data can also be organized in a decentral way within a domain, so there is no necessity for central steering. Moreover, the frameworks can be improved with additional explanations of its use, including examples, and using it in combination with a maturity assessment. Coming back to the main research question “*What framework can be designed that supports an organization in the implementation of a data governance platform by considering the situational factors of the organization?*”, both the narrative and visual frameworks together with their points of improvement can be used by an organization to help them determine which factors to include in the data governance platform implementation.

7. Discussion

7.1 Implications

This research addresses the implementation of data governance platforms, looking at it from both a theoretical and practical side. As Alhassan et al. (2016) state, there is limited evidence of the implementation part of data governance (platforms) in literature, which leads to a bridge between theory and practice. In this research, the aim was to combine the two.

Firstly, this study adds to the current theoretical foundations of the topic of data governance platforms. As lots of developments are being made in the topic of data governance platforms, current literature on this topic is scarce. Therefore, this study has focused on building theory that could be applied in the context of data governance platforms. Besides that, it draws on principles from contingency theory, as studies that determine what is important for data governance frequently neglect the dynamic environment in which organizations are situated. Additionally, this research contributes to practice as it provides guidelines for organizations for how and when to apply a data governance platform. As there are many organizations struggling with realizing the benefits that such a platform can provide, it is useful to have more guidance for what kind of situations a data governance platform is suitable for.

It should be noted that these are factors specific to the implementation of data governance platforms and not data governance on its own, and that this research contributes to the body of knowledge on platforms. The success factors which are leaning more towards the technical side (e.g. about classifying data sources) are not as relevant for data governance regardless of platforms. Moreover, the interviewees were asked to rate factors according to how important they were for the implementation of data governance platforms, so the ranking might have been different if focussing on data governance on its own.

7.2 Limitations

Some limitations in this study were identified and the most important ones are discussed in more detail in this section. The full list of limitations can be found in chapter 2.3, where the threats to the validity of the research are discussed and their ways to counter them are described.

A first important limitation about the study is the limited sample size which is a threat for the conclusion validity. As data governance programs vary greatly depending on the organization, in order to have an idea about the commonalities between the programs, it is necessary to include as many situations as possible. Answers from eleven participants were gathered for the first round of interviews, and six interviews took place for the second validation part. However, not every interviewee wanted to answer every question or go into detail about a certain organization during the interviews for privacy reasons due to the companies they worked for. This decreased the sample size for some answers even more. Moreover, it was difficult to find academics who did research on data governance platforms, as it is such a new topic. To counter this, one researcher with a data governance background from Forrester research (which is essentially a consultancy firm) and two academics with a data governance and information system background were reached (but not a data governance platform background). However, the academics without a data governance platform

could not be shown the visual framework for the data governance platform implementation, as this went into too much domain knowledge about the topic. Regarding the survey that was filled in for the validation part where participants could rank the factors and criteria on a 5-point Likert scale, there were too few participants in order to perform statistical analysis on it. Therefore, the study mainly focused on the qualitative analysis and the quantitative part was only used to rate the importance of the success and situational factors.

Another limitation was the possible bias for a specific data governance platform, namely Purview, which is a risk for the external validity of the research. SMEs were reached from contacts with the organization Avanade, which is a Microsoft oriented company which provides consultancy services for Purview specifically. Regarding the data governance tooling, Microsoft chose their own framework so it enforces a specific way to look at a data governance implementation. Therefore, some of the terminology used for data governance platforms in this research might be closely related to the terminology used in Purview and not necessarily other data governance platforms. This was confirmed by one of the practitioners during the validation interviews, who gave a comment that the terminology was a bit focused on how Purview was set up. This risk was countered by asking during every interview how Purview compared to other products. Moreover, there were three cases in which the main focus was not on Purview, but on Collibra, Informatica, and IBM. Therefore, perspectives from other products were also covered. Additionally, all participants in the validation interview did not have a specific affinity with Microsoft Purview, as they were not part of Avanade. Furthermore, triangulation of methods were used, and literature, case study work, and expert interviews with researchers were also used to form a conclusion.

Moreover, a risk for the construct validity was that during the first round of interviews, the categories were rated subjectively. This might have caused researcher bias in the ranking of the success factors. However, the interviewees were not asked to rate a pre-prepared list of factors because the intention of the interviews was to elicit which success factors there were. Giving the interviewees a predefined list of factors could have prevented some factors from being discovered. Therefore, the study was approached from a more qualitative viewpoint instead of quantitative analysis. To minimize this threat, during the validation rounds the interviewees were asked to rate the identified factors on a 5-point Likert scale.

Next, a risk is that the framework is too high-level which affects the practical applicability as it is not scoped to a specific sector. However, the purpose of the framework is not only to guide an implementation, but also to facilitate discussions about data governance and reach a common understanding within the organizations (Begg & Caira, 2012). A data governance solution can also be used for bringing together groups from different departments within an organization to communicate about the use of the data assets to increase data awareness (DeStefano et al., 2016). Since little is known about data governance platforms in practice as it is such a new area, it was chosen to focus on the more general factors so that first a common understanding could be created about data governance platform implementation. Moreover, organizations can create their own version of a framework by adapting a more general framework from other standards (Begg & Caira, 2012). By using the framework as a tool for communication and using it as a guide in combination with other best practices and standards, this limitation can be addressed. For example, it can be used in combination with maturity analysis to determine the level of maturity in the organization (Begg & Caira, 2012), or with an assessment of the organizational culture to determine what implementation strategy is most suitable (Cameron & Quinn, 2006).

Another point of attention is the difference between findings from theory and findings from the expert interviews with practitioners. The practical results were compared and validated with theory at the points in the research; the success factors that resulted from the expert interviews were confirmed by literature, and the preliminary framework results were validated with academics. This way, triangulation of the research was applied, which increased the external validity. One noticeable distinction between the outcomes of the literature review and the expert interviews was that in literature, there was a high emphasis on organizational culture to be the most important factor for the success of an implementation effort, while this was not the main emphasis in the expert interviews with the SMEs. During the interviews it was asked whether the culture and overcoming resistance posed a challenge during the implementation of the system, but multiple interviewees indicated that they had no trouble with this. An explanation for this might be that the subject matter experts (SMEs) mainly focused on the technical implementation of the system and did not get to experience the human resource perspective besides it. Another explanation might be that problems with culture and resistance mainly come after the implementation, during the adoption part which focuses on the actual use of the system (Van Linge, 2006). The interviewees focused on the implementation part of the data governance system, and did not go into detail about its use after the implementation, mainly because they did not get to experience that as the system was recently implemented and was put in use for a very short time. Another difference between practitioners and academics could be seen in the validation interviews. The main level of expertise and a big focus of the interviews for the academics included open data and the sharing of data. However, the focus of the interviews with the practitioners was more on the practicality of the framework, specifically with regards to the data landscape within the organization. More points of differences can be seen in Appendix I, where bar charts are given that show the difference in scores given by practitioners and academics to the success and situational factors.

7.3 Future research

By considering the limitations from the previous section, some suggestions can be made about topics for future research.

This study focused on data governance platform implementation, and not on data governance itself. Although data governance platforms can help improve data governance, they are not a solution (Petzold et al., 2020). Ladley (2012) mentions that a tool should only be used to improve something that is already being done in an organization. For future work, more research can be done about what an organization needs to have established before they start thinking about using a tool. Moreover, it can be examined to what extent an organization can establish good data governance without the use of a tool. This was confirmed during the validation interviews, in which it was said that some sort of maturity assessment should be made before using the frameworks, in order to check whether the organization is ready enough to actually use the frameworks.

As this study was mainly focused on qualitative analysis and eliciting success and situational factors, there are opportunities for quantitative research. A survey could be constructed that could be forwarded on a much larger scale to data governance professionals. This way, some kind of statistical analysis can be made about which factors score higher than others. As such, a more sound ranking can be made about the priority of success factors and situational factors.

Another point is that the visual framework could be extended to more dimensions that might affect the implementation of a data governance program. In the validation interviews it was frequently indicated that the scope or domain of the data governance program is very important for determining which parts of the platform to focus on first. Thus, it would be useful to combine the current visual framework with another framework that helps an organization to decide on the scope of the data governance program. Moreover, from the validation it became clear that the sector might be an important factor to consider as highly regulated industries require a different approach. Consequently, there is opportunity to adopt the framework to different sectors.

Since there was a high focus in this study on Purview, additional research could look into how Purview compares to other products, and make a recommendation for which product is required in what kind of situation. This way, some form of advice can be given to an organization which would help them in a selection of a product. It would also contribute to the current body of knowledge on data governance platforms by determining which functionalities are still lacking in some of the products, and how they can be improved.

8. Conclusion

The problem of this research was the complexity of the implementation of data governance platforms. New solutions for data governance are needed, due to the fastly increasing data volumes organizations have to manage and the pressure from new regulations. Additionally, there is no one-size-fits-all solution, and every organization needs a different approach for implementing a data governance (platform) program. If the organization does not have the capabilities to properly execute data governance activities, then the maximum benefit of a tool cannot be reached. Resultantly, the objective of this research was to design a data governance framework that could support an organization in implementing a data governance platform, by taking into account which factors would affect such an implementation. This led to the following main research question:

MRQ: "What framework can be designed that supports an organization in the implementation of a data governance platform by considering the situational factors of the organization?"

This question was answered by performing a multivocal literature review (MLR), conducting expert interviews with subject matter experts (SME's), designing the framework, and validating this in a case study and by means of expert interviews with researchers. In order to answer the main research question, sub-questions were established and in this chapter the conclusions for each of the sub-questions are drawn.

RQ1: What does the implementation of data governance mean for an organization?

The outcomes of the MLR showed that data governance has eight components that should all be addressed in a data governance program: cross-functional collaboration; devising a framework; managing data as a strategic enterprise asset; assigning decision rights and accountabilities; setting up data policies, standards, and procedures; and monitoring compliance. It is about decisions made about data, ensuring the data's right use, and how value can best be gained from this data. Data governance activities can be automated by a system for improving efficiency and consistency, but it is important that the organization already has the data governance foundations in place before a data governance platform is introduced.

RQ2: What current frameworks that can be used for the implementation of data governance platforms exist?

In most frameworks that are designed for data governance, data roles and responsibilities are an important element. Besides that, the organizational structure (e.g. the presence of a data governance board) takes up a large part of the focus in data governance frameworks. Other frameworks highlight the importance of data policies and standards, which comprise rules for access management and SLA's. There are few frameworks which take into account the presence of data governance systems, so frameworks for implementation of systems are also considered. For this, frameworks from the contingency approach can be used. This theory describes how an implementation depends on the situation of the organization and it is influenced by the organization's environment. Therefore, situational factors should be taken into account when implementing a data governance platform. Examples of frameworks from the contingency approach focus on classifying an organization based on some dimension (internal versus external focus and stability versus control) and suggest a set of characteristics that best fit this organization.

RQ3: What critical success factors in the organization are needed for the successful implementation of a data governance platform?

The success factors from literature were combined with success factors derived from SME interviews. From literature, it became clear that a substantial part of a data governance program is affected by the communication from leadership, and that change management is important for handling the change. During the interviews, there was a higher focus on how data governance platforms were implemented. It became clear that in every implementation project, the integration with the current systems was an important determinant for deciding which data governance platform to implement. Moreover, the data catalog and glossary was a starting point for ensuring the right use of a data governance platform. For this, there needed to be some kind of responsibilities for data. Every interviewee also stressed that a data governance platform is a work in progress, and that after its implementation it still has to be put in actual use.

RQ4: What situational aspects of the organization have the biggest influence on how a data governance platform is implemented?

Similarly to the success factors, the situational factors were derived from literature and SME interviews. The data governance maturity level had a big influence on whether the organization was ready to implement such a platform. Moreover, the sector (e.g. financial, energy) and the goal (e.g. increasing data quality, ensuring compliance) of the program were important determinants. The rest of the factors were more focused on the systems and data in the organization: the systems infrastructure (e.g. Cloud or on-premise systems), type of data (e.g. sensitive or business critical). Besides, the complexity and number of data sources in the organization had some influence on the need of the platform. Culture also influenced this to some extent.

RQ5: What framework can be developed that can support an organization in implementing a data governance platform?

A framework can be designed both in narrative form (i.e. textual description) and graphical representation. For the narrative form, the list of success factors and situational factors were prioritized based on their importance for the implementation of a data governance platform. Success factors that were considered highly important were the integration between data source and data governance system, and classifying a few sources first before scaling to the rest of the system. A situational factor that was considered most important was the data governance maturity level and this was taken into account for designing the visual framework. The visual representation consisted of two complementary frameworks, with framework A focussing on the process of the implementation and framework B focussing on picking an implementation strategy of focus. For framework A, groups of success factors were made based on at what stage of the implementation of a data governance platform they should be applied. For framework B, the success factors were organized along two dimensions related to the data environment of the organization. The frameworks were intended to support an organization into choosing which factors to focus on.

RQ6: How can the data governance platform implementation framework be validated?

In this research, the framework was tested with practitioners in a case study and researchers in expert interviews. During the validation of the framework, it became clear that the framework could mainly be used as a tool for communication. In order for the framework to be applied in practice, it

would be beneficial to include more detailed steps. Moreover, the framework should consider that the steps and guidance from the success factors are not applicable to all data, as some data is irrelevant from which no value can be gained and does not require rigorous data governance. Additionally, it would be beneficial to provide more explanation about the intention of the framework, give the reason for why it is needed, provide a description of platform specific terminology, and provide some examples about its use.

Taking the answers to each of these sub-questions together, it can be concluded that data governance platforms are becoming of increasing importance, as organizations can no longer handle data governance activities manually due to the growing volumes of data. To realize data governance at a big scale, platforms should be considered. However, when a data governance platform should be implemented and what functionalities should be included depend on the organization at hand. Before any kind of automation, it is important that the organization has the foundations of data governance in place, which include a plan for roles and responsibilities, policies for data use and access, and an idea of the data landscape. A data governance platform can increase the efficiency at which data governance is executed and it can ensure consistency of data. A framework that focuses on the data governance platform maturity level can be used as guidance for the implementation process of such a platform, and this should be used in combination with a framework that helps the organization in deciding the level of governance sophistication needed for a data set. However, at every step of the way, it should be noted that data governance is more than tools. Or, as frequently mentioned during the interviews: *“Technology is helpful, but without the others (people, processes) the strategy could fail.”*

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.
- Adams, W. C. (2015). Conducting semi-structured interviews. *Handbook of practical program evaluation*, 4, 492-505.
- Al-Badi, A., Tarhini, A., & Khan, A. I. (2018). Exploring big data governance frameworks. *Procedia computer science*, 141, 271-277.
- Alhassan, I., Sammon, D., & Daly, M. (2016). Data governance activities: an analysis of the literature. *Journal of Decision Systems*, 25(1), 64-75.
- Alhassan, I., Sammon, D., & Daly, M. (2018). Data governance activities: A comparison between scientific and practice-oriented literature. *Journal of Enterprise Information Management*.
- AlManej, M., Salonitis, K., & Tsinopoulos, C. (2018). A conceptual lean implementation framework based on change management theory. *Procedia cirp*, 72, 1160-1165.
- Al-Ruithe, M., Benkhelifa, E., & Hameed, K. (2016). A conceptual framework for designing data governance for cloud computing. *Procedia Computer Science*, 94, 160-167.
- Al-Ruithe, M., Benkhelifa, E., & Hameed, K. (2019). A systematic literature review of data governance and cloud data governance. *Personal and Ubiquitous Computing*, 23(5), 839-859.
- Ampatzoglou, A., Bibi, S., Avgeriou, P., Verbeek, M., & Chatzigeorgiou, A. (2019). Identifying, categorizing and mitigating threats to validity in software engineering secondary studies. *Information and Software Technology*, 106, 201-230.
- Arvidsson, V., Holmström, J., & Lyytinen, K. (2014). Information systems use as strategy practice: A multi-dimensional view of strategic information system implementation and use. *The Journal of Strategic Information Systems*, 23(1), 45-61.
- Atkins, L., Francis, J., Islam, R., O'Connor, D., Patey, A., Ivers, N., ... & Michie, S. (2017). A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. *Implementation science*, 12(1), 1-18.
- Avanade (2022). *Our story*. Retrieved March 20, 2022, from <https://www.avanade.com/nl-nl/about-avanade/story>
- Begg, C., & Caira, T. (2012). Exploring the SME Quandary: Data Governance in Practise in the Small to Medium-Sized Enterprise Sector. *Electronic Journal of Information Systems Evaluation*, 15(1), pp3-13.

- Birken, S. A., Rohweder, C. L., Powell, B. J., Shea, C. M., Scott, J., Leeman, J., ... & Presseau, J. (2018). T-CaST: an implementation theory comparison and selection tool. *Implementation Science*, *13*(1), 1-10.
- Brous, P., Janssen, M., & Vilminko-Heikkinen, R. (2016). Coordinating decision-making in data management activities: a systematic review of data governance principles. In *International Conference on Electronic Government*, 115-125. Springer, Cham.
- Cameron, K. S. & Quinn, R. E. (2006). *Diagnosing and changing organizational culture: Based on the competing values framework*. Addison-Wesley.
- Carter, N., Bryant-Lukosius, D., DiCenso, A., Blythe, J., & Neville, A. J. (2014). The use of triangulation in qualitative research. In *Oncol Nurs Forum*, *41*(5), 545.
- Cecchino, N. J. (2010). Google scholar. *Journal of the Medical Library Association: JMLA*, *98*(4), 320.
- Cheong, L. K., & Chang, V. (2007). The need for data governance: a case study. *ACIS 2007 Proceedings*, 100.
- Ćirić, Z., & Raković, L. (2010). Change Management in information system development and implementation projects. *Management Information System*, *5*(1), 23-28.
- Cook, B. G., & Odom, S. L. (2013). Evidence-based practices and implementation science in special education. *Exceptional children*, *79*(2), 135-144.
- Corbin, J., & Strauss, A. (2008). Theoretical sampling. *Basics of qualitative research*, 143-158.
- Dai, W., Wardlaw, I., Cui, Y., Mehdi, K., Li, Y., & Long, J. (2016). Data profiling technology of data governance regarding big data: review and rethinking. *Information Technology: New Generations*, 439-450.
- Davis, R., Campbell, R., Hildon, Z., Hobbs, L., & Michie, S. (2015). Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. *Health psychology review*, *9*(3), 323-344.
- DeStefano, R. J., Tao, L., & Gai, K. (2016). Improving data governance in large organizations through ontology and linked data. In *2016 IEEE 3rd International Conference on Cyber Security and Cloud Computing (CSCloud)*, 279-284.
- Figuroa-Flores, J. R., Acosta-Gonzaga, E., & Ruiz-Ledesma, E. F. (2020). Causes of failure in the implementation and functioning of information systems in organizations. *International Journal of Advanced Computer Science and Applications*, *11*(6), 137-142.

- Garousi, V., Felderer, M., & Mäntylä, M. V. (2019). Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. *Information and Software Technology, 106*, 101-121.
- Hafiz, K. (2008). Case study example. *The qualitative report, 13*(4), 544-559.
- Henriette, E., Feki, M., & Boughzala, I. (2016). Digital Transformation Challenges. *In MCIS, 33*.
- Hong, S., & Kim, J. (2004). Architectural criteria for website evaluation—conceptual framework and empirical validation. *Behaviour & Information Technology, 23*(5), 337-357.
- Hove, S. E., & Anda, B. (2005, September). Experiences from conducting semi-structured interviews in empirical software engineering research. *In 11th IEEE International Software Metrics Symposium*.
- Hussain, S. T., Lei, S., Akram, T., Haider, M. J., Hussain, S. H., & Ali, M. (2018). Kurt Lewin's change model: A critical review of the role of leadership and employee involvement in organizational change. *Journal of Innovation & Knowledge, 3*(3), 123-127.
- Imran, M. K., Rehman, C. A., Aslam, U., & Bilal, A. R. (2016). What's organization knowledge management strategy for successful change implementation? *Journal of Organizational Change Management, 19*(7).
- Jami Pour, M., & Hosseinzadeh, M. (2021). An integrated framework of change management for social CRM implementation. *Information Systems and e-Business Management, 19*(1), 43-75.
- Jones, R. A., Jimmieson, N. L., & Griffiths, A. (2005). The impact of organizational culture and reshaping capabilities on change implementation success: The mediating role of readiness for change. *Journal of management studies, 42*(2), 361-386.
- Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). *Likert scale: Explored and explained. British journal of applied science & technology, 7*(4), 396.
- Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM, 53*(1), 148-152.
- Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University, 33*(2004), 1-26.
- Wróbel, A., Komnata, K., & Rudek, K. (2017). IBM data governance solutions. *In 2017 International Conference on Behavioral, Economic, Socio-cultural Computing (BESOC), 1-3*.
- Ladley, J. (2012). *Data governance: How to design, deploy, and sustain an effective data governance program*. Academic Press.

