

ALIGNING DATA ARCHITECTURE AND DATA GOVERNANCE

MASTER THESIS BY

NEETU ARAVIND

ELECTRICAL ENGINEERING MATHEMATICS AND COMPUTER SCIENCE

BUSINESS INFORMATION TECHNOLOGY

EXAMINATION COMMITTEE

prof. dr. M. E. Iacob

prof. dr. ir. M. Van Sinderen

COMPANY SUPERVISOR

Andreas Wombacher



UNIVERSITY
OF TWENTE.

Acknowledgement

Being born in the "Information age", my brother and I have been constantly exposed to a technological life which has rapidly shifted from the traditional industry to an economy which is completely based on Information technology. My brother, being a PhD graduate who is keen on creating new technologies, had a different goal set for himself. Well, for myself, I had other plans. All the innovations and developments in the field of technology have kept me wondering as to how the manufacturing and service sectors could operate in a more efficient and convenient manner. These thoughts ignited many questions in my mind. My career aim of becoming a good Information Technology professional made me decide to pursue a master's program in Business Information Technology at the University of Twente and it helped me to further expand my knowledge in this field.

As a part of my program, I chose to do my Master thesis graduation assignment in collaboration with Aurelius Enterprise, Amsterdam, Netherlands. This report is about the topic "Aligning Data architecture and Data governance: Developing a model that covers and complements both the data architecture and the data governance". The thesis is carried out in the form of research.

Throughout the writing of this dissertation, I received a lot of help and encouragement. First, I would like to thank my University supervisors Maria Iacob and Marten Sinderen for their excellent guidance and for providing the right direction to complete my dissertation. I would particularly like to single out my COMPANY supervisor Andreas Wombacher, I want to thank him for his patience, support and time I was given to further my research. I would also like to extend my deepest gratuities to my student counselor Evelien Vink and my study advisor Bibian Rosink for being there to help me with all the struggles I faced during my research. Without them I couldn't have completed my studies.

I Cannot begin to express my thanks to Abhinaya and Ritu – Growing apart doesn't change the fact that for a long time we grew side by side; our roots will always be

tangled. I'm glad for that. Special thanks to my friends Deepak, Smrithi, Suriya and Venus for always being there for me, for making me feel supported, loved and cared for.

Much love to Nikhileta, chechu and Jithu for helping me to push my boundaries, lifting me up when I'm feeling down, and giving me comfort when I'm sad. Thank you so much.

Last but not least I would like to thank my parents, without whom none of this would have been possible, I owe you an immeasurable debt of gratitude for the long, hard hours you worked throughout our youth to give me and my brother the opportunity to follow our dreams, and for all of the tremendous love you've shown us along the way.

This thesis is dedicated to my Acha and Sheejama

For your endless love, support and encouragement

Executive Summary

In recent years, there has been a reawakening, with companies realizing the value of data as a strategic advantage as well as an organizational necessity. Managing and harnessing the power of data and processes, on the other hand, is becoming increasingly difficult. Companies are leveraging enterprise data for better efficiency and decision-making in today's rapidly evolving market environment. Data governance programs must be founded on a thorough understanding of business processes, a grasp of how data is moved and transformed within the enterprise, and a shared language to ensure efficient communication. Organizational data, procedures, business rules, priorities, and strategies must be carefully controlled. Data must be accessible and consumable, with adequate access and visibility based on roles and responsibilities. Hence, data governance is a basic foundation that must be prepared to enable data management, and data architecture is a fundamental thing in data management.

Many data architectures have ended up as shelfware in many organizations, never being deployed in the real world. There are a number of reasons for this, but one of the most common is that many data architecture projects lack business support and participation. Data architecture is often misunderstood by business people as an academic, abstract, technical practice of little or no relevance to them. Having business people in the development and implementation of the architecture would greatly increase the likelihood of a successful implementation. Like all data disciplines, data governance and data architecture may have a different emphasis and focus, but they are mutually reinforcing.

No prior studies have been found relating to the two disciplines as a topic. However, there are studies found in the context of the relationship, saying that if the aligning happens, it will be more effective at mitigating risk and avoiding steep penalties for non-compliance. But no further research is to be found. Therefore, a neat research to find the alignment in the topic is required.