Longhurst, R. (2003). Semi-structured interviews and focus groups. *Key methods in geography*, 3(2), 143-156.

Mahanti, R. (2018). Data Governance Implementation: Critical Success Factors. *Software Quality Professional*, 20(4).

Microsoft (2022). *Azure Purview*. Retrieved March 20, 2022, from <https://azure.microsoft.com/en-us/services/purview/#overview>

Moullin, J. C., Dickson, K. S., Stadnick, N. A., Albers, B., Nilsen, P., Broder-Fingert, S., ... & Aarons, G. A. (2020). Ten recommendations for using implementation frameworks in research and practice.

Naude, F. (2017). Comparing downloads, Mendeley readership and Google Scholar citations as indicators of article performance. *The Electronic Journal of Information Systems in Developing Countries*, 78(1), 1-25.

Otto, B. (2011). A morphology of the organization of data governance. *ECIS 2011 Proceedings*, 272.

Panian, Z. (2010). Some practical experiences in data governance. *World Academy of Science, Engineering and Technology*, 62(1), 939-946.

Peper, J., van Gunsteren, M., van Linge, R., & Robben, P. (2014). Implementeren als kenniscreatie. In *Management & Organisatie*, 1.

Petzold, B., Roggendorf, M., Rowshankish, K., & Sporleder, C. (2020). *Designing data governance that delivers value* [Whitepaper]. McKinsey.

Peyret, H., & Goetz, M. (2014). *The forrester wave™: Data governance tools, Q2 2014* [Whitepaper]. Cambridge, MA 02140: Forrester Research, Inc.

Reinsel, D., Gantz, J., & Rydning, J. (2018). *The digitization of the world: from edge to core* [Whitepaper]. IDC.

Ronnenberg, S. K., Graham, M. E., & Mahmoodi, F. (2011). The important role of change management in environmental management system implementation. *International Journal of Operations & Production Management*, 31(6).

Russom, P. (2008). Q&A: Data Governance Strategies. *Business Intelligence Journal*, 13(2), 13.

Sandelowski, M. (2000). Combining qualitative and quantitative sampling, data collection, and analysis techniques in mixed-method studies. *Research in nursing & health*, 23(3), 246-255.

Schneider, D., Huth, T., & Vietor, T. (2021). Development of an Industry 4.0 method and knowledge platform for strategic technology implementation. *Procedia CIRP*, 100, 613-618.

Sonenshein, S., & Dholakia, U. (2012). Explaining employee engagement with strategic change implementation: A meaning-making approach. *Organization Science*, 23(1), 1-23.

Strifler, L., Barnsley, J. M., Hillmer, M., & Straus, S. E. (2020). Identifying and selecting implementation theories, models and frameworks: a qualitative study to inform the development of a decision support tool. *BMC medical informatics and decision making*, 20(1), 1-12.

Tallon, P. P. (2013). Corporate governance of big data: Perspectives on value, risk, and cost. *Computer*, 46(6), 32-38.

Tenkasi, R. V., & Chesmore, M. C. (2003). Social networks and planned organizational change: The impact of strong network ties on effective change implementation and use. *The Journal of Applied Behavioral Science*, 39(3), 281-300

The Data Governance Institute. (2022). *Definitions of Data Governance* [Online]. Retrieved from <https://datagovernance.com/the-data-governance-basics/definitions-of-data-governance/>

Van de Ven, A. H., Ganco, M., & Hinings, C. R. (2013). Returning to the frontier of contingency theory of organizational and institutional designs. *Academy of Management Annals*, 7(1), 393-440.

Van Linge, R. (2006). *Innoveren in de gezondheidszorg: theorie, praktijk en onderzoek*. Reed business.

Verschuren, P., Doorewaard, H., & Mellion, M. (2010). *Designing a research project*, 2. The Hague: Eleven International Publishing.

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 (JDIQ)*, 1(1), 1-27.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly*, xiii-xxiii.

Wende, K., & Otto, B. (2007). A contingency approach to data governance. In *Proceedings of 12th International Conference on Information Quality*, 163-176.

Wieringa, R. J. (2014). *Design science methodology for information systems and software engineering*. Springer.

Wieringa, R., & Morali, A. (2012, May). Technical action research as a validation method in information systems design science. In *International Conference on Design Science Research in Information Systems* (pp. 220-238). Springer, Berlin, Heidelberg

Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th international conference on evaluation and assessment in software engineering* (pp. 1-10).

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). *Experimentation in software engineering*. Springer Science & Business Media.

Wynn, D. C., Eckert, C. M., & Clarkson, P. J. (2007). Modeling iteration in engineering design. In *DS 42: Proceedings of ICED 2007, the 16th International Conference on Engineering Design*, 693-694.

Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research*, 39(1), 93-112.

Appendices

Appendix A: Literature review

In figure 15, an example is given on how information is gathered for the literature review.

Title	Author(s)	Year	Topic	Type	Publisher	Number of citations	Resource	Link	Keywords	Snowballing	Relevance	Reason for inclusion/exclusion	Summary
Some Practical	Panian, Z.	2010	Data governance framework	Article	-	122	Google scholar	https://cit	Data governance (framework OR model)	-	2	Mostly reasons, concepts, corporate factors, very high-level	Gives reasons for why data key attributes (a security) and the built policies and processes and technological su
Data governi	Abraham, R., Sc	2019	Data governance framework	Article	-	120	Google scholar	https://d1	Data governance (framework OR model)	-	1	Gives an overview of all factors affecting data governance	Designs a conceptual governance mechanism scope, antecedents & possibilities for feature
Data governi	Alhassan, I., Sar	2016	Data governance definitions	Article	-	96	Google scholar	Data.gov	Data governance (framework OR model)	-	3	Only concepts as used within other literature	Performs a literature overview of the data data can be used as
A Conceptu	Al-Ruithe, M., Be	2016	Cloud data governance	Article	-	43	Google scholar	A Concept	Data governance (framework OR model)	-	3	Focussed mostly on cloud, very high-level	Gives an explanation cloud. Mentions the IT governance. Also
A systemati	Al-Ruithe, M., Be	2019	Data governance framework	Article	-	71	Google scholar	https://wv	Data governance (framework OR model)	-	1	Comprehensive overview of relevant frameworks	Gives an overview of factors that can ensu this to the cloud spec
Designing c	Khatri, V., & Brov	2010	Data governance framework	Article	-	619	Google scholar	https://dl	Data governance (framework OR model)	-	2	Good starting point, very high level	Provides a framework which can be used a of 5 domains: data pi lifecycle. These are li locus of accountabilit
The need fo	Cheong, L. K., &	2007	Data governance framework	Conferen ce paper	-	124	Google scholar	The Need	Data governance (framework OR model)	-	3	Highly focusses on data quality, proposes a solution only for data governance roles	States that in order fo a need for a data gov for financial reporting data quality. Relates lack of roles and res; study, and suggests ; governance program
Exploring b	Al-Badi, A., Tarhi	2018	Data governance framework	Conferen ce paper	-	54	Mendeley	https://wv	Data governance (framework OR model)	-	3	Focusses mostly on Big data	Relates data govern: governance framework and data governance

Figure 15. Preview of the Excel sheet used for the SLR and MLR.

In table 15, the data extraction form is given which gives a justification of each column from the spreadsheet.

Table 15. Data extraction form

Column	Description	Form
Title	The title of the selected source	Text
Author	The author(s) of the selected source	Text
Year	The year the source was published	Number
Topic	The main topic of the source	Text options: "data governance (framework OR definitions OR cloud OR guidelines OR tools)", "change implementation (framework OR definitions OR success factors)"
Type	The format in which the source was published	Text options: "article", "conference" "paper", "book", "whitepaper"
Publisher	In case of books or whitepapers; which organization published the source	Text
Number of citations	The number of times the source is cited; in case of sources found by the MLR, a "MLR" was noted	Number / "MLR"

Resource	The web engine that was used when searching for the source	Text options: "Google Scholar", "Mendeley", "Google"
Link	URL link to where the source was found	Text
Keywords	The keywords that were used when searching for the source	Text options: see review protocol in section 2.2.1
Snowballing	In case forward or backward snowballing was used, it was noted from which source	Text in the form "(Backwards OR forwards) snowballing from (...)"
Relevance	A priority score for how relevant this source was for this research; 1 indicating highly relevant vs 3 indicating slightly relevant	Number from 1 to 3
Reason for inclusion / exclusion	A justification of the relevance score; an explanation of why some papers were deemed relevant	Text
Summary	A summary of the abstract from the selected source	Text

Appendix B: Expert interview questions

Experience

- Q1: What is your role within the organization?
- Q2: Could you briefly describe your experience with data governance?
- Q3: Did you work on data governance projects at clients in which a data governance platform was introduced? *If not:*
 - Did you work on data governance projects without the introduction of a new system?
 - Did you work on other projects where a data-related system was implemented?
- Q4: Which project would you say was the most interesting to focus on for the rest of the interview?

Context project

- Q5: At which organization was the project carried out?
- Q6: What kind of organization is it; what are its characteristics?
- Q7: How experienced was the organization with data governance before the project started?
- Q8: What was the intended timeline of the project and when did the actual delivery of the data governance system start?
- Q9: What was the reason the project was initiated?
- Q10: From who or which part of the organization came the request to implement a data governance system?
- Q11: Which data governance functionalities did the organization implement in the data governance system?
- Q12: Which products were considered? Why was this specific product chosen?
- Q13: Did the organization have other data management related tools already in place?

Implementation proces

- Q14: Which changes did the organization need to make before the data governance system could be implemented?
- Q15: What data sources were used to connect the data governance system to?
- Q16: What were the main challenges during the implementation of the system?
- Q17: Are there any functionalities left that the organization wants to implement in the data governance system in the future? Which ones?
- Q18: If you compared this project with other projects in which a data governance platform (or another data management related tool) was introduced / implemented, what would you say was the main difference? Did these organizations have a very different approach on the implementation?

Outcome

- Q19: What is the current status of the project? Was the data governance system successfully implemented?
- Q20: How was the outcome / status of the data governance system measured?
- Q21: What benefits did the data governance system bring to the organization?

- Q22: Which factors in the organization had the biggest influence on the outcome / status of the project?

Other questions

- Q23: Is there something that you want to add to this topic, that we have not discussed yet?
- Q24: Do you have any recommendations for resources on this topic such as papers?
- Q25: Do you have any recommendations for any other contact persons?
- Q26: Do you have any questions left for my research?

Appendix C: Checklist from literature for input expert interviews

The following checklists were used as a guide during the interviews in order to elicit answers on whether this factor applied to the example organization of focus. For the questions which are marked with a “-”, there were either no answers with a fixed number of options, or there was not sufficient information from literature to base the answer categories on.

Context project

- Q5: At which organization was the project carried out? -
- Q6: What kind of organization is it; what are its characteristics? -
- Q7: How experienced was the organization with data governance before the project started?
 - Clear data strategy
 - Clear roles & ownership of data
 - Policies & standards
 - Central data management office / board in place
 - Formal data governance framework / plan / documentation
- Q8: What was the intended timeline of the project and when did the actual delivery of the data governance system start? -
- Q9: What was the reason the project was initiated?
 - Improving quality of organization’s data
 - Ensuring compliance / pressure from regulations
 - Ensuring security over data
 - Need for more efficient operations / processes
 - Need for improved data analytics
 - Need for one central overview of data to gain better understanding
 - Part of other transformation
 - To stay competitive in the long-run
- Q10: From who or which part of the organization came the request to implement a data governance system?
 - Top-level management
 - Chief data officer
 - IT department
 - External influence
- Q11: Which data governance functionalities did the organization implement in the data governance system?
 - Data catalog
 - Data glossary
 - Data lineage
 - Data classification
 - Data sharing
- Q12: Which products were considered? Why was this specific product chosen?
 - Microsoft Purview / Azure Data Catalog (predecessor)

- Collibra
- Informatica Data Governance & Catalog
- IBM Data Governance
- SAP Master Data Governance
- Q13: Did the organization have other data management related tools already in place?
 - Basic data systems
 - Master data management platform

Implementation proces

- Q14: Which changes did the organization need to make before the data governance system could be implemented?
 - Setting up a new organizational structure
 - Changes to data architecture / landscape
 - Changes in business processes
 - Changes to systems
- Q15: What data sources were used to connect the data governance system to? -
- Q16: What were the main challenges during the implementation of the system?
 - Resistance to change from employees
 - Low priority / lack of organizational commitment
 - Technology limitations
 - Complexity of data sources / data integration issues
 - Lack of resources
 - Costly / budget limitations
 - Lengthy implementation
 - Platform is difficult to learn / use
 - Business value is not clear
- Q17: Are there any functionalities left that the organization wants to implement in the data governance system in the future? Which ones? -
- Q18: If you compared this project with other projects in which a data governance platform (or another data management related tool) was introduced / implemented, what would you say was the main difference? Did these organizations have a very different approach on the implementation? -

Outcome

- Q19: What is the current status of the project? Was the data governance system successfully implemented?
 - Successfully implemented
 - Project was canceled
 - Work in progress
- Q20: How was the outcome / status of the data governance system measured? -
- Q21: What benefits did the data governance platform bring to the organization?
 - Financial

- Ensuring regulatory compliance
- Improvement in audit reports
- Higher data quality
- Q22: Which factors in the organization had the biggest influence on the outcome / status of the project?
 - Support from leadership
 - Commitment to change
 - Open culture
 - Well-defined KPIs
- Q23: Is there something that you want to add to this topic, that we have not discussed yet? -

Appendix D: Qualitative results of the interviews

In this Appendix, a summary is given of the most important parts of the eleven interviews that were conducted with subject matter experts. The more comprehensive answers to all interview questions are given in Appendix F, where answers were cited as much as possible.

Interview I1

During interview I1, the organization O1 implemented Collibra in order to improve their data quality across their different business lines.

- *Data governance experience interviewee:* The interviewee from interview I1 was an analytics consultant with a background as an enterprise data architect and had plenty of experience with Purview and setting up data governance frameworks, both from an organizational and a technical point of view.
- *Organization of focus:* Organization O1 was a start-up operating in the utility sector, and was a website that did comparative analysis on utility products and tried to recommend the best one.
- *Reason for project:* Since O1 was a start-up, their business grew in an organic manner around their product offerings, and they had never looked at an enterprise-wide strategy. At the time of the data governance undertaking, they had seven business lines for each of the utility products, and each line had their own systems. All systems were managed in isolation and they had massive issues around data quality.
- *Changes in organization:* As a first step, a committee was established that made sure that everything was running based on the new data governance strategy. Moreover, as part of the framework, the right processes were established which were discussed with different data owners and stewards. After this, they looked for tools that could improve data governance further.
- *Tools considered:* The organization looked at solutions in the Azure Cloud stack, but there was no solution yet from a data governance perspective except Azure Data Catalog. Instead, they looked at Collibra and Alation for their data catalog, and they finally went with Collibra. They also considered master data management (MDM) solutions for creating a single view.
- *Important factors:* A big factor that contributed to the initiative was leadership; data governance requires a bit of bureaucratic process with guidance from the top. This revolves around people and change management.
- *Other remarks:* Data governance solutions require a lot of advisory and proof of concepts (PoC), and only the last step is tools.

Interview I2

Interview I2 focused on organization O2, where Purview was in the process of being implemented in order to be able to use data to their advantage.

- *Data governance experience interviewee:* The interviewee had experience as a data architect and had lots of experience with helping organizations migrate to the Cloud, including migrations to Purview.

- *Organization of focus:* Organization O2 was operating in the financial industry, and recently started working on a data governance program. It had thought about data governance, and build their domain and systems in a direction that supported this.
- *Reason for project:* As O2 was a financial institution, they had a lot of personal information (PI) data, and they wanted to reduce the risk in order to protect the data of clients. Moreover, their data landscape was messy with lots of distributed data, and they wanted to have data governance to improve the data structure. They wanted more insights in this data structure, because they wanted to understand the business value of it and to use it to their advantage.
- *Changes in organization:* As a first step of the program, the organization wanted to set up the operating model. Secondly, to make the implementation possible, the organization had to change the structure of their data landscape, and they went with a data fabric or data mesh approach, which is a decentralized approach to the management of data.
- *Tools considered:* They went with Purview to start implementing their data catalog. No other systems were considered, as they had affinity with Microsoft technology and all their solutions were in the Azure Cloud so Purview was the easiest to implement.
- *Contribution factors:* It was a challenge to convince everyone of the new way of working, but what helped with this was visionary leadership, who were very proactive in carrying through the change.
- *Other remarks:* Data governance is gaining a far bigger scope, and it is necessary because of the growing volumes of data to have data governance in place. Earlier on data was not that distributed, contrary to now where data is coming in from many different systems.

Interview I3

Organization O3 was the focus of interview I3, where Azure Data Catalog, the predecessor of Purview, was implemented.

- *Data governance experience interviewee:* The interviewee was a solution architect for data platform implementations. The interviewee had practical experience with data governance, and was also familiar with the theory behind it.
- *Organization of focus:* O3 was part of the financial sector. It was a hierarchical organization regarding decisions that have to be made, but it was flat and informal regarding communication between employees and departments. The organization had already established a central data governance board.
- *Reason for project:* The organization wanted to have one central overview of their data, which is why the project was initiated.
- *Changes in organization:* Some technical challenges arised due to the limited functionalities and sources that could be scanned in Azure Data Catalog, and they had to build workarounds for this regarding the connection to data sources. Moreover, ownership had to be assigned.
- *Tools considered:* The project started with implementing a data platform, and Azure Data Catalog (predecessor of Purview) was a part of it. Since the organization already had a Microsoft Azure license, the choice for this particular platform was made quickly as their data was already in the Azure Cloud.
- *Contribution factors:* One thing that helped with the success of the project was that the Chief Data Officer was very clear in the communication about the changes that needed to be

made. Another factor that contributed to the project was that the tool was easy to apply for business users.

- *Other remarks:* Reporting, data cleaning, adding of business rules and assigning of policies were not yet possible. A desired functionality for the future was to have access to data centrally managed from this platform, as this can save a lot of work in manually sending emails requesting access.

Interview I4

The interviewee from interview I4 had seen many data governance programs and therefore provided insights on three organization cases, and additionally gave remarks on general data governance platform success factors. In Appendix D, the quantitative results for these three example organizations are accumulated under the one column for interview 4 (I4). In Appendix E with the qualitative results, it is described which specific example it applied to.

- *Data governance experience interviewee:* The interviewee was a consultant in the area of data platform implementation; the governance & management part, the architecture, and the technical implementation. I4 had seen many data governance implementation programs at clients and had a great deal of knowledge about general success factors that applied to all these different programs, and therefore during the interview kept it at a more general level instead of zooming in on one specific organization.
- Organization O4.1:
 - *Organization of focus:* The first organization, O4.1, was operating in the health sector and had many different member states operating in different countries, with data varying dramatically between member states. They used mostly Excel spreadsheets for their files.
 - *Reason for project:* They had problems with data quality, as they could not consistently ingest data from all the member states, since everyone collected data in different ways.
 - *Tools considered:* They wanted Purview for their data catalog.
 - *Changes in organization:* They built a portal to allow the simple uploading of data into the data platform, so that the person uploading the file could declare the data. These declarations were checked against the Purview catalog.
- Organization O4.2:
 - *Organization of focus:* O4.2, was operating in the consumer goods sector. This organization had a great deal of data in the Cloud, on-premise (mostly in SQL), and in SAP.
 - *Reason for project:* They wanted to build a data catalog of their data with Purview in order to get an idea of where they kept data across their data estate. They needed this understanding of their data because they wanted to apply policies and access control in the future.
 - *Tools considered:* Purview, as it integrated with the existing infrastructure.
 - *Changes in organization:* The data came in too raw of a form. Purview was not able to correlate data between different systems and it would have taken a lot of work to align data. As a workaround, they implemented a master data management (MDM)

tool that correlated data that was scanned by Purview using its own artificial intelligence (AI) solution.

- Organization O4.3:
 - *Organization of focus*: Organization O4.3 was operating in the manufacturing sector, and they acquired a great deal of companies over the years which resulted in many different brands under one organization.
 - *Reason for project*: Data was owned and controlled at brand level, and sometimes also per country. Brands were asking for data from other brands within the same company, but access to data was often denied.
 - *Changes in organization*: Minimal.
 - *Tools considered*: Purview.
 - *Contribution factors*: Purview did not help them, as they first needed to have a management and governance structure to enable them to open up and have free access to data.
- *Other remarks*: Demonstrated by these example cases, the interviewee drew some conclusions during the interview about the success factors within an organization in order to implement a data governance platform. Firstly, the level of data understanding influences the choice for a tool. Before a tool can be implemented, the organization should understand its data and put data ownership in place. Next, an indicator of a right level of maturity in the organization is if the organization has some sort of Chief Data Officer (CDO), as then there is commitment from the board to drive the organization in the right direction. Moreover, it is not necessarily about the size of the organization, but more about the data architecture and the extent of the data sources. In order to start seeing benefits of a data governance tool, there needs to be at least about five to seven complicated data sources, which can be scaled. Another option is to use a data swamp or data lake which is controlled by some kind of master data management, before it is registered in Purview. Furthermore, there needs to be established agreements between parts of the organization that own data for which political negotiations are important, as there need to be agreements about sharing data. Some general trends can also be said about the sector which might be an indicator of the readiness to start using a tool.

Interview I5

During interview I5, the interviewee explained that Informatica was implemented at organization O5 as part of a data platform transformation because they had to adapt to a new regulation.

- *Data governance experience interviewee*: Regarding data governance, the interviewee had worked on multiple data governance implementations programs at clients, but on an overall strategic level instead of technology
- *Organization of focus*: The interviewee chose to focus on organization O5 during the interview, which had quite some challenges during their data governance platform implementation project. O5 was a midsize financial organization, with a central office in the US, but with multiple subsidiaries spread far across different countries.
- *Reason for project*: O5 already had experience with data governance and had roles & responsibilities, data ownership, and policies already in place, but this needed a complete

revision. This was due to pressure from a new regulation, which required the organization to report on their compliance on a far more detailed level.

- *Changes in organization:* Revisions were made to the policies and ownership. A big data lake was implemented to get data in from multiple sources, which could act as a quality check or some sort of gatekeeping before the data would go to the data governance platform.
- *Tools considered:* The tool that O5 chose was Informatica, because of the technical capabilities, but also because the organization already had some other Informatica tools which made integration easier.
- *Contribution factors:* Subsidiaries were located in different countries so it was difficult to determine common definitions and establish a new way of working. A key factor in overcoming these challenges was a very good sponsor and mandate from top-level management. Also, they had buy-in from the rest of the organization since it was part of another transformational effort.

Interview 16

Interview 16 focused on a data governance program at an organization that wanted to improve its data governance structure, where Purview was not successfully implemented.

- *Data governance experience interviewee:* The interviewee had both experience with data governance projects at clients (in particular Purview), as well as a theoretical background on data governance from a research perspective.
- *Organization of focus:* The interviewee chose to zoom in on the project from organization O6, where Purview was not successfully implemented but where much learning experience was gained. O6 was part of the tourism sector, which due to the pandemic did not have a lot of business the past two years.
- *Reason for project:* The organization wanted to establish a stronger data management and governance model in order to stay competitive with the rest of the industry
- *Changes in organization:* Personas and business cases were established in the first week of the program to determine glossary terms. No other changes were made; the platform was first supposed to include one source and ten use cases, and from there on scale it.
- *Tools considered:* The organization chose Purview, because they were looking for an Azure native platform that already had an existing integration with Azure, and because Purview was significantly cheaper than competitors.
- *Contribution factors:* The organization was not mature enough in their data governance understanding in order to scale the system to more data sources. They first had to increase their data understanding.
- *Other remarks:* The interviewee added to the interview that an organization is very mature in their understanding of data governance if they recognize the additional value that a master data management (MDM) platform can provide. With a data governance platform you establish rules and policies, but with a MDM platform these are enforced.

Interview 17

In interview 17, organization O7 was discussed which wanted to implement Purview to improve their current processes and get insights into their PowerBI models with data lineage.

- *Data governance experience interviewee:* The interviewee from interview I7 worked as a data platform solution architect and had two years of experience on data governance projects at clients. The interviewee's role was to help oversee the three parts of data governance (people, processes, technology) and build a data governance strategy.
- *Organization of focus:* Organization O7 was a midsize financial company. They already started thinking about data governance years before the Purview implementation took place, for which they first laid the foundational work: implementing data stores and managing data integrations.
- *Reason for project:* The organization wanted technology solutions to help further drive their data governance journey, but also the advisory side on how to keep everything running smoothly and to help identify the owner structure of data. The objective of the Purview implementation was to leverage a PowerBI solution; to bring PowerBI models in, get insights in the lineage, and build a glossary around that.
- *Changes in organization:* Challenges were mainly technology related, so some workarounds had to be built. Next to that, the organization started a master management program to get a master data management structure in place to help organize their data first. The intention was that data would go from the sources to the master data management platform to a data warehouse, which leveraged the data to Purview.
- *Tools considered:* They chose Purview because they were migrating to Azure, and because it was cheaper than tools such as Informatica.
- *Contribution factors:* One big factor that contributed to the outcome of the project was the commitment to doing data governance within the organization. Moreover, another factor that contributed to the project was support from the business. The project started with IT, but all parts of the organization were quickly brought on the same page and funding helped to make it successful.

Interview I8

The focus of interview I8 was on an organization that did a proof of concept with Purview.

- *Data governance experience interviewee:* The interviewee from interview I8 was a data platform consultant, with a big focus on the technical side of data platform modernization. The interviewee had limited experience with data governance platforms, since the interviewee only did a proof of concept with Purview.
- *Organization of focus:* Organization O8 was operating in the energy sector, in particular the gas and oil industry. The organization had many different data sources, mostly coming from specialized on-premise systems.
- *Reason for project:* The proof of concept of Purview was part of a data platform implementation. Because the project was so big and so much data was involved, they needed a tool like Purview to be able to have an overview of their data estate.
- *Changes in organization:* For the data that the organization wanted an overview of, they imported this data in Azure and for this, Azure Data Lake storage was used. Purview can only connect to a few data sources (e.g. connectors for Oracle, SAP), but it cannot scan on-premise file shares, so importing it in the Cloud was a way to work with these restrictions.

- *Tools considered:* They also considered Informatica, Collibra, and a master data management (MDM) solution like Profisee, but they went with Purview because it was more cost-effective and it was geared towards the Azure Cloud.
- *Contribution factors:* The proof of concept was purely a technical set-up. To be able to use it further, a plan needs to be made for how to organize data in the tool (i.e. hierarchy, collections), what classifications should be considered, and who gets access. For the PoC, the lack of guidance from the business was intentional, but if an organization wants to put it into production, there needs to be a clear understanding of the metadata to enrich the data with.
- *Other remarks:* The interviewee highlighted at the end of the interview the growing importance of data governance. In the past there were very traditional structured data warehouses, which meant that there was less need for data governance. If there is a small project, the organization probably does not need a data catalog because it can still have a good overview without it. For bigger projects with lots of data involved, a tool like Purview becomes more critical. This does not have to replace the traditional data warehouse. A data warehouse is still good for ensuring data quality and it is easy to use. Moreover, the interviewee noted that MDM is not a prerequisite for Purview to be successful. An MDM solution is useful if there are many sources and data needs to be confirmed to ensure data quality.

Interview I9

Interview I9 was about the same program from organization O3 as discussed in interview I3, where I9 took over the maintenance and platform optimization trajectory from I3. The Azure Data Catalog was transformed to Purview during the implementation as discussed in interview I9.

- *Data governance experience interviewee:* The interviewee from interview I9 had approximately three years of experience with data governance due to a data platform implementation project in which Purview was implemented.
- *Organization of focus:* Organization O3 was operating in the financial sector and provided services such as leasing and vendor financing. O3 was a big organization operating worldwide. The Purview implementation was intended as a part of the new data platform, which had been running for two years.
- *Reason for project:* O3 knew a new version of the Data Catalog was coming with lots more functionalities and they wanted to have this, so it was driven by the organization itself.
- *Changes in organization:* At the beginning, the implementation was mainly focused on registering all of the data in the data catalog correctly, as it was necessary as a first step to have the metadata in the catalog in order to be able to use the platform properly. The biggest problems were mainly related to technical issues as some file extensions could not be scanned into Purview. Resultantly, it was chosen to start scanning sources at a curated data layer. This way, data could be standardized and a structure could be prepared that could be scanned into Purview more easily. Other changes that had to be made was the formation of a new team and hiring new people to fulfill specific roles related to the use of the new platform, next to the platform maintenance team who resolved any technical issues.
- *Tools considered:* Before this implementation, a tool selection evaluation was made to check all the competing product options, in which the main other product that was considered was

Collibra. However, the choice for Purview was made as it could be easily integrated with the existing Azure environment.

- *Contribution factors:* What was important for the implementation to be successful was the support from leadership. Purview was a complex undertaking, so enough resources and budget had to go to this implementation in order to be able to work on it.
- *Other remarks:* In the future, O3 is excited for functionalities such as managing access control from within Purview, and having more options for tracking data lineage.

Interview I10

At organization O9, Purview was implemented to gain a higher understanding so that the organization could ensure compliance and improve efficiency, which was discussed during interview I10.

- *Data governance experience interviewee:* Interview I10 was conducted with a solutions architect who had multiple years of experience with data governance projects at clients.
- *Organization of focus:* During the interview, it was chosen to focus on organization O9, which was a retail company operating in many countries, offering both offline and online products.
- *Reason for project:* The organization was recently looking at implementing Purview, because they had big issues with data management; they were not sure what data they had, whether they were compliant with regulations, and if they were using resources efficiently (they had high costs for data storage). This happened after their journey to the cloud, as they then realized they had to improve data management.
- *Changes in organization:* To prepare for the implementation, a data governance steering committee was set up with stakeholders who had rights to change the way of working. Moreover, a data protection officer was created, who was responsible for reviewing GDPR practices for sensitive data. Besides that, defining domains and their owners was also part of preparations. This is where most issues occurred, as it was not clear as to why some data was there and who its owners were. Other kinds of actions that needed to be taken to implement the platform were not clear yet, because first the current data landscape needed to be understood in order to make decisions about changes.
- *Tools considered:* No other considerations than Purview, as this was the quickest, cheapest to implement, and delivered the most value with the least amount of time.
- *Contribution factors:* What made this project different from previous projects was the type of data, namely sensitive client data, for which strong regulations needed to be followed.
- *Other remarks:* The interviewee noted that it was important to try and include as many data sources as possible in the data governance platform in order to make better decisions about data. If for example only one data lake is implemented, you do not get a big overview

Interview I11

In interview I11, the interviewee focused on the data governance program at organization O10, where Purview was implemented to help standardize data across different business units.

- *Data governance experience interviewee:* The interviewee from interview I11 was a solution architect, and had worked with clients for many years to deliver products for them, specifically by looking at the data architecture.

- *Organization of focus:* The client project in which Purview was implemented was at organization O10, which was operating in the energy industry. Although headquarters were in Europe, O10 had business units (BU) upcoast in different countries. Each BU had its own local market with different branding, and resultantly each BU had varying levels of maturity.
- *Reason for project:* The main goal was to standardize data governance for each BU. O10 wanted to improve their data analytics by bringing all BUs on the same level with guidance from headquarters, so that every unit followed the same standards and way of working. Everything was still a manual process, and there was a need to automate this so that it became manageable for the central governance function to exercise governance.
- *Changes in organization:* As a first step, O10 needed to set up access rules and guidelines, in order to have a standardized way of working in place. For example, the level of restricted access for production needed to be decided. New processes had to be implemented to make sure data was cataloged well, metadata could be discovered, and that the glossary was filled properly. O10 implemented a data platform first to standardize everything so that Purview could be connected to this platform.
- *Tools considered:* Purview is part of the Microsoft landscape, and since all of O10's solutions were moving towards Azure, the integration with Microsoft was important. Also, Collibra is much more expensive.
- *Contribution factors:* The request for this specific system came from the IT department, and it took a lot of effort to convince the business side as they did not understand the value of it. To solve this, pilot phases were started where it was explained to the users how the system should be used. O10 was a lot more open to try out the system before requesting the end value. They recognized they needed a solution as they found it was challenging to understand what data they had, and they were willing to try to see how the system fitted in the organization. They were open to first using it as a pilot before going forward.
- *Other remarks:* The interviewee remarked that a data governance solution is necessary for organizations to have in place. You cannot scale if data governance only takes place through manual effort. Some kind of automation is needed to realize this on a big scale, specifically with the current context in which so much is happening in the Cloud.

Appendix E: Quantitative results expert interviews

In this appendix, the answers from the interviews are labeled and categorized in groups. For each category, it is noted in the tables in which interviews they occurred. An explanation by means of a citation from the interview of how the answer category was decided is given in Appendix F. The categories that are noted in italics are categories that were only identified in the interviews, not in literature (see the factors from Appendix C).

The following notations were used for documenting the occurrences of the categories in the interviews:

- “ + ” was used when the category was seen as important (with “++” indicating a relatively higher importance).
- “ - ” was used when the category was lacking (with “- -” indicating a relatively bigger lack).
- “ 0 ” was used when the category did not have an effect or was not taken into account.
- “ x ” was used when the category was exclusive; it could be either present or absent, and in this case it was present.
- Cells were left empty when the category was not discussed during the interviews.

In the column at the end of the tables (Σ), the scores are summed up, which gives a general indication of the relative importance of each category. For each total sum, the following rules were used:

- Each category started with a score of 0.
- For “+”, +1 was added to the score, for “++” +2 was added to the score.
- For “-”, -1 was subtracted from the score, for “- -” -2 was subtracted from the score.
- In case of questions where answers were indicated with “x” and thus were exclusive and not relative, “x” translated to an addition of +1 to the score.
- Empty cells were ignored. “0” was treated as an empty cell.

The scores were used to make a ranking of how much each category was applicable.

Q7: How experienced was the organization with data governance before the project started?

Options	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	Σ
Central data governance office / board in place	--		++			--	+			-	+	-1
Clear data strategy		-						-		--	+	-3
<i>Established data definitions</i>		-	++			--		-			-	-3
Policies & standards		+	--	-	+	--	++		--	-		-4
Formal data governance framework / plan / documentation	--			--		--			-		-	-8
Clear roles & ownership of data			--		+	--	--		--	-	-	-9

Q8: What was the intended timeline of the project and when did the actual delivery of the data governance system start?

Options	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	Σ
<i>1-6 months</i>			x		x		x		x	x	x	6
<i>1-4 weeks</i>						x						1

Q9: What was the reason the project was initiated?

Options	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	Σ
Need for one central overview of data to gain better understanding	++	++	++	++	+			++		++	++	15
<i>Using the same data in multiple departments</i>	+		++	++							++	7
To stay competitive in the long-run		+				++		++	++			7
Ensuring compliance / pressure from regulations					++					++		4
Need for more efficient operations / processes							++			++		4
Improving quality of organization's data	++		0	++							-	3
Need for improved data analytics		+									++	3
<i>Reduce risk involved with data security</i>		++										2
Part of other transformation			+	0		0	0		+	0	0	2

Q10: From who or which part of the organization came the request to implement a data governance system?

Options	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	Σ
IT department							++			++	++	6
Chief data officer	++		++									4

<i>Enterprise architect</i>	+		+										2
External influence							+						1

Q11: Which data governance functionalities did the organization implement in the data governance system?

Options	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	Σ
Data catalog	x	x	x	x	x	x	x	x	x	x	x	10
Data glossary	x	x	x	x	x	x	x		x	x	x	9
Data lineage							x			x	x	3
Data classification				x	x						x	2
<i>Workflow management</i>					x							1
<i>Notifications</i>					x							1

Q12.1: Which products were considered?

Options	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	Σ
Microsoft Purview / Azure Data Catalog (predecessor)		x	x	x	x	x	x	x	x	x	x	4
Collibra	x							x	x		x	4
Informatica Data Governance & Catalog	x				x		x	x				4
<i>Alation Enterprise Data Catalog & Data Governance</i>	x											1
<i>Master data management (MDM) solutions</i>	x							x				2

Q12.2: Why was this specific product chosen?

Options	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	Σ
<i>Easy integration with the existing data infrastructure</i>		++	++	++	++	++	++	++		++	++	18
<i>Available functionalities in product</i>	++	+	+		++				++			8

<i>Cheaper option</i>	+					+	+	++			+	6
<i>Easy integration with other data management tools</i>						++						2

Q13: Did the organization have other data management related tools already in place? -

Options	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	Σ
Basic data systems	x	x	x	x	x	x	x	x	x	x	x	11
Master data management system									x			1

Q14: Which changes did the organization need to make before the data governance system could be implemented?

Options	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	Σ
Changes to data architecture / landscape		++	++	++	++				++		++	12
<i>Assigning roles and responsibilities / data ownership</i>			++		++	++				++	++	10
<i>Establish glossary terms / data definitions</i>		++		++	+	++					++	9
Changes in business processes	++	+		++	+	+				+	+	9
Setting up a new organizational structure	++								+	++		5
Changes to systems				++	+					+		4
<i>Changes in strategy</i>		++										2
<i>To be decided</i>										++		2

Q15: What data sources were used to connect the data governance system to?

Options	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	Σ
<i>Sources in the Cloud</i>		x				x	x	x	x	x	x	7
<i>Data management platform</i>				x	x		x		x	x	x	6

<i>Original source</i>				x		x						2
<i>External sources</i>										x		1

Q16: What were the main challenges during the implementation of the system?

Options	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	Σ
Technology limitations		+	++	++			++		++	-	+	9
<i>Confusion / reluctant about new way of working</i>		++	++		++					+	+	8
<i>Lack of data understanding</i>	+			++		++		++				7
Complexity of data sources / data integration issues			+	++		+			+	-		4
<i>Adoption / integration into everyday business</i>					++						++	4
<i>Identifying roles and responsibilities</i>					+					++		3
<i>Deciding on a common set of definitions</i>					++							2
Business value is not clear											++	2
Low priority / lack of support from business								+	-			0
Resistance to change from employees			-				--		-			-4

Q17: Are there any functionalities left that the organization wants to implement in the data governance system in the future? Which ones?

Options	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	Σ
<i>Data access control</i>		++	++	++					++		++	10
<i>Data lineage</i>			++		++			++	++			8
<i>Data classification</i>							++	++				4
<i>Data sharing</i>				++								2

<i>Data quality management</i>		++										2
<i>Data privacy / policy enforcement</i>		++										2
<i>Workflow management</i>											+	1
<i>More automation</i>					+							1

Q18: If you compared this project with other projects in which a data governance platform (or another data management related tool) was introduced / implemented, what would you say was the main difference? Did these organizations have a very different approach on the implementation?

Options	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	Σ
<i>Data governance maturity level</i>			++			+	+					4
<i>Willingness to change</i>							++				++	4
<i>Visionary leadership</i>		++										2
<i>Involvement from the business</i>									++			2
<i>Type of data</i>								+		++		2

Q19: What is the current status of the project? Was the data governance system successfully implemented?

Options	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	Σ
Work in progress	x	x	x	x	x		x	x	x	x		9
Successfully implemented	x		x		x		x		x		x	6
Project was canceled				x		x						2

Q20: How was the outcome / status of the data governance system measured?

Options	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	Σ
<i>Users can easily find data</i>			++		++				++			6
<i>All / most data in platform</i>					++						++	4
<i>Positive attitude of business users</i>	++	+										3

<i>towards the system</i>												
<i>Savings in costs</i>										++		2
<i>Data issues are discovered for resolving</i>	+									+		2
<i>More efficient communication</i>					++							2
<i>Data from the system is used for analysis</i>		++										2
<i>Delivery formalities</i>									+			1
<i>Correct use of the system</i>		+										1

Q21: What benefits did the data governance platform bring to the organization?