OBJECTIVE:

Organizations with governance processes have a relationship with data architecture processes that come into alignment. The architecture is completely implicit but correlates to the prime functions of data governance. In the end, an organization needs both. We need data governance and data architecture, and we don't want to be in a situation where one stands in the way of the other. For both, there is a different model, but the aim is to construct another model that allows combining the two. The main objective of this thesis is to "Develop a model that covers and complements both the data architecture and the data governance".

METHODOLOGY:

According to (Wieringa, 2014) this research can be classified as a design science problem because the goal of this research is to solve a specific problem by developing an artifact (Method proposed can be found in [Chapter 4](#)). As a result, as a design science research, this thesis will structure the chapters of this master thesis report using the Design Science Research Methodology (DSRM) developed by (Peppers, Tunnanen, Rothenberger, & Chatterjee, 2007).

KEY FINDINGS:

The **CHAPTER 3** Literature Review provides several aspects relating the to the state-of-the-art literature available for Data governance and Data architecture. Both Data governance and Data architecture are two different domains which are still developing and validating research. We went deeper into the literature articles for the research question. One aspect that is common to all the selected literature is that most of these studies are still specialized and focused individually while not paying too much attention to further connections between both the disciplines. The findings can be found in the Literature section.

In **CHAPTER 4** the aim is to achieve the alignment data architecture and the data governance is constructed through a model and how the achievement is made can be extracted from the model. The alignment model was designed with the use of literature review and mainly based on the contents of the two books that are widely used for Data management DAMA (DAMA-DMBOK2, 2017) and DCAM (DCAM, 2017). The model gave clarity on how the areas are aligned and the relationship between them.

In **CHAPTER 5** the alignment model of data governance and data architecture have been evaluated to determine the correctness, quality, utility, efficacy and understandability of the designed model. This is accomplished by qualitatively evaluating what an expert thinks of the model. With the gathered feedback from the interview, the model has been adjusted accordingly and the final version of the developed alignment model is presented.

A case study was then performed to observe the usefulness of the developed alignment model in practice. The model is validated case study evaluation was to observe and identify how much the developed alignment model is followed in the real organization and the benefits of using the model is given. To measure the extent to which the proposed alignment model meets the benefits, it is quantitatively validated by a panel of experts. The results are provided in this section.

In **CHAPTER 6** this master's thesis report concludes, which is followed by contributions to scientific and practical contributions, limitations, and recommendations for future work.

Table of Contents

Acknowledgement	ii
Executive Summary.....	iv
List of Figures	x
List of Tables	xi
List of Acronyms.....	xii
1 Introduction.....	2
1.1 Thesis objective	3
1.2 Research question	4
1.3 Thesis structure	6
2 Research Methodology	10
2.1 Research design	10
2.2 Research methods	12
2.2.1 Literature review	13
2.2.2 Case study	15
2.3 Data collection method	15
2.3.1 Questionnaire.....	15
2.3.2 Observation	16
2.3.3 Interview	16
2.4 Summary and Conclusion.....	17
3 Literature review	19
3.1 Data governance and Data architecture	22
3.2 The areas of overlap in data governance and data architecture	29
3.3 The impact on business process performance when there is no alignment.....	36
3.4 Aligning Data Architecture and Data Governance	41

3.5	Summary and Conclusion.....	43
4	Design and development	46
4.1	ArchiMate.....	47
4.2	Defining activities of Data Governance and Data Architecture	48
4.2.1	Data Architecture Activities	48
4.2.2	Data Governance Activities.....	50
4.3	The Alignment Model	58
4.3.1	DA and DG processes with relevant tasks and roles.....	65
4.4	Summary and Conclusion.....	68
5	Model Evaluation and Validation	70
5.1	Evaluation approach	70
5.1.1	Selection of the Interview, Questionnaire and the material used	70
5.1.2	Profiles of the interviewees	71
5.1.3	Evaluation results and the final version of the alignment model	72
5.1.4	Suggestions and feedback gathered from the interview	73
5.1.5	Evaluation Conclusion.....	77
5.2	Case study.....	77
5.2.1	Case Description	78
5.2.2	Case Observation.....	79
5.2.3	Validation approach.....	83
5.2.4	Measurement Design.....	85
5.2.5	Analysis and Result	87
5.2.6	Case study conclusion	92
5.3	Conclusion: Evaluation and Validation	92
6	Conclusion	94
6.1	Summary & Conclusion.....	94