Options	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	Σ
<i>Not measurable yet</i>		x	x				x	x	x	x	x	7
<i>Ensuring regulatory compliance</i>					++							2
<i>Improvement in compliance reports</i>					+							1

Q22: Which factors in the organization had the biggest influence on the outcome / status of the project?

Options	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	Σ
<i>Support from leadership</i>	++	++	++		++							8
<i>Commitment to change from stakeholders</i>	+						++		++		++	7
<i>Good sponsorship</i>					++		++		++			6
<i>Extent to which data landscape is covered</i>				+						++		3
<i>Clear data strategy</i>	+										++	3
<i>Linking it to other transformation efforts</i>		+			++							3

<i>Ease of use of tool</i>			++									2
Open culture					0						+	1

Q23: Is there something that you want to add to this topic, that we have not discussed yet?

Options	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	Σ
<i>Pillars of data governance</i>	x	x		x		x	x	x		x		7
<i>Combination with other data management tools</i>	x			x		x		x		x		5
<i>Data governance maturity</i>	x			x		x		x				4
<i>Importance of data governance tools</i>		x		x				x			x	4
<i>Product choice for Purview</i>	x	x		x				x				4
<i>Complexity of data governance</i>	x			x							x	3
<i>Evolution of data governance</i>		x						x				2
<i>Thinking about data as a strategic asset / product</i>				x								1

Appendix F: Citations from interviews

In this appendix, all qualitative results are summarized in tables for each question. The tables are organized per answer category. For each answer category, it was noted in which interview it was addressed, together with an explanation of why it was addressed. These were used to decide on the quantitative results. Citations were used as much as possible (some answers were shortened, and some were translated to English).

Q2: Could you briefly describe your experience with data governance?

Answer	Interview #	Explanation
Multiple years of experience with data platforms and data governance platforms at clients, background as an enterprise data architect	I1	I1 was an analytics consultant, leading data governance and data management offerings, and had plenty of experience with Purview and setting up data governance frameworks. I1 also had an extensive background as an enterprise data architect, and therefore had plenty of technical knowledge about data governance as well.
Multiple years of experience with data platforms, recently a data governance program	I2	I2 was a consultant working on Azure technology offerings. The interviewee had experience as a data architect and had lots of experience with helping organizations migrate to the Cloud. I2's experience with data governance was not that long; I2 recently started on a data governance program at a client.
Multiple years of experience with data platforms and data governance platforms at clients	I3	I3 worked as a solution architect on the implementation of data platforms. I3 had multiple years of experience of implementing platforms at clients, including data governance solutions such as Microsoft Purview, Collibra. Besides, I3 had implemented separate data governance modules such as data catalogs as part of a data platform solution. I3 was also familiar with the theory behind data governance.
Multiple years of experience with data platforms and data governance platforms at clients	I4	I4 was a consultant in the area of data platform implementation. This involved the governance & management part, the architecture, and the technical implementation. I4 had many years of experience with these kinds of implementation programs at clients.
Worked on multiple data governance implementation projects at clients, from an overall strategic level instead of technology	I5	I5 was part of the technology advisory department, and gave clients advice about data, cloud, security, and enterprise technology transformations. Regarding data governance, the interviewee had worked on multiple data governance implementations programs at clients, but on an overall strategic level instead of technology. I5 had both experience at extremely big companies, as well as small companies, mostly in the financial sector. I5 had experience with projects where Purview was used, but did not implement it from a technical viewpoint. Additionally, I5 had experience with varying tools at clients, including Collibra, Informatica, and initial adoption of Purview.
Multiple years of experience with data governance projects at clients and a data governance research background	I6	I6 had multiple years of experience with data governance as a business technology consultant. I6 had both experience with data governance projects at clients, as well as a theoretical background on data governance from a research perspective. The client projects ranged from small (i.e. implementing one source) to an extremely large establishment of data governance across the entire enterprise.
Two years of experience with data	I7	I7 worked as a data platform solutions architect, and had worked on data governance projects for clients for the past two years. I7's role was to help

governance projects at clients		oversee the three parts of data governance (people, process, technology) at organizations, and helped on building a data governance strategy, not just technology. Regarding the experience with Purview, I7 did multiple Purview demos with clients.
Multiple years of experience with data platform implementation projects, limited experience with data governance	I8	I8 was a data platform consultant, with a big focus on the technical side of data platform modernization and a strong background in relational and analytical data platforms. I8 had limited experience with data governance, I8 only recently worked on a proof of concept with Purview at organization O8.
Three years of experience with a data platform project	I9	I9 was an analytics consultant and had about three years of experience with data governance when I9 started working on a data platform project. There, it started with filling Data Catalog Gen1 with metadata. Besides data governance, I9 had lots of experience with systems such as data warehouses.
Multiple years of experience with data governance projects at clients	I10	I10 had experience as a solution architect and worked on several projects that involved data governance. In a previous job, I10 gained experience on the implementation of a data governance strategy at a client, where multiple systems had to be integrated. With respect to Purview, I10 worked on proof of concepts and integrations with customer applications. Additionally, the interviewee had a great deal of experience with managing master data management (MDM) offerings.
Multiple years of experience with data platforms at clients	I11	I11 was a solution architect who had worked with clients for multiple years, and provided advice from an architectural perspective, as well as delivered products to clients. In the context of data, I11 was very familiar with data platforms, and focused on architecture best practices, setting up the right way of working, and AI-related topics such as machine learning. With respect to data governance, I11 had hands-on experience with Purview as I11 was working for the last half year on a data catalog solution for a client.

Q4: Which project would you say was the most interesting to focus on for the rest of the interview?

Answer	Interview #	Explanation
O1	I1	I1 had experience as an enterprise data architect at O1, where I1 experienced a data governance initiative. I1 became Chief Data Officer (CDO) of this initiative and worked on setting up a data governance framework.
O2	I2	I2 was working on a program at organization O2 during the time of the interview.
O3	I3	At O3, the project started as an implementation of a data platform, and Microsoft Purview (at the time called Data Catalog) was implemented as part of it.
O4.1	I4	In O4.1, they wanted to implement Purview as part of their data platform.
O.42	I4	O4.2 sought a solution for data discovery and had to look for a workaround as the initial Purview implementation did not work.
O4.3	I4	In O4.3, Purview was not implemented successfully.
O5	I5	I5 was now working on a data transformation project, where data governance was part of it but not the main focus. So instead, I5 chose to focus on a previous project at O5, where they had quite some challenges with data governance.

O6	I6	O6 started a Purview implementation which did not turn out to be successful, but did provide a big learning experience.
O7	I7	At O7 they implemented Purview with a PowerBI solution. They wanted to bring PowerBI models in, get insights into the lineage, and build a glossary around that.
O8	I8	I8 did a proof of concept at O8 with Purview.
O3	I9	I9 took over the project at organization O3 from I3. The data platform was expanded and optimized, which included an implementation of Microsoft Purview.
O9	I10	At O9, three to four years ago a data platform was introduced and implemented. The organization was recently looking at implementing Purview. This project is about to start soon.
O10	I11	At O10, a data platform was implemented, and Microsoft Purview was implemented as a solution for the data catalog.

Q6: What kind of organization is it; what are its characteristics?

Answer	Interview #	Explanation
Start-up which grew organically, different business lines for each utility product	I1	O1 was a start-up operating in the utility sector, and was essentially a website that did comparative analysis on utility products and tried to recommend the best one. As O1 had not been established quite recently and started as a start-up, O1 was built up organically and their business grew around their product offerings. They had seven different business lines for each utility product at the time of the program, each with different systems.
Distributed data, thought about data governance but no central data overview	I2	O2 had a messy data landscape as the data was very distributed. They had some form of governance in place, but not enough to drive the business further. They needed a central overview of their data to gain a better understanding.
Hierarchical for decisions, flat for communication	I3	Formally, O3 was a hierarchical organization regarding its decision-making, but everyone helped each other and there was informal communication.
No centrally managed program around data, very basic data systems (mostly Excel)	I4.1	O4.1 was operating in the health sector and had many different member states operating in different countries, with data varying dramatically between member states.
Many different systems without one centrally managed program	I4.2	O4.2 was operating in the consumer goods sector and did a lot of work regarding product development. This organization had a great deal of data in the Cloud, on-premise (mostly in SQL), and in SAP.
No central management or governance structure around data	I4.3	O4.3 was operating in the manufacturing sector, and they acquired a great deal of companies over the years which resulted in many different brands under one organization. Data was owned and controlled at brand level, and sometimes also per country.
Central office which did reporting, multiple subsidiaries	I5	O5 was a midsize financial organization. O5 had a group office in the US, but also a lot of subsidiaries spread far across multiple countries.
Lack of organizational	I6	O6 was operating in the tourism sector and had not been thriving the last few

structure surrounding data management		years due to the pandemic. O6 had a lack of formal organizational structure surrounding data management. Maybe this was because of the sector (tourism) or because they did not yet have to think about their data landscape.
Flat organization	I7	O7 was operating in the financial sector. The organization did not have a deep hierarchy, it was rather flat.
Many different systems without one centrally managed program	I8	O8 was operating in the oil and gas industry. The organization had many different data sources, mostly coming from specialized on-premise systems, and therefore data in the organization was very distributed without one central overview.
Operating worldwide	I9	O3 was a financial organization offering services such as leasing and vendor financing. O3 was a big organization with offices in more than 30 countries, operating worldwide.
Operating in multiple European countries	I10	O9 was a retail company which was operating in multiple European countries. The company offered products both offline and online.
Central programs from headquarters, but local business units	I11	Although headquarters were located in Europe, O10 had business units (BU) upcoast in different countries. The headquarters managed the standard data governance or any kinds of new programs. However, the BUs each had their own local market with their own branding, and resultantly they each had varying levels of maturity.

Q7: How experienced was the organization with data governance before the project started?

Answer category	Interview #	Explanation
Clear data strategy	I2	O2 had some form of data governance in place in the organization, but not to the extent necessary to drive their business further as they were lacking some central direction.
	I8	O8 did not define what the main goal was for the implementation of Purview.
	I10	O9 was not very mature with data governance. Something that interviewee I10 saw in many organizations, and was also the case at O9, is that organizations tend to start using new systems at an early stage because technology is becoming available quicker and cheaper. However, at a later stage they find out they have issues with data. At O9, data governance was not a strategy at the implementation of the Cloud journey, but something they did later when they realized they needed to improve the situation.
	I11	O10 had a clear strategy to standardize. The idea was to get the BUs on the same level of maturity and have headquarters guide them through the process. All datasets should be in place in the central system to get an idea of how to govern it properly. This strategy was specifically relevant for the acquisition of new companies, as they wanted to know the data estate of newly acquired companies in order to know how to govern it.
Clear roles & ownership of data	I5	O5 had data governance (like all big companies), and already had established roles & responsibilities in the past. However, this needed to completely change.
	I7	O7 did not yet have data owners (for example data stewards and other organizational bodies). That was what the advisory part of the project mainly was about.
Policies & standards	I2	O2 had some form of data governance in place, mostly in the form of policies.
	I4	O4.2 needed this understanding of their data because they wanted to apply

		policies and access control in the future.
	I5	O5 also had data policies in place, although they had to recreate policies for the new way of working.
	I7	O7 had already established policies & standards in place before the platform was implemented.
Central data governance office / board in place	I1	As a first step of the data governance undertaking, a central committee was established, which was not there beforehand.
	I3	O3 had already established a central data governance board before the data platform project.
	I11	When O10 started on the journey, they knew about the Cloud but they did not know much about data governance. They had a central function for the security and governance department, but they did not know whether every BU was complying to that. Everything was still a manual process, and there was a need to automate this so that it became manageable for the central governance function to exercise governance.
Established data definitions	I2	O2 did not have a proper way for defining a common vocabulary for their data.
	I3	As a tool for data governance, O3 used an Excel sheet with all data definitions, which was placed on Sharepoint.
Formal data governance framework / plan / documentation	I1	The initiative was intended to set up a formal data governance framework, as O1 did not have a plan for this.
	I4	Purview did not help O4.3, as they first needed to have a management and governance structure to enable them to open up and have free access to data.
	I6	There was no experience with data governance at O6. Two people were familiar with Purview (they knew a few key points), but they only understood data governance concepts (e.g. about architecture, design) from a very high level. There were no plans or documentation for personas (e.g. data stewards), glossary terms, or classifications, and the organization had no business use cases for this.
	I9	There was not a formal plan for how to manage data at O3. This is now a work in progress, and recently formal documentation was made about data retention.

Q8: What was the intended timeline of the project and when did the actual delivery of the data governance system start?

Answer category	Interview #	Explanation
1-4 weeks	I6	The project was supposed to take four weeks, as it was scoped to include only one data source and ten use cases (for each use case, glossary terms were found and lineage was tracked). From there on, the project could be scaled and more data sources could be added. In this case, it would turn into a project of a few months. At week 0, time was spent on deciding who the stakeholders and data owners were, and what the business use cases consisted of in order to establish the glossary terms.
1-6 months	I3	Initially, the project started with the implementation of a data platform (including a data lake, data factory, databricks solution), and Purview was a part of it. At the time, it was called Azure Data Catalog, as Purview did not yet exist.

		Azure Data Catalog was similar in functionality to Purview, only with fewer options. The platform was implemented in about 3 to 4 months. After this it was transferred to a maintenance and platform optimization trajectory.
	I5	The data governance platform implementation program was part of a bigger program. They had to adapt to a new regulation, and for that they had to keep a data platform up and running. It took 3 to 4 months to define what would be the new way of working; including more data sources, aligning data exchange more frequently, and assigning more responsibility to data owners.
	I7	O7 already started thinking about data governance years ago, when they did the foundational work: implementing data stores and managing data integrations. This year they started on Purview, and as a first step they wanted a service model for PowerBI to leverage lineage. They implemented the service & sales data first in Purview. Their master management program kicked off last month in order to get a master data management in place to help them organize their data.
	I9	It was difficult to put a timeline on the project. The implementation had been running for two years, including the private preview part of Purview. First, the data platform was implemented, and Data Catalog Gen1 was part of this from the beginning, and all data had to be registered correctly in the data catalog. The last half year was really spent on getting Purview into the production environment. The implementation process to make the system ready for end users is finished, and now is the phase where they put the system to use.
	I10	O9 is currently doing a proof of concept (PoC). After this, they have a month for the initial implementation of the data catalog. However, this did not include full implementation of the data governance strategy. Besides the implementation, an analysis of the current state is made, together with a decision of where O9 wants to be in the future, for which there are 3 months. Data governance is a recurring activity, as it is never finished.
	I11	Implementing Azure Purview took 3 months. That was all it takes to set it up technically. However, to make people start using the system was an ongoing process. People needed to be trained and helped with what they had to do. For example, different roles should use the system differently (e.g. a compliance officer has a different use than a developer or business user). These segments needed to be educated differently about the use of the system. So the entire data governance journey is much longer.

Q9: What was the reason the project was initiated?

Answer category	Interview #	Explanation
Improving quality of organization's data	I1	They built a lot of point to point integrations to make sure systems / data were syncing to a certain extent, but they still had massive issues around data quality. For example, the same customer that was registered in different systems was considered as two different people within the organization.
	I4	O4.1 had problems with data quality, as they could not consistently ingest data from all the member states, since everyone collected data in different ways.
Ensuring compliance / pressure from regulations	I5	O5 had to adapt to a new regulation. This new regulation required O5 to report on a far more detailed level. As a result, O5 had to gather data from every subsidiary on a very detailed level. This created pressure on current technology, but also on how to manage this quality.
	I10	A driver of the strategy was compliance. Because of the GDPR, O9 was required to tell customers what to do with data and have processes in place to ensure

		this. However, O9 was not sure all data was compliant.
Reduce risk involved with data security	I2	As O2 was a financial institution, they had a lot of personal information (PI) data, and they wanted to reduce the risk in order to protect the data of clients.
Need for more efficient operations / processes	I7	O7 wanted technology solutions to help further drive their data governance journey. The project was not just technology, but also advisory; how to keep everything running smoothly and to help identify the owner structure of data.
	I10	Because O9 did not have visibility in what data they had, it sometimes happened that data was requested that was not necessary or had been requested in the past. A better visibility in this would save time and effort.
Need for improved data analytics	I2	O2 wanted more insights in their data structure, because they wanted to understand the business value of it and to use it to their advantage.
	I11	The strategy was to greatly improve data analytics. The intention was to bring all BUs on the same level with guidance from headquarters, so that every unit followed the same standards and way of working. They wanted to have a data platform that every BU could work with, so that applications could be built with the same security practices and the same data ingestion patterns.
Need for one central overview of data to gain better understanding	I1	O1 had a messy data estate, and a lot of their IT processes were running in a tangled manner. Since they were a start-up, their business grew in an organic manner and data started piling up. They had never looked at an enterprise wide strategy about how to manage data, and how to manage processes around it.
	I2	The data landscape of O2 was messy with lots of distributed data, and they wanted to have data governance to improve the data structure.
	I3	At O3, there were lots of point-to-point solutions. Data was divided over the whole organization. The intention was to have data accessible from one central place.
	I4	O4.2 wanted to build a data catalog of their data with Purview in order to get an idea of where they kept data across their data estate.
	I8	O8 wanted to create a data platform based on Azure where they were pulling the data out of the systems and into the data platform. The implementation of the data platform was a massive project with around 200 people involved. Due to their large amount of different data, they needed a solution to organize and find data. Because the project was so big and so much data was involved, they needed a tool like Purview to be able to have an overview of their data estate. As part of this project, they worked on a proof of concept of Purview.
	I10	The big issue at O9 was related to the management of data; they were not sure what data they had, whether they were compliant with regulations, and they were not sure if they were using resources efficiently (they had high storage costs). O9 selected Purview as a technology to get an overview of their data landscape to understand what they had. Once they had that, they could make decisions regarding compliance and improving efficiency.
	I11	As part of the data platform, Purview was implemented as the data catalog solution. It was meant as a way to scan the entire data estate of all BUs and build meta-data around. The fundamental goal was to standardize.
Using the same data in multiple departments	I1	O1 had seven business lines for each of the utility products, and each line had their own systems (e.g. their own CRM systems). All systems were managed in isolation, which made it difficult to manage data between departments.

	I3	It was the case that in O3 they wanted to use synergies of data in different departments.
	I4	Brands were asking for data from other brands at O4.3 within the same company, but access to data was often denied.
	I11	O3 wanted to make it faster for teams to build their solutions (e.g. reports) and make time to market faster.
Part of other transformation	I3	O3 had to make a decision; either buy new hardware, or move data to the Cloud. Advice was given about moving to the Cloud.
	I7	The journey to the Cloud was part of the transformation at the organization, but it was not part of the data governance part.
	I9	The data platform was meant to bring a solution for how to manage data at O3. The journey to the Cloud was directly linked to this.
	I10	O9 did not start the Purview implementation because they were moving to the cloud. They did it way after the transformation to the Cloud, because they realized they had to improve data management due to all the issues.
To stay competitive in the long-run	I6	O6 had just moved to the Cloud, and wanted to establish a stronger data management and governance model. The reason for this was to keep up with the industry and to stay competitive.
	I8	O8 wanted to transform to an industry-relevant format for their data and stay competitive. They want to be compatible in data format with other organizations.
	I9	The project was really driven by the customer. O3 knew a new version of the Data Catalog was coming with lots more functionalities and they wanted to have this.

Q10: From who or which part of the organization came the request to implement a data governance system?

Answer category	Interview #	Explanation
Chief Data Officer	I1	I1 became Chief Data Officer to drive the data governance initiative further.
	I3	The Chief Data Officer, who now has become vice president of O3, pushed through the request for a data governance system.
IT department	I7	The IT department drove the initial request for Purview, but they quickly built consensus with the CEO and c-level management. When they got the system, they had buy-in already.
	I10	The request for Purview came from the IT department. Usually initiatives come from IT because they face project challenges. However, despite the request from IT, the business side was required to be involved. For decisions about changing applications or changing the way of working, support from the business was required.
	I11	The IT department chose to implement this system. Microsoft supported them by informing them about what kind of challenges the system solves. It took a lot of time to convince the business as they did not understand the value of it. To convince them, a pilot phase was started. There were pilots with a small subset of users, their feedback was gathered, and it was explained how the system is

		helping them. It is still a process and people still have to be trained.
Enterprise architect	I1	I1 was working as an enterprise architect at the time of the initiative, and became very involved with setting up the data governance framework.
	I3	There was collaboration between the enterprise architect of O3 to implement the system.
External influence	I7	Initially, a person who owned data for O7's client came over, and was the driver for the project. This person had an analytics background and knew what good data governance looked like, and recommended a roadmap for improving data governance.

Q11: Which data governance functionalities did the organization implement in the data governance system?

Answer category	Interview #	Explanation
Data catalog	I1	O1 wanted to have a data catalog in their data governance solution.
	I2	O2 started with implementing the data catalog.
	I3	The data governance components started with Azure Data Catalog. All that could be done with Azure Data Catalog was the labeling of raw data. Purview brought more possibilities and these were implemented at a later stage.
	I4	O4.1 used Purview for metadata definitions and building a data dictionary.
	I7	The plan was to first implement a data management platform, and build the catalog on top of that.
	I8	The organization was looking at Purview to implement a data catalog. At the time of the interview, only the data catalog was implemented.
	I9	At the beginning, data had to be moved to the Cloud. In the reference architecture, it was described that when you move data to the Cloud, there must first be a data catalog. The biggest focus of the implementation, which also took the most time, was getting all the data in the data catalog. To be able to technically use the platform and search for data, it was necessary to have the metadata in the data catalog. This was the part where most technical problems occurred.
Data glossary	I3	The data glossary functionality was implemented, but in Azure Data Catalog this was reduced to 5000 glossary items. These limitations were why a switch was made to Purview.
	I7	O7 built the data glossary around their PowerBI models.
	I6	The functionality that O6 wanted was the data glossary. A lot of organizations start with the data glossary, because they do not know where to go from there. This was also the case for O6; due to their lack of data governance knowledge they were only looking for glossary terms. To have the data glossary, data source connections had to be set up, data had to be imported, and the glossary had to be organized with terms and classifications.
Data lineage	I7	Insights from the data lineage helped drive business analysis in terms of policies and client retention.
	I10	O9 did not know what data they had, so they did not know the source. This was

		something O9 wanted to know, although it was not the main driver. Purview provided by default lineage for data that was in the Cloud, but for data sources for on-premise data, the data lineage could not be shown.
Data classifications	I4	At O4.2, they wanted to declare data in order to gain a higher understanding of the data landscape.

Q12.1: Which products were considered?

Answer category	Interview #	Explanation
Collibra	I1	O1 looked at solutions in the Azure Cloud stack, but there was no solution yet from a data governance perspective except Azure Data Catalog, but that offers just one of the capabilities required to establish data governance. Instead, they looked at Collibra and Alation for their data catalog. They finally went with Collibra.
	I9	A tooling selection evaluation was made to check all the competing product options. Considerations of these products were the features that they offered. One of the products that was mainly considered was Collibra, which was a widely used product with lots of people experienced in it.
	I11	Collibra was also considered. O10 chose Purview for two reasons.
Informatica Data Governance & Catalog	I5	Informatica was implemented at O5.
	I8	Informatica is the most comprehensive solution and includes most functionalities.
Master data management (MDM) solutions	I1	During the initiative they also looked at master data management (MDM) solutions for creating a single view of the customer, which manages master data and enables the organization to create goals and records for a domain with consolidated data. This involved matching processes, standardization, and improvement of data quality.

Q12.2: Why was this specific product chosen?

Answer category	Interview #	Explanation
Easy integration with the existing data environment	I2	No other systems were considered at O2, as they had affinity with Microsoft technology and all their solutions were in the Azure Cloud, so Purview was the easiest to implement. Purview had a cloud-based model that could be activated in 10 minutes in a cost-effective way, so this provided scalability and flexibility.
	I3	The choice for Purview was made very quickly, because O3 had everything in the Azure cloud. Another tool would require another license, and transfer to another cloud. Having everything in the Azure cloud also meant that the lineage could be tracked over the whole pipeline.
	I4	Purview is the most logical tool to implement when the organization makes use of the Azure Cloud.
	I5	O5 already used a few other Informatica tools, so integration was easy.
	I6	Six months ago, people did not trust Purview yet because there were many more mature options that had been tested. However, now a lot more organizations, including O6, are on Azure so it makes sense to continue with an Azure native platform that already has that integration.

	I7	O7 was migrating to Azure already.
	I8	They also considered Informatica, Collibra, and a master data management (MDM) solution like Profisee, but they went with Purview because it was more cost-effective and it was geared towards the Azure Cloud.
	I9	The choice to use Purview was made because it integrated with the Azure environment. O3 already had an Azure license, so the tool could be easily used within the Azure environment.
	I10	O9 worked fully with the Azure Cloud, so they only considered Microsoft technologies. This was the quickest, cheapest to implement, and delivered the most value with the least amount of time.
	I11	Purview is part of the Microsoft landscape, and since all of O10's solutions were moving towards Azure, the integration with Microsoft was important.
Available functionalities in product	I1	The choice for a tool was decided by the use case at the organization. It depended on how the requirements evolved and for what data it was going to be used.
	I2	Purview had lots of future prospects as it was quickly catching up with other services, so O2 bought into the future perspective of the functionalities that were coming to Purview.
	I5	The technical capabilities of Informatica were good.
	I9	The choice was mainly driven by technical reasons, not process reasons. Purview could offer specific functionalities that were considered important for O3. The private preview of Purview played an important part in this choice, since that could demonstrate how the product worked.
Cheaper option	I1	Informatica was also considered, as that is the most comprehensive solution, but it is not available as a single component to plug in and requires a lot of investments and money.
	I7	Cost was the main driver to go for Purview instead of tools like Informatica.
	I11	Collibra is much more expensive.
Easy integration with other data management tools	I6	Purview partnered with Profisee and a lot of other MDM platforms. When an organization knows that a MDM platform has to be added eventually and Purview already has all the connections and integrations with these MDM platforms, the choice for Purview is straightforward.

Q13: Did the organization have other data management related tools already in place?

Answer category	Interview #	Explanation
Basic data systems	I1	The data in O1 was stored and used in the different systems used by different business lines, e.g. CRM systems.
	I2	O2 had no other data management tools in place. They had thought about data governance and started building their data domain and hierarchy in that direction, but they had no tools for this in their enterprise system.
	I4	O4.1 had very basic data systems, as most of the files they used were Excel spreadsheets.

	I5	O5 had Excel sheets, and a few scattered data tools (e.g. data modeling tool for ownership management), but not a professional tool.
	I6	O6 was so new to data governance that they were not looking for a super fast acceleration plan to also establish a data management platform.
	I9	A data warehouse was running on premise for 10 to 15 years and was fundamental to O3. This was needed for their compliance reporting.
	I10	O9 did not have other data management tools in place, but they wanted to have a master data management (MDM) tool.
	I11	O10 used tools such as SQL servers for managing data, but nothing comparable to a data management platform.
Master data management system	I9	At a very small scale, they had some separate master data management systems, but not in the context of an organization-wide initiative. They are now considering a master data management platform for master data management and data quality. There were 2 main candidates considered in the tool selection evaluation: Informatica and Ataccama. They finally went with Informatica because it satisfied all requirements.

Q14: Which changes did the organization need to make before the data governance system could be implemented?

Answer category	Interview #	Explanation
Setting up a new organizational structure	I1	As a first step, a committee was established at O1 that made sure that everything was running based on the new data governance strategy. The committee looked at the different initiatives within the organization, and made the strategic plans about what needed to be executed and how this could be aligned with existing initiatives.
	I9	Next to Purview working well technically, new roles had to be filled. A new team had to be formed, and people had to be hired in order to fulfill a specific role, next to the platform team that takes care of any technical issues.
	I10	O9 created a data governance steering committee. After the implementation, more actions needed to be taken. However, in order to potentially change the way of working, the right stakeholders needed to be part of this initiative, and they needed to be included in the steering committee. Besides, they created a data protection officer. This person was responsible for implementing and reviewing GDPR practices for sensitive data.
Changes to data architecture / landscape	I2	To make the implementation possible, the organization had to change the structure of their data landscape, and they went with a data fabric or data mesh approach, which is a decentralized approach to the management of data. This was an initiative from the organization itself. They wanted to adapt to what was latest in the market to be able to grow in the future quickly.
	I3	For Azure Data Catalog, a data platform architecture was designed at O3 which was focused on modularity. It could be seen as a hotel, where you need to check into rooms. There is a central hotel lobby where all data is kept. Yet you can rent your own room and put your own Azure solution there.
	I4	As a solution, O4.1 built a portal to allow the simple uploading of data into the data platform, so that the person uploading the file could declare the data.
	I5	The right architecture had to be decided.

	I9	A curated data layer had to be prepared where data was standardized to a certain extent and where there was a fixed structure.
	I11	O10 implemented a data platform first to standardize everything. They created automation around it, so that every time they set up a platform it is through automated scripts. Then they put a catalog around it. More teams could bring in data to the centralized platform with the standardized way of working.
Assigning roles and responsibilities	I3	People had to show ownership when the system was implemented at O3. They were told they were the owner of the data source, and as a result also the owner of the data, and that was new for them.
	I5	Every subsidiary needed to send data to the central office, and the central office would do the reporting. New rules had to be created for who was responsible for the end reports, who was responsible for sending the right data, and who was responsible for the data quality in the reports. One of the key roles was the defining of the data stewards, so that the business could understand who could take ownership of the data. The organizational alignment and buy-in (i.e. agreement) for the responsibilities of that role was also very important, because people also had to accept and respect the new responsibilities. It did not have to be a full-time role (although it was always better to have it as a full-time role), since sometimes the dataset was so small that someone could take it as a +1 responsibility.
	I10	The data catalog allowed O9 to scan the system, to identify owners, and set up processes needed to manage datasets. O9 set up owners by domain (e.g. customer data, management data), and for each of those domains, someone was responsible for making sure that dataset was compliant. Defining domains and owners was part of the MDM strategy after Purview was implemented.
	I11	As a first step, O10 needed to set up access rules and guidelines, in order to have a standardized way of working in place. For example, the level of restricted access for production needed to be decided.
Establish glossary terms / data definitions	I2	During the program, O2 wanted to unify how to label the data.
	I4	At O4.1, they had to decide on uniform definitions for all member states.
	I6	At week 0, time was spent on deciding who the stakeholders and data owners were, and what the business use cases consisted of in order to establish the glossary terms.
Changes in strategy	I2	As a first step of the program, O2 wanted to set up the operating model of the organization and change the way they looked at their data.
Changes in business processes	I1	At O1, a data governance framework was made in which the right processes were established for maintaining good data governance. These were discussed with data owners and data stewards, and led to a lot of conversation about data governance.
	I4	O4.3 first needed to have a management and governance structure to enable them to open up and have free access to data.
	I5	There was a change in the operating model. O5 had to align on policies about how data was brought to the central office, and how to govern data quality.
	I10	Processes were set in place about who can make changes to data, approve, etc.
	I11	New processes had to be implemented to make sure data was cataloged well, metadata could be discovered, and that the glossary was filled properly.

Changes to systems	I4	O4.2 first had to implement a master data management solution before data could be declared in Purview.
	I5	A big data lake and data warehouse was implemented at O5. This was done to get data in from multiple sources. This acted as a way for specifying how to store the data, how to define the right controls for who owns what data, and it could act as a quality check or gate keeping. However, a tool is just a tool, but it needed integration with the right processes and organizational structure to make it useful.
To be decided	I10	First, it was necessary to understand the current data landscape. After that is more clear, it will be decided which changes to make. Probably changes need to be made to systems, the way of working, and processes.

Q15: What data sources were used to connect the data governance system to?

Answer category	Interview #	Explanation
Original source	I4	At O4.2, the master data management platform looked at all the data that was scanned by Purview (from its original source), and then correlated that data.
Data management platform	I4	O4.1 built a portal to allow the simple uploading of data into the data platform.
	I5	A big data lake was implemented where the data governance platform was connected to.
	I7	There were several custom applications (e.g. CRM) at O7 with data for Purview. These systems were moving to master data management, for which O7 chose Profisee as it was a direct integration with Purview. Data was going from these sources to master data management to a data warehouse, which leveraged data to Purview.
	I9	The question was where to start with scanning; raw data or curated data. In the curated data layer, data can be prepared by standardizing it and giving it a fixed structure. The scanning of raw data was at some point considered, but there were too many technical limitations. Some file extensions are not supported in Purview for scanning (e.g. .praq extension, multiple separators in flat files). Consequently, the choice for a curated layer was made so that all files could be registered and scanned.
	I11	Purview was connected to the central data management platform that was implemented instead of the original source.
Sources in the Cloud	I2	The implementation of Purview was part of a bigger program, where data from their existing enterprise system was moved to the Cloud. Purview was connected to these sources.
	I8	There were big repositories of data on-premise with high access-control, so very few people could access it. For the data that the organization wanted an overview of, they imported this data in Azure. For this, Azure Data Lake storage was used, and Purview was looking at different areas in separate data lake storages in Azure. Purview can only connect to a few data sources (e.g. connectors for Oracle, SAP), but it cannot scan on-premise file shares, so importing it in the Cloud was a way to work with these restrictions.
	I10	For the sources that were already in the Azure Cloud, they were very easy to connect.
External sources	I10	Data sources included SAP, the online sales application, data lakes, databases, and data factories. For the external sources they had to set up connectivity, but

		no changes were required to the source.
--	--	---

Q16: What were the main challenges during the implementation of the system?

Answer category	Interview #	Explanation
Technology limitations	I2	Another challenge was to choose which service to use. Because Purview was a new product, this caused some worry in the organization. However, this was solved by showing all the motivations of Microsoft for the product, so that the organization and Microsoft were aligned.
	I3	There were some problems because Purview was a new product and had limited functionality. It still keeps getting better. When it first was introduced, it was a good tool for metadata management, but it was not really a tool that could be used for reporting, data cleaning, adding business rules, and assigning policies. This took a while.
	I4	The data came in too raw of a form at O4.2. Purview could identify the metadata, but a lot of the identification and association in Purview did not happen as it should. Purview is good with generic terms (e.g. credit card numbers, date of births), but not for data associated with product understanding. Resultantly, Purview was not able to correlate data between different systems and it would have taken a lot of work to align data.
	I7	The main challenges during the implementation were technology related. Some analysis services did not support lineage, and there were similar values for different glossary terms (e.g. invoice day, receive day, etc). The lineage itself even had a bug, for which Microsoft had to be contacted. Purview is still a young immature product which has room to grow. Every organization will face similar challenges when using such a new tool
	I9	The technical challenges were the biggest problem, which sometimes took multiple months. In order to resolve some issues, Microsoft had to be contacted and discussions with partners took place in order to come up with workarounds.
	I10	The implementation was not very complex because O9 already had everything from Microsoft.
Complexity of data sources / data integration issues	I3	Sometimes the scanning of sources did not go well. Particular .zip formats were used that could not be scanned in Purview, so other solutions had to be found.
	I4	Due to the variety of data sources at O4.1, they had to build a portal to allow the simple uploading of data into the data platform.
	I6	O6 did not know where the data was coming from. They scoped the implementation to 1 or 2 data sources, but they did not know where those were. Many organizations have this problem.
	I9	Certain file extensions could not be imported in Purview so workarounds from these data sources had to be found.
Deciding on a common set of definitions	I5	There were so many subsidiaries, that it was difficult to determine a common set of definitions, rules, and data classifications.
Identifying roles and responsibilities	I10	As a result of the data catalog, sources were found for which owners had to be identified. However, for a lot of data O9 did not know why it was there or if it was obsolete. Identifying owners was difficult for this.

Low priority / lack of support from business	I8	As O8 only did a proof of concept leaning on the technical side and not a full implementation, some organizational challenges arose. They did not have any guidance from the business as it was a purely technical set-up. Therefore, they were experimenting with Purview and trying to figure out what was best without a clear goal from the business. However, for Purview it is important to know how to organize data; what kind of hierarchy should be created for collections (i.e. where data comes from), how data interrelates, and who is allowed to have access to collections. For another organization they should think carefully about how this data is organized prior to the implementation.
Confusion / reluctance about new way of working	I2	The organization was big in size, and it was a challenge to convince every person in the organization of the new way of working, as there are always people who prefer to work the traditional way.
	I3	There were some challenges with the assignment of ownership of data. Employees had to be helped with this.
	I5	There were some political challenges at O5, and change management was challenging. Due to the different subsidiaries being located in different countries, there were different cultures, and different approaches had to be used to make something happen. To bring about new change, it was important to do lobbying, to have implications and clear communication in place, and to make a mandate (i.e. need to this this, not optional).
Business value is not clear	I11	It was an IT-driven project, and it was not exactly clear how the business would use it. As a result, requirements had to be understood and validated from a business perspective. Purview is now in General Availability (GA), but there are still many private preview features which have to be validated. To show the business value, an initial version had to be implemented to demonstrate it to the business (which can be seen as a chicken and egg problem). Once it is found out how the business uses it, the system can be scaled.
Adoption / integration into everyday business processes	I5	It was a challenge to come up with a balanced approach for transitioning into the new way of working. Instead of reporting once a month, weekly reports needed to be made. How to create this transition roadmap was tricky, but also one of the most important parts. You can get the best tool, but if there is no adoption journey, it can get challenging.
	I11	How much value the business could get out of it depends on its use. It is good to know the information exists and that you can look for it, but until it gets integrated in the actual business processes and way of working, it is still an external solution. As an example, after every product, teams were expected to fill all the glossary parts, which is a task that is encouraged to do but it is not yet happening. Integration into the core process lifecycle still poses a challenge.
Lack of data understanding	I4	O4.3 first had to understand the data they had before they could implement a tool.
	I6	O6 had no clear organization for its data. O6 had very little knowledge about where data was coming from.
	I8	O8 did not define what the main goal was of what they wanted to see in their data catalog. However, in order for a data governance tool to work, there needs to be a plan for data classifications and how to deploy these classifications to the data. If Purview is just implemented without a particular use in mind, the usability of the system is quite small. Resultantly, it was not very easy to find data in the catalog that was implemented. For the PoC, the lack of guidance from the business was intentional, but if an organization wants to put it into production, there needs to be a clear understanding of the metadata to enrich the data with. Without metadata, it becomes difficult to find anything.

Q17: Are there any functionalities left that the organization wants to implement in the data governance system in the future? Which ones?

Answer category	Interview #	Explanation
Data lineage	I3	The wish of O3 was to also see the data lineage, but that was not yet possible in Azure Data Catalog.
	I5	Data lineage was a challenge at O5. There was not enough time to implement it properly, because so many other features were implemented too.
	I8	O8 wanted data lineage so that they could see where data originates from, and what kind of transformation steps it goes through. They wanted this to be easily visible for business users.
	I9	There are certain parts of the platform that cannot give insights into the data lineage. This can only be solved by programming these manually. This is unlikely to happen, as it is not cost-effective to do this. The out of the box functionality for data lineage is used and this is accepted as a nice functionality, but it would be better if there are more options for this.
	I10	Data lineage was very limited. For the Data Factory, lineage was quite straightforward. However, engineering was done in most cases with Databricks or Notebook, and this could not be easily seen in Purview. Data lineage therefore needs to evolve in Purview.
Data classification	I7	O7 has not yet implemented data classification, but it is on their roadmap.
	I8	I8 wanted to implement data rules to be able to identify what sensitive data is.
Data sharing	I4	At O4.3 they wanted to share data between different brands.
Data access control	I3	Functionalities that O3 still wanted to have were related to the policies. To be more particular, there is a need for giving access to data from a central place. What you currently need to do to get access is to send an email manually to request access. However, when you use the policy engine to send an email from Purview and have it automated with Power Apps, it saves a lot of work.
	I4	O4.2 wanted to apply policies and access control in the future.
	I9	What O3 is very excited for includes the management of access control from within Purview.
	I11	Two years from now, the catalog has metadata of the entire landscape, and then it is valuable to do access management through Purview. Someone could create a workflow in Purview requesting access for interesting data by sending a request to the data owner, which could be approved or rejected by the data owner depending on the profile of the user. In the current situation it takes about three to four stages to do this. Firstly, they do not know what data there is, so they do not know what to explore. Secondly, they do not know who the owner is. Thirdly, the owner does not have an easy way to give permissions, as an ID for a service request is needed for this.
Data privacy / policy enforcement	I2	As Purview was at the time of the interview a limited product, not all functionality was available and the organization started with what was available. However, due to the investments in Purview there was lots of future perspective. Their two year plan was to have most of the data governance functionalities, such as data quality, data privacy, and data access management.