6.2	Scientific and practical contribution.....	96
6.3	Limitations	97
6.4	Recommendations.....	98
	Appendix A: Identified Literature paper	104
	Appendix B: Interview Agenda for model evaluation	110
	Appendix C: Evaluation Questions.....	111
	Appendix D: Evaluation of Benefits	117

List of Figures

Figure 1.1 Sub-questions of the research	6
Figure 1.2 Outline of the thesis	8
Figure 1.3 Structure of the thesis	9
Figure 2.1 Design Science Research Methodology (Peppers, Tunnanen, Rothenberger, & Chatterjee, 2007)	10
Figure 3.1 The process used to conduct the systematic literature review	22
Figure 3.2 Data governance organization	25
Figure 3.3 DAMA-DMBOK2 Data Management Framework	29
Figure 3.4 Analysis of a Survey on the latest trends in Data Architecture	31
Figure 3.5 DM model based on DAMA Knowledge Areas (Steenbeek, 2019) ...	33
Figure 3.6 Relation between DA,DG & business process (Steenbeek, 2019)	37
Figure 3.7 Conceptualization	38
Figure 4.1 Business Process Notation	47
Figure 4.2 Triggering Notation	47
Figure 4.3 Serving Notation	48
Figure 4.4 Flow Notation	48
Figure 4.5 Specialization Notation	48
Figure 4.6 AND Junction	48
Figure 4.7 Relation between identify and define the data	51
Figure 4.8 Identify the Data	61
Figure 4.9 Define the Data	64
Figure 5.1 The final version of Alignment model - Identify the data	75
Figure 5.2 The final version of Alignment model – Define the data	76
Figure 5.3 UTATUT Research Models (Venkatesh, Morris, Davis, & Davis, 2003)	84

List of Tables

Table 1 Guidelines for DSR research and application.	12
Table 2 Overview of research methods & data collection methods used.....	13
Table 3 SLR Activities.....	14
Table 4 Database of sources.....	19
Table 5 Inclusion and Exclusion Criteria	21
Table 6 DA processes and description	55
Table 7 DG processes and description.....	58
Table 8 Responsible people for the task.....	67
Table 9 Overview of evaluation interviews	72
Table 10 UTAUT constructs used in the questionnaire	87
Table 11 Questionnaire result	89
Table 12 Questionnaire result of PE	90
Table 13 Questionnaire result of EE	91
Table 14 Quality Assessment Form	109
Table 15 Response Q1.....	118
Table 16 Response Q2.....	118
Table 17 Response Q3.....	120
Table 18 Response Q4.....	121
Table 19 Response Q5.....	122
Table 20 Response Q6.....	123

List of Acronyms

Abbreviation ***Explanation***

<i>DSRM</i>	Design Science Research Methodology
<i>SLR</i>	Systematic Literature Review
<i>DA</i>	Data Architecture
<i>DG</i>	Data Governance
<i>DM</i>	Data Management
<i>DGT</i>	Data Governance Team
<i>SSI</i>	Semi Structure Interview
<i>UTAUT</i>	Unified Theory of Acceptance and Use of Technology
<i>PE</i>	Performance Expectancy
<i>EE</i>	Effort Expectancy

PART I

Introduction to the topic

1 Introduction

In recent years, there has been a reawakening, with companies realizing the value of data as a strategic advantage as well as an organizational necessity. Managing and harnessing the power of data and processes, on the other hand, is becoming increasingly difficult. Companies are leveraging enterprise data for better efficiency and decision-making in today's rapidly evolving market environment. Data governance programs must be founded on a thorough understanding of business processes, a grasp of how data is moved and transformed within the enterprise, and a shared language to ensure efficient communication. Organizational data, procedures, business rules, priorities, and strategies must be carefully controlled. It must be accessible and consumable in the company, with adequate access and visibility based on roles and responsibilities. Hence, data governance is a basic foundation that must be prepared to enable data management, and data architecture is a fundamental thing in data management. With the DAMA wheel of 2017 (DAMA, 2017) and 11 disciplines placed, there is data governance right at the center of the data management activities. The implication is that you really cannot do anything without having a solid core around data governance in the middle, as it is required for consistency within and balance between the functions. The other noticeable implication from the wheel is that if one needs to deliver a data architecture, it is almost impossible to do without having Data Governance in place to drive the business leadership of that data architecture.