Workflow management	I11	A lot of features O10 asked for are technically not possible. For example, implementing workflows.
---------------------	-----	--

Q18: If you compared this project with other projects in which a data governance platform (or another data management related tool) was introduced / implemented, what would you say was the main difference? Did these organizations have a very different approach on the implementation?

Answer category	Interview #	Explanation
Data governance maturity level	I3	In another Purview implementation project, that organization had more detailed plans about what they wanted. They also acknowledged the difference between metadata management and data quality, and linked this to data governance. They were more mature in their data governance than O3.
	I6	With projects at clients from the government sector, there was often a much higher understanding of data governance. You can talk about the architecture, design, and necessary transformations and they understand it. This was not the case at O6, as they only understood it from a very high level.
	I7	At O7, they did not have a good view in terms of the kind of platform they wanted to leverage. Other clients wanted lots of testing and PoCs to figure out which product they wanted to go with (e.g. based on costs, maturity of product, functionalities of product). O7 wanted to immediately try out Purview and give it a shot.
Visionary leadership	I2	In this organization, they already started the cloud migration and actions for data governance had to be taken immediately. What helped with this was visionary leadership, who were very proactive in carrying through the change.
Involvement from the business	I9	In another big data management program, the implementation was way more focused on the technical side. It was about data integration (e.g.. getting the definitions into the system, bulk loading). For the project of O3, there was a higher focus on the organizational side and the processes. It should be noted that the technical side is still crucial; if the system does not work, you cannot continue.
Type of data	I10	At O9, they worked with sensitive client data, and for this there were strong regulations that the organization needed to follow. Previous projects focused more on how to integrate systems to improve performance in organizations, with no sensitive information so the process was managed differently. Depending on the scope and the data, different actions needed to be taken. In this case, a data protection officer had to be introduced.
Willingness to change	I7	In O7 there was a high commitment to doing data governance. With a similar program at another client in the electric utility industry, they loved the value that such a data governance program could bring, but there was no commitment to doing the actual work. In O7, they had commitment, they had buy-in, and they had conversations about where they needed to go.
	I11	O10 was a lot more open to try out the system before requesting the end value. They recognized they needed a solution as they found it was challenging to understand what data they had, and they were willing to try to see how the system fitted in the organization. They were open to first using it as a pilot before going forward.

Q19: What is the current status of the project? Was the data governance system successfully implemented?

Answer category	Interview #	Explanation
Project was canceled	I4	O4.3 first needed to work on their governance and management structure.
	I6	The organization was not mature enough in their data governance understanding in order to scale the system to more data sources and did not continue with a full implementation.
Work in progress	I2	The Purview implementation (i.e. actual deliverable) just started at the time of the interview, so it was still in progress and it was too early to see any benefits. However, the consultancy part was already going on for a long time to help them make a decision about which changes to make.
	I7	As it is an ongoing effort, there is much more work to be done. Purview itself has been implemented, and the glossary and catalog have been completed. However, master data management is ongoing and the advisory part is to be continued.
	I8	At the time of the interview, the PoC was left on the Azure subscription for what it was with just the data catalog. There were other areas of development which had higher priority. However, the organization can come back to the PoC and enhance it if they want to continue with it. Purview was not the transformation, but rather the catalog to be able to get an overview of the data as a result of the data platform implementation.
	I9	The system is live and in production. However, it is now the time to look at its actual use. Whether the system is going to be used as it is intended remains to be seen. The implementation so far has been very technical.
	I10	The proof of concept was selected, and soon the implementation of the platform would start. There are 4 weeks for setting up the data catalog, getting all scans in place, and connecting all systems. After that, there are two months for taking decisions and actions for improving the organization.
	I11	The data platform was implemented at 14 BUs and lots of teams have been brought onboard. However, it is very much in progress. O10 was looking to slowly expand to more sources and to increase the spread of the scans. Purview is not only for Azure, and O10 wanted to also bring in data from SAP. SAP has large amounts of data, and data will never be moved to Azure as it will reside in SAP

Q20: How was the outcome / status of the data governance system measured?

Answer category	Interview #	Explanation
Users can easily find data	I3	The system was considered a success when the data was labeled, and when users could easily find the data they were looking for.
	I5	Top-level management wanted to have more clarity into what data comes in, and what data goes out.
	I9	The real success criteria was in the use of the system. Now findings can be gained about its actual use, so it could be decided whether it fits with the processes of the organization.
Positive attitude of business users towards the system	I1	To measure the success of the initiative, factors involving data quality and the feeling among business users could be used to identify whether the function was doing well.
	I2	Smaller milestones included a well functioning operating model, which meant

		stakeholders are happy with the result; they are working together, and processes are functioning correctly.
Delivery formalities	I9	There were some delivery formalities; for the implementation, definitions of done were established. This had to come from the product owner and any other related stakeholders.
Savings in costs	I10	Cost savings determine the success of the implementation, as the project was driven on cost.
Data issues are discovered for resolving	I10	O9 wanted to make sure data was compliant with regulations. Management will be happy to find out when there is data that is not compliant, so it can be resolved.
All / most data in platform	I5	It was measured whether all sources were onboarding and were mapped to the data catalog.
	I11	Data is currently being moved into the platform. Now everytime the data landscape is scanned, there are 1000 discovered entities while this was 0 at the beginning. The goal is to enrich the system with more data.
More efficient communication	I5	The goal was to have more efficient communication powered by a single view of all data.
Data from the system is used for analysis	I2	The main goal of the organization was to use the data to drive business analysis.
Correct use of the system	I2	A smaller goal is that the system is used correctly (e.g. getting a data hierarchy in place, label data correctly, classify data correctly, business can check the data glossary definitions).

Q21: What benefits did the data governance system bring to the organization?

Answer category	Interview #	Explanation
Not measurable yet	I7	It is a little too early to see benefits. For now there are five to six analytical models and that is a small set to draw a conclusion from. Also, there are different needs at both project level and overall program level, but so far feedback from management had been positive.
	I11	Benefits are not measurable yet. It will be at least a year before benefits can be seen. First, the response has to be seen and the system has to be used properly. However, there had been demos with three teams, and the teams were very excited. For one team specifically who is responsible for acquiring other organizations it is going to be valuable, to see what they are bringing into the data landscape and how they plan to integrate and manage this.

Q22: Which factors in the organization had the biggest influence on the outcome / status of the project?

Answer category	Interview #	Explanation
Support from leadership	I1	An important factor that contributed to the initiative was leadership. It was important that leadership was updated on why governance is important, and how they should do data governance.
	I2	The leaders at O2 were visionary, which added to the success of the process.

	I3	The Chief Data Officer was very clear in communicating what needed to happen and when.
	I5	A key factor in overcoming the challenges was a mandate from top-level management.
Good sponsorship	I5	Although you can have good leadership, if they do not sponsor the project then you can never get it done.
	I7	The funding helped to make it successful.
	I9	It was important that the team had the capacity and the budget to be able to work on it.
Commitment to change from stakeholders	I1	A lot of the initiative had to do with people and change management.
	I7	All stakeholders were on the same page in O7. The project started with someone from IT, but the business side was quickly on the same path. At O7, there was a high commitment to do the work necessary for good data governance.
	I9	One of the biggest influences on the outcome of the project was the support from stakeholders. Purview could be very complex and require a lot of work, and next to this the data platform also had to be kept running. If the stakeholders did not pay any attention to some tasks, then it would disappear from the backlog.
	I11	Whether the people in the organization were open to follow the plan to achieve the strategy was an important factor.
Extent to which data landscape is covered	I4	At O4.2, the implementation of the MDM platform looked at all data that was scanned by Purview, correlated the data, and gave the options to data owners. The more data was scanned by Purview, the more useful it was.
	I10	It was important to try and include as many sources as possible so most of the data landscape is covered. If for example only one data lake is implemented, you do not get a big overview. The more systems were covered, the better decisions could be made.
Clear data strategy	I1	Data governance required a bit of bureaucratic process with guidance from the top, in order to show the importance of why the organization has to do it.
	I11	It was important that O10 was seeing beyond the immediate things and had a clear long-term vision.
Ease of use of tool	I3	The data governance tool was easy to use. The data catalog has lots of technical solutions for adding data, but it is also very easy for the user to add something. Other tools are more focused on technical details, while Purview is very focused on a business perspective which makes it easier for a business user. The downside about this is that sometimes you have to find workarounds for technical issues.
Linking it to other transformation efforts	I2	The implementation of Purview was part of an overall program to help migrate the existing enterprise system to the Cloud. This made it easier to introduce Purview, as they were already making changes in the organization.
	I5	If you do data governance for the sake of data governance, it is not that effective. But if you couple it with another transformation effort to see the urgency of a data governance implementation, you have more buy-in and more drive. This was the case for O5 when the new way of working was introduced.

--	--	--

Q23: Is there something that you want to add to this topic, that we have not discussed yet?

Answer category	Interview #	Explanation
Data governance maturity	I1	The minimum requirement before an organization starts looking at a platform is that a data governance framework should be there. By this, it is meant that there should be clear ownership of data within different business groups, there should be a definition of who owns the data and who are the stewards and there should be a decision-making party that is able to take data initiatives around data governance. Data governance solutions require a lot of advisory and proof of concepts (PoC), and only the last step is tools. In the practical world, tools are easily available, and it is something that provides a lot of benefits. Sometimes a platform is up and running before a data governance framework is set up. Benefits that are gained from this platform are then on a level that could have been way better if there was some kind of governance body.
	I4	The choice for a data governance tool is related to the level of data understanding within the organization. Before a tool can be implemented, the organization should understand its data and put ownership in place. Governance processes should be implemented around who uses the data, and where it comes from. An organization which does not understand their data and wants a tool to tell them what they got, is bound to fail. Applying a technical tool to a lack of data understanding, only results in costs and confusion, not in a gain in business value. An indicator of a right level of maturity in the organization is if the organization has some sort of Chief Data Officer (CDO), as then there is commitment from the board to drive the organization in the right direction. Moreover, it is not necessarily about the size of the organization, but more about the data architecture and the extent of the data sources. In order to start seeing benefits of a data governance tool, there needs to be at least about five to seven complicated data sources, which can be scaled. Another option is to use a data swamp or data lake which is controlled by some kind of master data management, before it is registered in Purview. Furthermore, there needs to be established agreements between parts of the organization that own data. A mature organization is one that understands data is owned by the organization, not by a particular team, and has agreements about sharing data in place. Good governance is a political negotiation, as it needs to break down barriers to look at data from an organizational perspective. It is not about central control because you do not have that in a federated organization, but a highly centralized or controlled organization is probably successful because they have a good idea where their data is, but they need a tool to manage this. Some general trends can also be said about the sector. Product companies tend to be most mature regarding their readiness for a data governance tool, not necessarily the financial sector. The public sector is also very mature in their data understanding. However, a lot of personal identification (PI) data is handled in this sector, and while Purview may help, a great deal of data is in silos so it is difficult to scan on premise.
	I6	Most clients want to establish the glossary terms, because of their unfamiliarity with data governance and they do not know where to go from there. After the glossary is established, organizations can start adding policies, adding rules, tracking lineage, connecting to PowerBI, etc. The more the client is familiar with the tool, the more they ask.
	I8	If there is a small project, the organization probably does not need a data catalog because it can still have a good overview without it. This also applies if the organization has a single data warehouse, because everything is already centralized there. However, for bigger projects with lots of data involved, a tool

		like Purview becomes more critical. This does not have to replace the data warehouse. A data warehouse is still good for ensuring data quality and it is easy to use. However, tools like Purview can be used to get an overview of the data next to the data warehouse.
Pillars of data governance	I1	There are six pillars for data governance platforms: data catalog, data lineage, data sharing (with other organizations), data quality, master data management (MDM), and data use. Whenever you talk about data governance, you talk about all perspectives (including data management).
	I2	Data governance focuses on policies and rules, whereas data management is about using the policies and rules for managing the data.
	I4	Data governance is about the organization and its people, not about tools. It is about organizational structure and organizations understanding the value of data. To make data governance successful, you need an organizational change. Data governance focuses on ownership, control and agreements. Regarding ownership, someone from the business should own data, not IT.
	I6	Changes needed to be made in organizations when implementing a data governance platform should be approached from a people, process, and technology standpoint. Technology is about where the sources come from. People is about who is managing data and who should get access, which can be discovered by establishing personas. The process is about what are the terms, what should be in the glossary, and are the definitions going to be.
	I7	Data governance is not just a technological solution, but it is a connection between three parts; people, processes, and technology. Advisory is a big part of the implementation.
	I8	Data governance is about control (who has access to data and with what restrictions) and how to approach personal information (PI) data.
	I10	There are three pillars of a good data governance strategy; technology (e.g. Purview), people (e.g. data owners) and processes (what has to be done). The organization relies on all three. Technology is helpful, but without the others the strategy could fail.
Combination with other data management tools	I1	Another stream of work that is happening is customer data platforms (CDP). This works the same as MDM, but only on customer data. Many customers are going in the CDP direction, and an example of this is Microsoft Customer Insights.
	I4	Data governance is all around business owners. Purview does not give that functionality, so it is useful to have other master data management (MDM) tools, e.g. Profisee or Boomi. This can provide a good workflow to look at data from a domain perspective. Some clients just go with MDM, and not with Purview. This can also work, but it does not give future ability to do both things.
	I6	A client is very mature in their understanding of data management if they start thinking about a master data management (MDM) tool. A MDM platform (e.g. Profisee) can be complementary to a data governance platform (e.g. Purview). With Purview you establish rules and policies, but with a MDM platform these are enforced. Most value can be gained if they work side to side together, and this also provides opportunities for scaling. Regarding the success rate for either establishing Purview or Profisee first; I6 had seen both at clients, but did not see a great difference in its success. The order also depends on how fast the client wants to scale.
	I8	I8 noted that MDM is not a prerequisite for Purview to be successful. An MDM

		<p>solution is useful if there are many sources and data needs to be confirmed to ensure data quality. However, in some scenarios Purview is enough. Purview does not import data, it just imports metadata so you know what data there is. Data will still be wherever it is (e.g. on-premise, in Cloud), so you need other systems for managing it. Purview only gives a label that says what data it is and what it is connected to, but not the data itself, otherwise it would explode.</p>
	I10	<p>It depends on the situation whether a master data management (MDM) tool is necessary for data governance. There are two types of organizations. Firstly, an organization just starting with their journey to the Cloud probably does not have many issues with data management, and does not necessarily require a tool. Secondly, an organization that already moved their workloads to the Cloud and are advanced in IT probably have data issues. For them, it is strongly recommended to have Purview and a MDM tool. Data governance should be implemented before a MDM tool, as first needs to be understood what data the organization has, before it can be improved and changed. Although data governance and MDM are linked, you do not necessarily need both (but will probably get the best value with both).</p>
Complexity of data governance	I1	<p>The challenging part about data governance is that it is not a project you start and finish. People, organizations, and leadership change, and the data governance framework should be strong enough to adopt as these changes are happening. There needs to be strong principles and roles. There is not a single set of solutions for this, as there are always different scenarios which require different solutions. Everyone wants the outcomes of data governance, but they often do not want to do the small tasks to make it happen.</p>
	I4	<p>I4 said that it was difficult to say for which client a data governance platform works. It varies massively per client, and it depends also on their history.</p>
	I11	<p>I11 noted that data governance is much more than a data catalog, it is also about access management, making sure there is a clear understanding of how data is managed, and compliance (e.g. with GDPR). There were multiple discussions at O10 for following best practices about access management such as the principle of least privilege (i.e. not everyone has access to data).</p>
Thinking about data as a strategic asset / product	I4	<p>Organizations should start thinking about data as a product. Data should be governed the same as any product creation. When governance is managed around data domains, it proves to be successful. Domain subject matter experts manage the data in the domain. Bits of each domain are used to get to the data product. The product owners at the end are responsible for these data products. The value of the data should be understood at both domain level (interim level) as well as product level in order evaluate whether the data is used in the right, cost-efficient way and that they are handled properly with the right access. The business needs to understand how much waste goes into the product line. The focus on data organized around domains is sometimes referred to as a data mesh or a data fabric approach.</p>
Importance of data governance tools	I2	<p>Data governance is a very loosely used term. Some consider data governance as having a few processes or policies in place. However, when looking beyond what is necessary now, it is necessary because of the growing volumes of data to have data governance in place, as data governance is gaining a far bigger scope. This is the consumerism of data governance; earlier only the CDO and a few analysts were using data, but currently the user group has become much broader and also roles such as privacy officers and information managers want to use data governance processes. Earlier on data was not that distributed, contrary to now where data is coming from many different processes. This makes it more relevant to have a solution for gaining a central overview.</p>
	I4	<p>Data governance is about organizational understanding and data maturity. There needs to be an idea about who uses the data, and where it is coming</p>

		from. It can be a person or a tool who gains insights in this. You can automate it, but generally a person is needed as there needs to be a conscious decision about who manages the data.
	I8	In order to address the new problem of trying to find the right data due to the explosion in volumes of data, tools like Purview can be used. Many companies now see the value of bringing the data together, especially doing it in the Cloud which is cheaper than in the past and enables more processing capability.
	I11	I11 indicated that a data governance tool is necessary for organizations to have in place. You cannot scale if data governance only takes place through manual effort. Some kind of automation is needed to realize this on a big scale, specifically with the current context in which so much is happening in the Cloud. Having a standardized platform, having a data catalog, defining and automating roles & responsibilities is key to making sure the organization's vision can be realized.
Product choice for Purview	I1	A challenge for the choice in a platform is that there are so many options available. Purview does not have some features like data quality, MDM, and interest validation. However, Microsoft does lots of marketing and investments around Purview to make sure it stays updated and better than other offerings. In the future, Purview will likely be the go-to tool for data catalogs and data lineage. When starting from scratch, Purview is the best choice. If there are already tools in place from the other vendors, then it is more likely to consider tools from these vendors for data governance.
	I2	This organization quickly made the decision for Azure as they were already on their Azure journey. However, when organizations have to choose which cloud to use, then the choice for a product becomes more difficult. When an organization already has data governance in place, it might be difficult to sell Purview to. However, if that organization decides to migrate to the Cloud, their existing tool might lack that particular functionality for the Cloud, and that is where Purview comes in. However, even though Purview's scope is expanding, Purview still has a long way to go for example regarding policies.
	I4	Purview gives organizations the ability to manage some of the data from a business perspective. It has a metadata engine, and allows you to discover data in the form of a data dictionary. However, it can lure clients into false sense of security that they got all the data. Purview cannot look at everything as it does not have that many scanners. Although there are API's and you can write your own scanners, it is not the same as having these scanner connections already in place. When Microsoft is able to scan most of the operational stores at clients, then it becomes a much more attractive tool. Moreover, Purview acts as a middle piece and enables good governance to a point, but it does not do good stakeholder management around data. As Purview matures, it will become more useful, specifically as a data catalog and dictionary for data platforms. What would be useful to have in Purview are functionalities for applying policies (regarding security and personal information (PI) data) and options for sharing and modeling data in Purview, which are more master data management functionalities. The choice for Purview is mainly based on its future potential as Microsoft is investing a great deal of money in it, and because it is a cloud native platform.
	I8	Areas in which Purview can improve is the ability to create custom properties to enrich metadata with, an improved user interface (UI) for business users to increase usability, more functionality like data quality and rules to verify if data meets certain quality requirement, more connectors to get more data sources in, and better statistics about how often data is used. However, other solutions are not geared towards Azure and Purview is very competitive from a pricing point of view.

Evolution of data governance	I2	Recently with the explosion in volumes of data, the focus of organizations is changing. Not only do they want to build analytical capabilities to see how the business is doing, but they also want to have a better structure and management of the data to unlock insights from it. As the volume of data is growing fastly, organizations should take a forward approach in governing it.
	I8	In the past there were very traditional structured data warehouses, which meant that there was less need for data governance. Only now that there is an enormous amount of data coming from everywhere in every form with lots of data silos that are spread, there is more of a need for systems with data catalogs and glossaries in order to understand what data there is. This means a fast and sudden shift to the data governance domain. Instead of preparing data upfront and then importing it in a structured data warehouse, there is a new paradigm of getting all data together and importing it cheaply into a data lake. The effort now comes afterwards if the organization wants to do something with this data. This explains why there is a sudden need to be able to find the data, as in the first structured model data was already easy to find.

Appendix G: Framework validation expert interview questions

Part A

For case study:

Background interviewee

- Q1: What is your role within the organization?
- Q2: What is your experience with data governance?
- Q3: Have you worked on other data platform related projects besides the Purview implementation?
- Q4: Does this data governance implementation project differ a lot from other data related projects you have experienced? If so, in what way?

Background data governance platform implementation project

- Q5: What were the most important issues in the organization that needed to be solved by the data governance platform?
- Q6: Which functionalities did the organization want to have in the data governance platform?
- Q7: What was the most challenging part of the implementation of the data governance platform?
- Q8: What is the current status of the data governance platform program?
- Q9: What is the attitude of people in the organization towards the use of the platform?

For academics:

Background interviewee

- Q1: Can you tell me more about your research background?
- Q2: How is your background connected to data governance?
- Q3: How much experience do you have with data governance?
- Q4: Besides your theoretical experience, do you have practical experience with data governance or related data management (solutions / systems) at organizations?

Background theory on data governance (platform) implementation

- Q5: What are the main problems within organizations surrounding data governance?
- Q6: How can these problems be solved?
- Q7: What is necessary for organizations to have in place in order to establish good data governance?
- Q8: What is the role of systems for establishing good data governance?
- Q9: To what extent do you think that you can apply frameworks or models from theory to practice when a new system is implemented in an organization?

Part B

Testing success & situational factors (link to Google forms: <https://forms.gle/EJqnwVhQmHwqMSuJ6>)

- Q1: From literature and the expert interviews, a list of success factors have been derived that are necessary to have in place for a successful implementation of a data governance

platform. For the following success factors, could you rank them on a scale from 1 to 5, according to how important they are for a data governance platform implementation?

1. *Make sure there is a direct integration between the data source and the data governance platform so data sources can easily be imported in the system.*
 2. *Assign roles and responsibilities to data.*
 3. *Get the data catalog and glossary in the system first before adding other functionalities.*
 4. *Create a central and unified set of data definitions.*
 5. *Set up policies that ensure the right data access and use.*
 6. *Classify a few sources successfully first before scaling to more data sources.*
 7. *Have high support from top-level management to guide the implementation (top-down).*
 8. *Create personas and use cases to determine what should be included in the system.*
 9. *Have visionary leadership that clearly communicates the data strategy, vision, and intended outcome across all departments.*
 10. *Have a high understanding of the origin of the data.*
 11. *Create an open culture which encourages embracing change.*
 12. *Have support from a sponsor that is willing to invest enough resources until the outcome is reached.*
 13. *Connect the system to a central data platform that first standardizes data instead of the original data source.*
 14. *Create data architecture where data is treated as a product around a data domain (i.e. data mesh approach).*
 15. *Combine platform with data quality management system or functionality to manage the quality of data.*
 16. *Implement a pilot version / proof of concept to demonstrate the platform's use for the business.*
 17. *Include as many sources as possible in order to cover the whole data landscape.*
 18. *Set up a data governance committee that manages the implementation from a central point.*
- Q2: Besides the success factors, some situational factors have also been derived from literature and expert interviews. These are characteristics from the organization that might have an effect on the implementation of a data governance platform. Could you rank the following situational factors on a scale from 1 to 5, according to how much effect they have on an implementation of data governance (platforms)?
 1. *Data governance maturity level.*
 2. *Sector.*
 3. *Goal.*
 4. *Hierarchy / division.*
 5. *Systems infrastructure (e.g. Cloud, on-premise).*
 6. *Type of data.*
 7. *Number of data sources.*
 8. *Complexity of data sources (e.g. structured, unstructured).*
 9. *Culture.*

Part C

Testing framework (link to Google forms: <https://forms.gle/YG7HadXJbRYAYiTH6>)

- Q1: A framework has been designed which specifies the stages / characteristics of a data governance platform implementation. In which stage / quadrant would you place the organization?
- Q2: For the following evaluation criteria, could you give an indication on a scale from 1 to 5, according to how much this applies to the framework?
 1. Clarity of guidelines / success factors: to what extent are the guidelines and success factors clear
 2. Clarity of relationships: to what extent are the relations between concepts clear
 3. Accuracy of categories: to what extent are the guidelines / success factors accurate for the chosen classification
 4. Understandability: to what extent is the framework understandable
 5. Applicability: to what extent can it be applied to the organization
 6. Usability: to what extent it uses relevant constructs
 7. Robustness: to what extent does it handle exceptional situations
 8. Generalizability: to what extent can it be generalized to other organizations
- Q3: Do you have any suggestions for improvement of the framework?

Appendix H: Survey results validation interviews

In this Appendix, the results are given from the survey in which the interview participants could rate the factors and criteria on a 5-point Likert scale.

Part B

Success factors

In table 16, the results from the survey are shown. The participants were asked to rate each of the factors based on how important they are for the implementation of a data governance platform, with 1 indicating not very important, and 5 indicating very important.

Table 16. Survey results for the importance of the success factors

Factor number	Factor	Interviewee					
		I1.1	I1.2	I2.1	I2.2	I3.1	I3.2
1	Create a central and unified set of data definitions used in all organizational domains	1	4	1	5	5	4
2	Assign roles and responsibilities to data	5	1	1	5	5	5
3	Get the data catalog and glossary in the system first before adding other functionalities	4	5	4	5	5	3
4	Set up policies that ensure the right data access and use	5	4	4	5	5	5
5	Classify a few sources successfully first before scaling to more sources	5	4	5	5	5	3
6	Have high support from top-level management to guide the change (top-down)	5	4	4	5	5	3
7	Create personas and use cases to determine what should be included in the system	4	5	4	4	5	5
8	Have visionary leadership that clearly communicates the data strategy, vision, and intended outcome across all departments	3	5	4	5	5	5
9	Have a high understanding of the origin of the data	2	5	3	5	5	4
10	Create an open culture which encourages embracing change	3	2	3	5	5	3
11	Have support from a sponsor that is willing to invest enough resources until the outcome is reached	5		5	5	5	2
12	Connect the system to a central data platform that first standardizes data instead of the original data source	3		2	4	5	2
13	Create data architecture where data is treated as a product around a data domain (i.e. data mesh approach)			1	5	5	
14	Combine platform with data quality management system or functionality to manage the quality of data	1	3		5	5	4
15	Implement pilot version / proof of concept first to demonstrate the platform's use for the business	4	2	5	5	5	2
16	Include as many sources as possible in order to cover the whole data landscape	1	1	1	3	4	3
17	Set up a data governance committee that manages the changes from a central point	4	3	5	3	5	5
18	Make sure there is a direct integration between the data source and the data governance platform so data sources can easily be imported in the system	5	5	2	5	4	5

Situational factors

In table 17, the survey results for the situational factors are shown. The participants were asked to rate on a 5-point Likert scale how much the situational factor has an effect on the implementation of a data governance platform, with 1 indicating not influential and 5 indicating very influential.

Table 17. Survey results for the effect of the situational factors

Factor number	Factor	Interviewee					
		I1.1	I1.2	I2.1	I2.2	I3.1	I3.2
1	Data governance maturity level	4	5	5	5	5	5
2	Sector	5	5	1	5	5	4
3	Goal	4	5	5	5	5	5
4	Division (e.g. central / decentral)	2	1	2	5	5	2
5	Systems infrastructure				5	4	
6	Type of data (e.g. sensitive, business critical)	4	2	4	5	5	5
7	Number of data sources	2	2	1	5	4	2
8	Complexity of data sources	4	4	1	5	4	4
9	Culture	4	3	3	5	5	4

Part C

Framework

In table 18, the survey results for the framework criteria are given. The participants could rank each criteria on a 5-point Likert scale, with 1 indicating the framework could use improvement on this criteria and 5 indicating that the framework was good in this criteria. The column that ends with A belongs to a score given to framework A, and likewise the column ending with B belongs to the score for framework B. Interviewee I2.2 and I3.2 were not given this survey, as they were knowledgeable on data governance and the implementation of new systems, but not specific data governance platforms in practical situations.

Table 18. Survey results for the framework criteria scores

Criteria number	Criteria	Interviewee									
		I1.1A	I1.1B	I1.2A	I1.2B	I2.1A	I2.1B	I2.2	I3.1A	I3.1B	I3.2
1	Clarity of concepts	3	4	4	4	4	3		5	4	
2	Clarity of relationships	4	3		2	4	2		4	4	
3	Accuracy of classifications	4	3	3	2	1	1		3	4	
4	Understandability	4	4		1	3	1		4	3	
5	Applicability	2	4	3	4	3	3		4	4	
6	Usability	3	4	3	2	3	2		4	5	
7	Robustness	3	4	4	1	2	3		3	4	
8	Generalizability	4	4	4	3	4	3		4	4	

Appendix I: Visual results of validation interviews

This Appendix visualizes the survey results in bar charts. For the success factors, situational factors, and evaluation criteria, the average scores are given. Besides that, for the success factors and the situational factors, a distinction is also made between the scores given by practitioners from the case study organizations and the scores given by the academics. For the evaluation criteria, a distinction is made between the different iterations. 0 indicated the factor / criteria was not important, and 5 indicated the factor / criteria was very important.

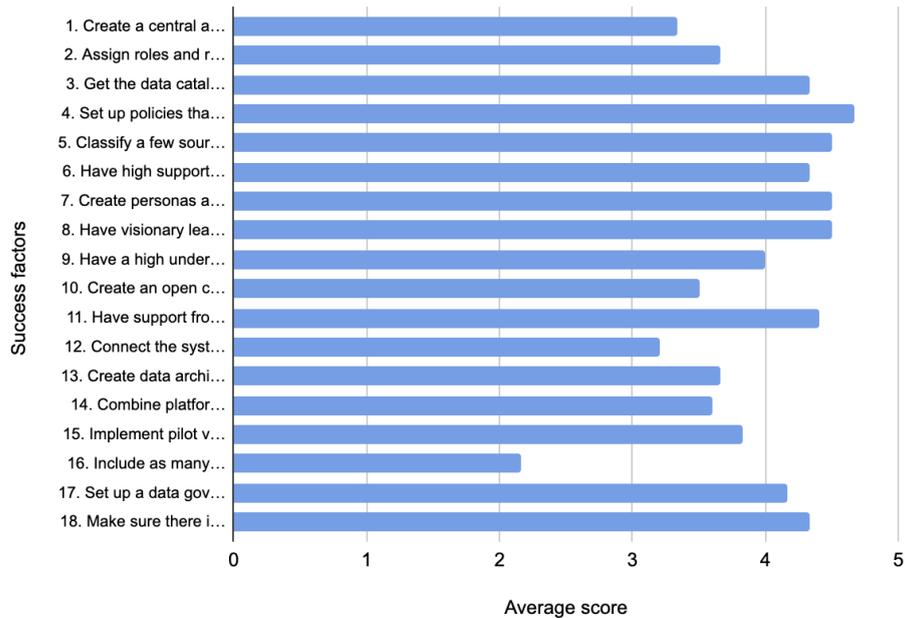


Figure 16. Average scores for the data governance platform implementation success factors

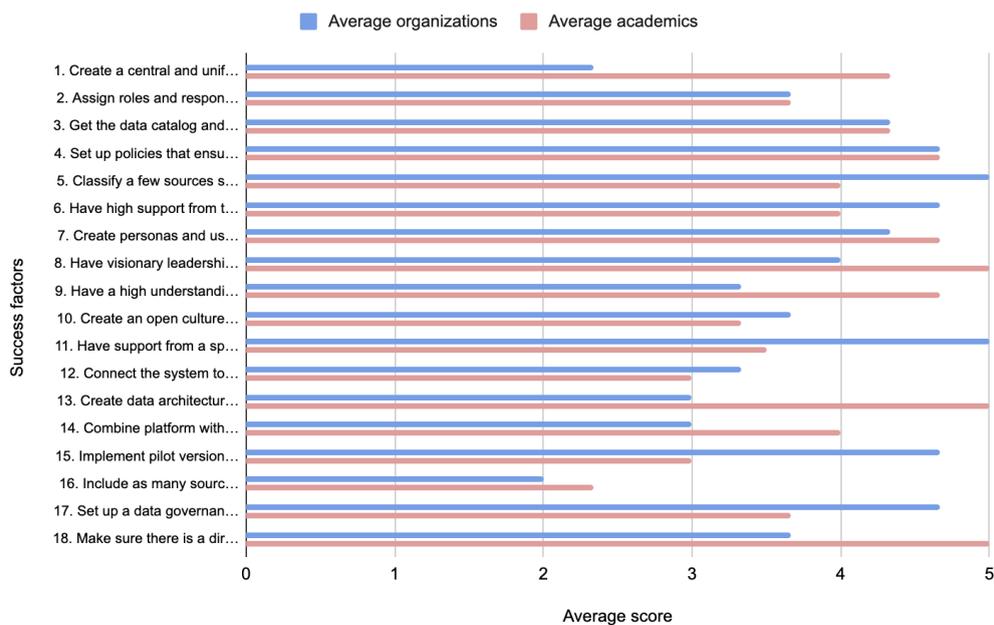


Figure 17. Average scores for the data governance platform implementation success factors with a distinction between practitioners and academics

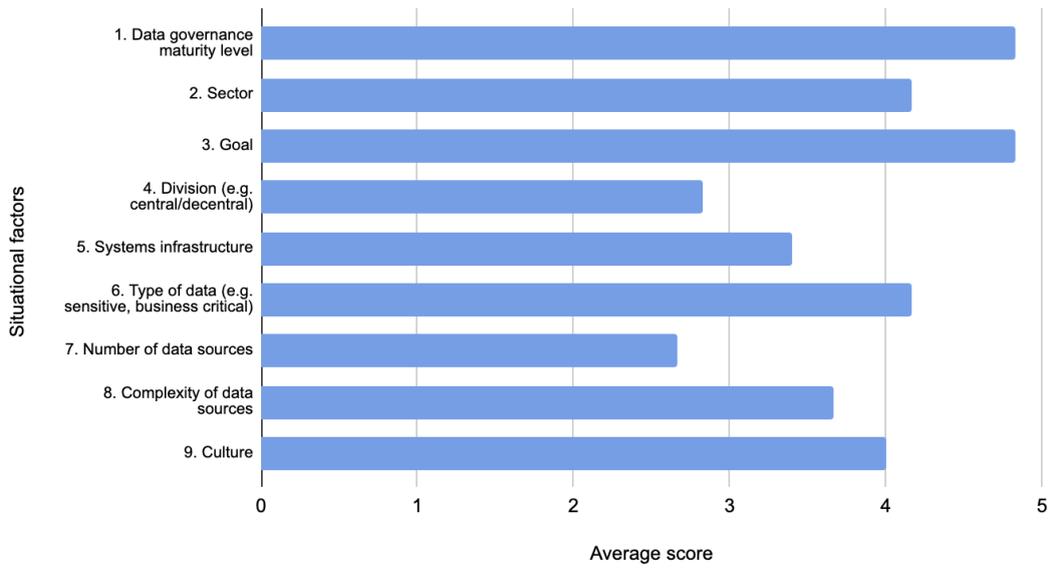


Figure 18. Average scores for the data governance platform implementation situational factors

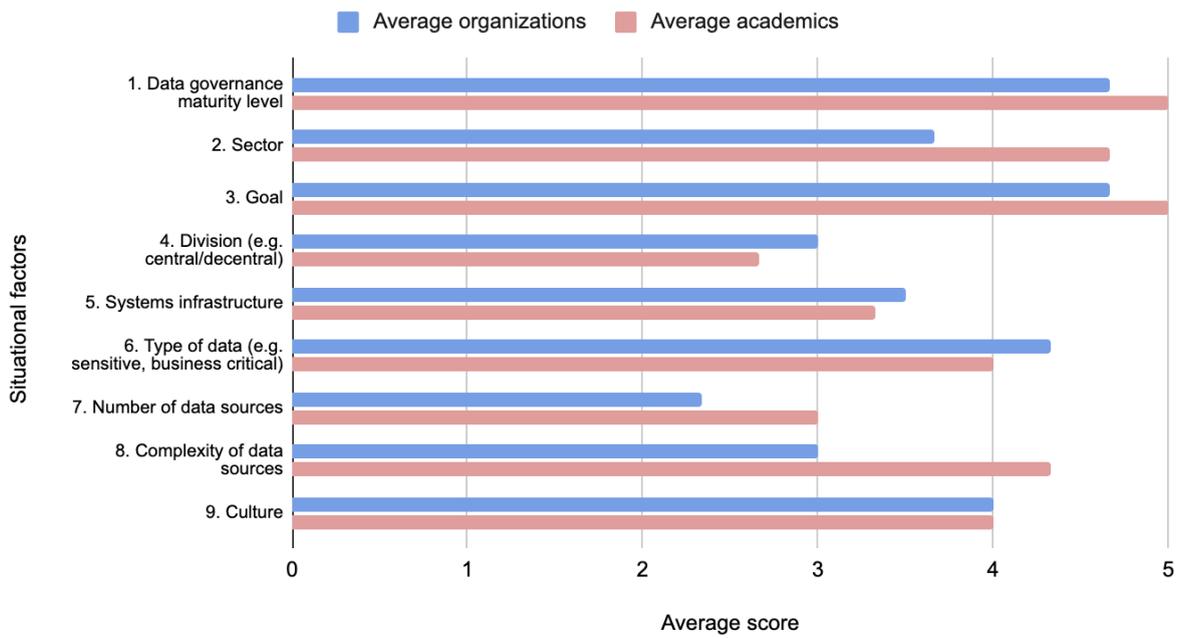


Figure 19. Average scores for the data governance platform implementation situational factors with a distinction between practitioners and academics

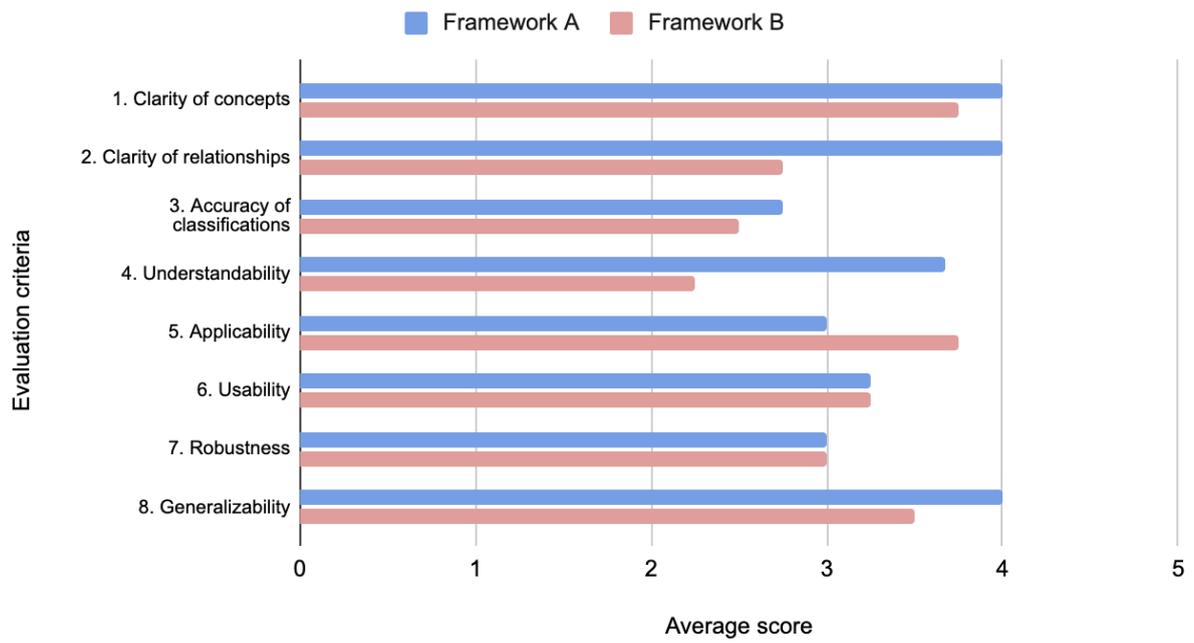


Figure 20. Average scores for the evaluation criteria for framework A and framework B