Much of the data architecture has ended up as shelfware in many organizations, never being deployed in the real world. There are a number of reasons for this, but one of the most common is that many data architecture projects lack business support and participation. Data architecture is often misunderstood by business people as an academic, abstract, technical practice of little or no relevance to them. Having business people in the development and implementation of the architecture would greatly increase the likelihood of a successful implementation. Like all data disciplines, data governance and data architecture may have a

different emphasis and focus, but they are mutually reinforcing. Organizations with governance processes have a relationship with data architecture processes coming into alignment. But the architecture is completely implicit. Some organizations try to address the need to bring out the benefits of aligning holistically by defining an enterprise data architecture.

In the context of Data Governance, Data Architecture, a seemingly simple job, becomes as difficult as six blind men constructing an elephant model. Each blind man sees the elephant from a different angle. This is analogous to stakeholders and employees who are dispersed around the organization, each with their own interpretations and implementations of Data Governance. As a result, many businesses end up with fundamentally diverse knowledge silos, each owned by a separate group and used for distinct purposes. This poses a risk and cost to an organization that values Data Governance. The data architect is in the middle of it all, and he or she often has the most mature and holistic picture of information and data. Data architects, on the other hand, have a hard time bringing different silos and teams together. No prior studies have been found relating the two disciplines as a topic. However, there are studies found in the context of the relationship, saying that if the aligning happens, it will be more effective at mitigating risk and avoiding steep penalties for non-compliance. But no further research is to be found. Therefore, a neat research to find the alignment in the topic is required.

1.1 Thesis objective

Data Governance is a management structure that is layered on top of data to ensure that it is identified, registered, categorized, and handled in a consistent manner. Its role is to make decisions about data ownership and data maintenance over the course of the data life cycle, data quality, and data compliance. It also defines and regulates the rules for data usage, access, aggregation, and flow. Data architecture is becoming increasingly crucial for enterprises to design, improve, record, and maintain. This is partly due to a growing desire for access, data integration, and data exchange with other parties, as well as legally mandated

insights into internal data flows. It brings standardization to names and, most importantly, definitions of entities across the organization. Data governance and data architecture may have a different emphasis and focus, but they are mutually reinforcing. Organizations with governance processes have a relationship with data architecture processes that come into alignment. The architecture is completely implicit but correlates to the prime functions of data governance. In the end, an organization needs both. We need data governance and data architecture, and we don't want to be in a situation where one stands in the way of the other. For both, there is a different model, but the aim is to construct another model that allows combining the two. The main objective of this thesis is to "Develop a model that covers and complements both the data architecture and the data governance".

To achieve this, a bunch of sub-objectives are set. Having an awareness of where they overlap, how they relate to each other, is the first step in that direction. What is the intersection? How are they, enhancing each other, or contradicting each other? Why is aligning important? How can alignment be achieved?

1.2 Research question

The main goal of this thesis is to propose a model that covers both the data architecture and the data governance. Because organizations with governance have processes that have a relationship with data architecture processes coming into alignment. Data architecture clearly supports data governance, but it must also be acknowledged that it's not a one-way relationship. In the end, an organization needs both. Even though there are two different models for the two disciplines, the ultimate goal is to propose a model that allows them to combine the two. Based on this goal, the main question of the research can be formulated as follows:

RQ: “How to align data governance and data architecture and how it can be achieved?”

A main research topic was established, as a result of which various sub-questions had to be formed Figure 1.1. The result of answering these sub-questions is the generation of the main deliverables of the research. Sub-questions, labeled as knowledge questions (K) or as Design questions, are shown by (D). When trying to find out the answer to a knowledge question, it is answered by investigating the state of the art surrounding a subject or an artifact. In contrast, when trying to answer a design question, it is answered by identifying design criteria, investigating possible solutions to the research problem, and examining trade-offs between various solutions (Wieringa, 2014).

SQ 1: What is Data governance and Data architecture? (K)

The main goal of this sub-question is to investigate the state-of-art theories, models, methods, and techniques regarding both disciplines. Furthermore, it helps investigate the currently widely accepted and used model to present its limitations.

SQ 2: What are the areas of overlap in data governance and data architecture? (K)&(D)

In order to answer this research question, both literature and practice are to be used. First, investigate the state-of-the-art regarding data governance and data architecture in literature, what the literature defines as the overlap areas, and how it enhances and contradicts it. Second, using the results of the literature review as a basis for a targeted investigation with practitioners. From both literature and practice, this will provide a good amount of information in order to form the methodology developed from both sources.

SQ 3: How to achieve the alignment between data governance and data architecture? (D)

The main deliverable of this research question is to determine the alignment model to bring the implicit data architecture outside, revealing the relationship with the case of data governance. The answer is based on the results of the previous research questions and includes guidelines, accompanied by formal modelling.

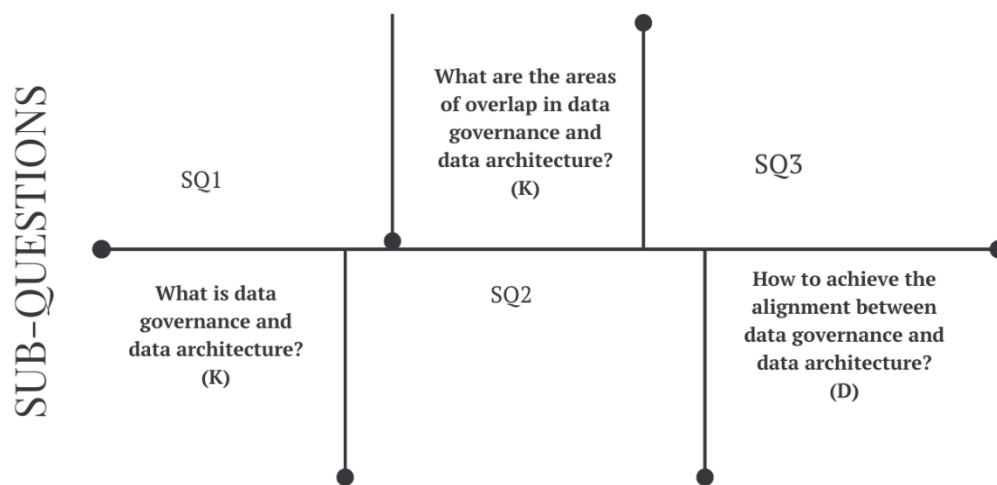


Figure 1.1 Sub-questions of the research

1.3 Thesis structure

This dissertation adheres to most of the guidelines of the Design Science Research Methodology (DSRM) by Peffers et al. (Peffers, Tunnanen, Rothenberger, & Chatterjee, 2007) which follows the five steps: problem identification and motivation; defining the objectives for a solution; design and development; demonstration; evaluation and communication. I have relied on the study conducted by Peffers et al. (Peffers, Tunnanen, Rothenberger, & Chatterjee, 2007) in order to conduct this research and to frame this dissertation,

as described in the following paragraphs. Furthermore, it employs the DSRM to assist in answering the research questions we have set out for our research (Section 1.2). Furthermore, these chapters are grouped into three parts, beginning with an introduction to the topic, followed by a proposed solution to the research problem, and ending with the validation and evaluation of the proposed solution, as proposed by (Wieringa, 2014) according to the DSRM. Figure 1.2 is the outline of the thesis and figure 3 depicts the structure of the thesis chapter wise.

Problem identification and motivation: Before research can be done, the problem must be clearly defined, and the value of the suggested solution must be communicated. The goal of this thesis is to offer a clear overview of the problem identification and motivating activity, which can be found in [Chapter 1](#), and a concrete investigation, which takes place in [Chapter 3](#). The DSRM work helps to provide a partial response to the questions of SQ1 and SQ2. As discussed in [Chapter 2](#), the research approach which is followed throughout this dissertation is elaborated upon in further detail in this chapter.

Defining the objectives for a solution: It is critical that research objectives be established on the basis of the problem definition. These objectives can be regarded as quantitative when they describe how the proposed solution can outperform existing ones, or when they describe how the suggested technique can help solve problems that have never been addressed before. According to Peffers et al. (Peffers, Tunnanen, Rothenberger, & Chatterjee, 2007), the resources needed to undertake this task include knowledge about the current state of research and possible solutions. Once again, it can be observed in [Chapter 3](#), where the literature that is available at the moment is thoroughly reviewed, which provides detailed responses to all of the knowledge research questions ([SQ1,SQ2](#)).