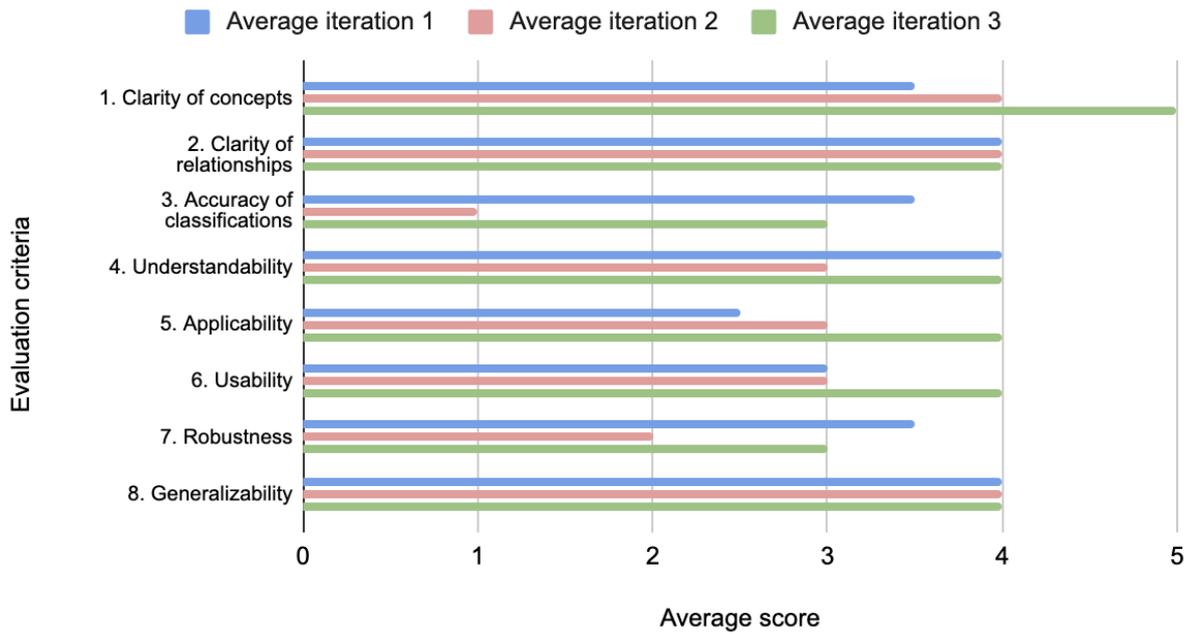


Figure 21. Average scores for the evaluation criteria for framework A with a distinction between the iterations

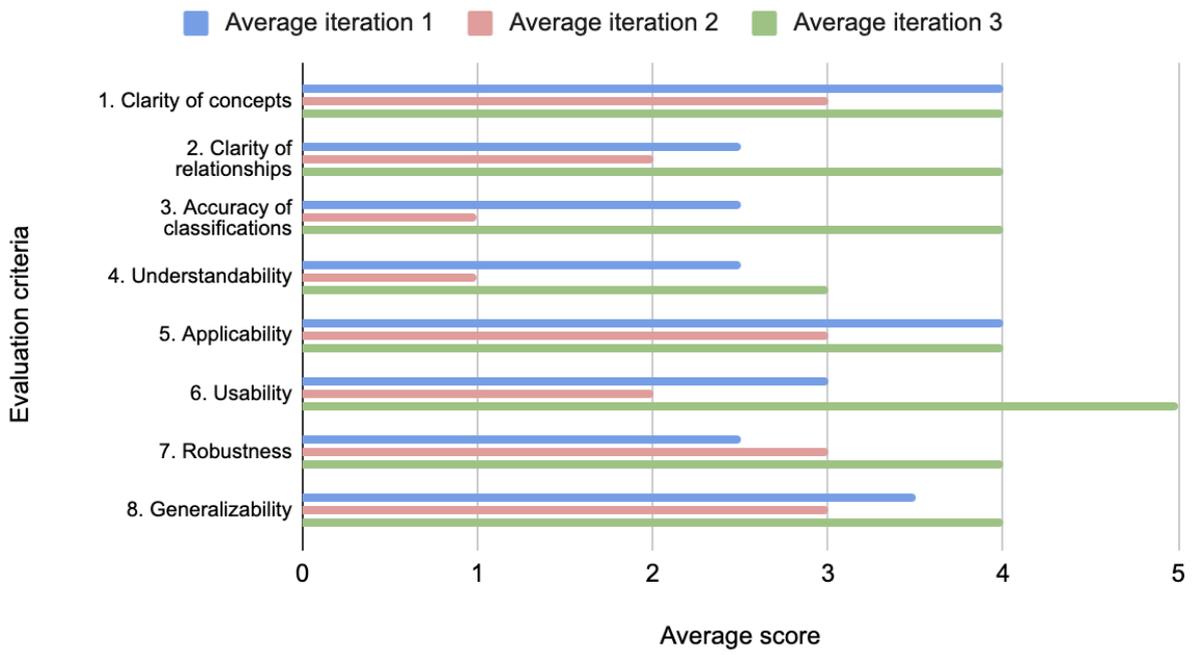


Figure 22. Average scores for the evaluation criteria for framework B with a distinction between the iterations

Appendix J: Points of improvements from validation interviews

The tables below show a summary of the qualitative answers provided during the interviews. For part B, the explanations for why an interviewee gave a certain score are shown. Similarly, for part C, the explanations for why the evaluation criteria scored a certain way are given. These also include points of improvement for the frameworks. If there are no comments to a factor / evaluation criteria, the interviewee did not provide a qualitative comment about this score.

Part B

Success factors

In table 19, the comments about scores given for success factors are shown.

Table 19. Explanations from interviews of scores given for success factors

Success factors	Interview #	Explanation
1. Create a central and unified set of data definitions	I1.1	Should be established in Purview.
	I3.2	Not for different domains, but within the same domain this is very important.
2. Assign roles and responsibilities to data	I1.1	This was highly desirable before Purview was introduced.
	I1.2	Roles and responsibilities are for a person and not data. Data might have user preferences, metadata, access rights and privileges assigned to it.
	I2.1	You can assign this to users, but not to data. Besides, it is valuable to pick it up piece by piece, so this should not apply to all data. It is a pitfall to try to apply it all at once.
3. Get the data catalog and glossary in the system first before adding other functionalities		
4. Set up policies that ensure the right data access and use	I1.1	Can release Purview if access is not yet determined but that depends on what is scanned in Purview.
	I1.2	Maybe not the most important for a successful implementation, but for compliance, risk management and security it is very important.
5. Classify a few sources successfully first before scaling to more sources		
6. Have high support from top-level management to guide the change (top-down)	I1.2	There are other ways to do it, but organizational change is essential.
7. Create personas and use cases to determine what should be included in the system		
8. Have visionary leadership that clearly communicates the data strategy, vision, and intended outcome across all	I1.1	Communication is important, but attention should be paid to what is communicated in case the wrong information is brought out.

departments	I1.2	You do not need visionary leadership, but communication is essential.
	I3.1	O11 recognizes that data is very important and data functions are assigned, but decentral data strategies are made (leadership does not do this).
	I3.2	Important to convince people to actually use the system.
9. Have a high understanding of the origin of the data	I1.2	Essential for good data quality, good data oversight and data quality.
	I2.1	No one has an idea where every data in the organization is; it should always be related to a business problem.
10. Create an open culture which encourages embracing change	I1.1	O3 did not yet have trainings in place for learning how to use the new system.
	I1.2	A hierarchical culture can also be very successful.
11. Have support from a sponsor that is willing to invest enough resources until the outcome is reached	I2.2	This weighs up against people's willingness to change; your money for the project can run low while still achieving change in the end.
	I3.2	With this you can endlessly allocate a budget even if the system is not usable, which is not good.
12. Connect the system to a central data platform that first standardizes data instead of the original data source	I1.1	O3 did this with Purview, but it is not necessarily a prerequisite.
	I2.2	Depends on what you want to do with the data.
	I3.2	Raw data can also be interesting.
13. Create data architecture where data is treated as a product around a data domain (i.e. data mesh approach)		
14. Combine platform with data quality management system or functionality to manage the quality of data	I1.1	Not necessarily required as catalog is sufficient, but it would be a nice addition.
	I1.2	Not necessary but it is a good idea.
	I2.1	A data quality functionality can be part of the data governance system, so this should not be necessarily seen as something separate.
15. Implement pilot version / proof of concept first to demonstrate the platform's use for the business	I1.1	A pilot was introduced for the data governance team, but not a pilot for the whole organization. However, they started small in Purview and did some demos to gather feedback.
	I1.2	It is not necessary, as it depends on the complexity of the system.
	I3.1	It is important to start small first in order to align everyone.
	I3.2	If the needs of the end users are considered in the design, then a pilot version is not necessary.
16. Include as many sources as possible in order to cover the whole data	I1.1	Start small, but grow big.

landscape	I1.2	You can have a good data platform with a few specific sources.
	I3.2	Not about as many as possible, but more about that it is usable.
17. Set up a data governance committee that manages the changes from a central point	I1.1	This was the case for the Purview implementation, but it relied heavily on a few people.
	I1.2	There should be oversight, but there are ways to do it in a different way.
	I2.2	You can have different areas in the organization that manage data in a different way.
	I3.1	The past few years, there was a high focus in the organization on decentrality, as the aim was to have more distributed systems. This conflicts with central leadership.
18. Make sure there is a direct integration between the data source and the data governance platform so data sources can easily be imported in the system	I2.1	Purview is source based (e.g. data sources, connectors) so terminology is very Purview focused. Talking about a source usually means IT deals with it. It is better to talk about entities regardless of source; for example, customer data is an asset..
	I2.2	Take for example a social media platform like Twitter; it is very important that the data can be integrated easily for this.
	I3.2	It is very important to import data sources.

Situational factors

In table 20, the interview responses about the situational factors are given.

Table 20. Explanations from interviews of scores given for situational factors

Situational factor	Interview #	Explanation
1. Data governance maturity level	I1.1	There is no use to introduce tooling if the organization has a low level of maturity.
	I2.1	Most important factor, because if you are a starter there are so many other things you have to look at rather than if you are in the middle.
	I3.1	Extremely important, if you are on a level 1 then you need to take very different steps before implementing something.
2. Sector	I1.2	If there are a lot of regulations than sector matters, as lots of sophistication is required. Because of an organization's history it should not matter, but often in practicality it seems to matter.
	I3.2	Important because it has something to do with standards that are in place in the industry.
3. Goal	I2.1	Data governance platforms can be unnecessary, as data governance mostly revolves around complexity and culture. You can use Excel to organize data governance. A system has to support processes and this depends on what it is going to be used for. Sometimes ten people are needed to maintain the system and the focus goes away from the problem it is intended to solve.
4. Division (e.g. central / decentral)	I1.2	Whether a company is hierarchical or flat does not matter a lot, as an open culture can also work.

	I2.1	With a decentral organization it means there is more room for different definitions, systems, or quality requirements, while with a central organization there is one approach.
	I3.2	It does not really matter whether it is central or decentral for the success of the implementation.
5. Systems infrastructure		
6. Type of data (e.g. sensitive, business critical)	I1.1	It is an important factor for thinking about a data catalog, but it was not necessarily looked at during the implementation.
	I1.2	This may come into play if there are regulatory requirements, but with high maturity it should not matter as much.
7. Number of data sources	I1.1	O3 started with only one data source, the enterprise data warehouse, but all systems leverage data to this system.
	I3.2	This does not have to be important, it is more about usable data than the number of data.
8. Complexity of data sources	I1.1	How many attributes the data has and how many data consumers there are is more important than the number of sources.
9. Culture	I1.1	O3 is a pragmatic organization that wants to have things as concrete as possible, so if something works fine, employees wonder why something different has to be implemented. Change management is important here, but not determining.
	I2.1	For an organization with a military regime, it is easier to introduce a data governance platform than with an organization where everyone has an opinion.

Part C

Frameworks

In table 21, the comments given on the evaluation criteria for framework A are shown.

Table 21. Explanations from interviews of scores given for evaluation criteria for Framework A

Evaluation criteria	Interview #	Explanation
1. Clarity of concepts	I1.2	There is not a lot of detail.
2. Clarity of relationships	I1.1	Conceptually, it is clear, but more guidance is preferred about how to go to the next step.
	I2.1	Now it looks like you do everything all at once, but you need to take it up piece by piece. Make it clear that you can repeat certain stages.
	I3.1	There should be a list of actions which have to be executed to go from one level to a higher maturity level.
3. Accuracy of classifications	I1.1	O3 could be classified between the first and second stage, as it now focuses on the glossary. A glossary can do a lot; masking, access rights, data quality. Per data catalog there should be such a diagram or make it even more high-level.
	I1.2	A company might be in stage 3 before they go to stage 2. The order can be switched for some classifications.

	I2.1	Stage 'Getting support in the organization' needs a factor which says that scope should be defined, as not all data should be treated the same. This should be in relation to the business challenge.
	I3.1	Classifications are a bit unclear, because some terms are not straightforward and hard to understand.
4. Understandability	I1.2	It is a start, but applicability comes with the next level of detail.
	I3.1	This framework is similar to a growth model that was used at O11, so it was easy to understand.
5. Applicability	I1.1	The framework helps to make a plan for how to tackle a data governance platform project. It cannot be applied in the stage for deciding whether the system is suitable or what should be included.
	I1.2	It can be used for planning to accomplish the implementation.
	I2.1	If an organization wants to work on data governance, you have to go through the cycle multiple times. The steps are recognizable, but it is presented as one framework while there should be many iterations. You should not apply it to all data.
	I3.1	An organization can start with sponsorship, after that implement a few sources, and then gradually grow, which is very applicable for all organizations.
6. Usability	I1.1	Usable if very immature in data governance platform implementation and can help in devising a plan, but needs more guidance for organizations who are more mature.
	I1.2	It is a good way to enter the conversation and set a base-line.
	I2.1	Frameworks are sometimes used to defend why a 3 year long project is needed, but you need to look at what the problem is, how you can solve it, and make your own roadmap.
	I3.1	There should be another model behind it (e.g. RADAR) to determine the maturity of the organization and at which stage they should start. Sometimes you have to improve the architecture and sometimes you have to adjust something on the organizational side (e.g. culture), and this influences your growth model.
7. Robustness	I1.1	The world changes fastly, so shorter steps help make it more adaptable.
	I1.2	You need to check off each of the items to go to the next phase, but maybe it should be possible to move out of order.
8. Generalizability	I1.1	The framework is generalizable, but it would be interesting to see it in the context of the sector. For the financial sector, there are additional requirements for regulations.

Table 22 shows the comments given on the evaluation criteria for framework B.

Table 22. Explanations from interviews of scores given for evaluation criteria for Framework B

Evaluation criteria	Interview #	Explanation
1. Clarity of concepts	I1.1	You have to specify what you mean by "transactional", as multiple people within the organization use this term in a different way.

	I3.1	Large unstructured sources are about email, social media, and directories which increases complexity a lot, but not sure how data governance should handle this for an organization.
2. Clarity of relationships	I1.1	These might not be the right dimensions as the intersection between the two are not logical or practical. Why a platform project was important for O3 was based on rules and legislation, which depends on the sector.
	I1.2	Relations are less clear; not sure how to go from one quadrant to the other.
3. Accuracy of classifications	I1.1	If you want to get all data in the system, you limit your time to make improvements on data, so this quadrant might not be accurate.
	I1.2	There could be companies that are very transactional, but they could also be focused on insights even if they are transactionally focused.
	I2.1	Governance is not relevant for some data (e.g. it is not needed for euro signs). It can be imported in Purview as you need to find it, but it is not important to do something with it. They are in your catalog, maybe add a definition, but there is a high chance that you filter it from the catalog because no one uses it. This should be taken into account in the framework.
	I3.1	It is not clear why large unstructured sources require to include as many sources as possible in order to cover the whole data landscape.
4. Understandability	I3.1	The framework requires an explanation of its use, some examples, and a reason for why this needs to be considered as it is not self-explanatory.
5. Applicability	I1.1	This is more of a self-reflection for thinking about what to include.
	I1.2	It can be applied as a starting point. It is more a description or a status, but not a strategy or a process. It can be used to prioritize investment or tactics.
6. Usability	I1.1	Two dimensions are difficult as there is a limited set of predefined quadrants.
	I1.2	Not sure what to do with it; if you are placed in one quadrant, do you want to go to the other quadrant?
	I2.1	The relation with the business problem needs to be clear.
	I3.1	The dimensions seem relevant, although it is difficult to specify what data governance must look like for unstructured data.
7. Robustness	I1.2	Lots of businesses are going to say that they can be placed in multiple boxes, so how to know which box describes them best.
	I3.1	There is a difference between simple structured sources and complex structured sources, so complexity should gradually increase.
8. Generalizability	I1.1	It is questionable whether it can be applied to different sectors. As there are many regulations in the financial industry, there are few requirements for e.g. the energy sector. The framework might stay the same, but it should be dependent on the sector how you would complete it.

Appendix K: Research planning

In figure 23 the planning for the research project is shown. In the first phase, the long proposal was finalized which includes a planning and justification for the research problem and the research methods. In the second phase, the research plan was executed. This included the performing of the literature review, the conducting of interviews, the design of the framework and the validation of the framework. In the wrap-up phase, the report was finalized and the thesis defense was prepared.

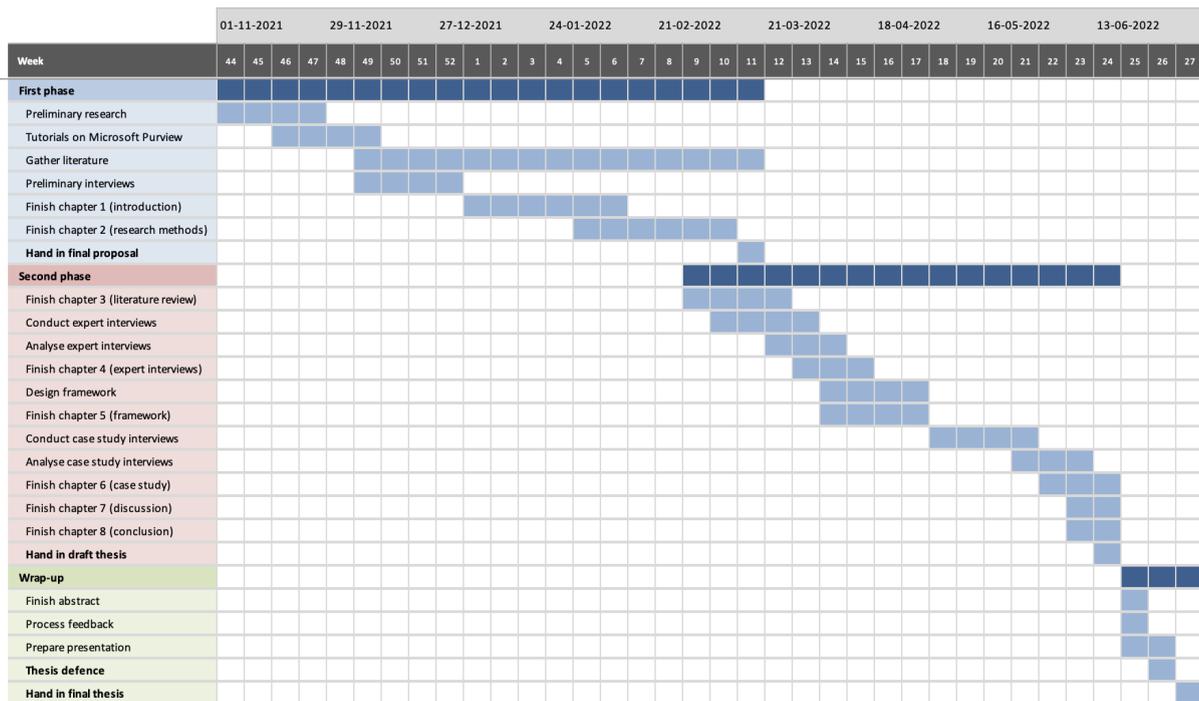


Figure 23. Gantt chart of project planning