Design and development: The method that is presented as a solution to the problem is developed in this activity. Based on the literature review, this includes

determining the method's functionality and architecture. The design and development activity in this dissertation may be observed in [Chapter 4](#), where the suggested method's design is provided. This DSRM activity contributes to the solution of the design research problem ([SQ2, SQ3](#)).

CHAPTERS	MAIN TOPICS	RQS	
Introduction to the topic <i>DSRM: Problem identification</i>	<ul style="list-style-type: none"> • Problem statement • Research questions 	SQ1,SQ2	
	Research Methodology	<ul style="list-style-type: none"> • Research design and methods 	
	Literature Review <i>DSRM: Objectives for the solutions</i>	<ul style="list-style-type: none"> • Data architecture, Data governance • Areas of overlap • Misalignment impact on business performances • Alignment of DA & DG 	SQ1 SQ2 SQ3
Solution to research problem <i>DSRM: Design and development</i>	<ul style="list-style-type: none"> • Activities of Data Governance and Data Architecture <ul style="list-style-type: none"> • Model with ArchiMate • The Alignment Model • Relevant tasks and roles of DA and DG processes 	SQ2,SQ3	
Validation & Evaluation <i>DSRM: Demonstration</i>	Validation of the Model <i>DSRM: Demonstration</i>	<ul style="list-style-type: none"> • Case study 	SQ3
	Evaluation of the Model <i>DSRM: Evaluation</i>	<ul style="list-style-type: none"> • Selection of the Interview • Suggestions and feedback gathered from the evaluation • Final version of the alignment model 	SQ2,SQ3
	Conclusions	<ul style="list-style-type: none"> • Summary & Conclusion • Scientific and practical contribution • Limitations • Recommendations 	

Figure 1.2 Outline of the thesis

Validation: To establish the ability of the proposed method, it must be proven. Experimentation, simulation, case study, evidence, and other methods can be used to accomplish this. In this thesis, the validation is applied to one case study and it is presented in [Chapter 5](#). This DSRM activity contributes to the solution of the design research problem ([SQ3](#)).

Evaluation: In order to see if the proposed strategy is effective, it must be evaluated how nicely it accompanies the issue. This requires comparing the research aims to the demonstration activity's observable results. The evaluation

of our suggested approach is presented in [Chapter 5](#) which includes a semi-structured interview with a professional. This DSRM activity contributes to the solution of the core design research topic ([SQ2, SQ3](#)).

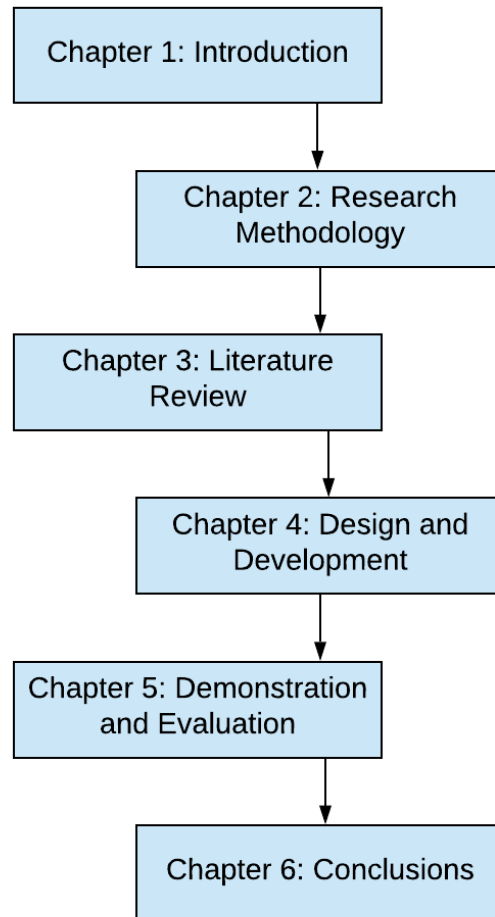


Figure 1.3 Structure of the thesis

2 Research Methodology

This chapter introduces the research design that has been employed, along with the research methodologies that have been employed.

2.1 Research design

In order to fulfill the goal of this dissertation, we have decided to apply design science research methodology since it is aligned with the overall objectives of the thesis. That is, we intend to address and solve a specific problem by creating an artefact ([Chapter 4](#)). Design science is a research methodology that emphasizes the connection between theoretical knowledge and practical application by showing that scientific knowledge can be produced by designing useful things (Wieringa, 2014). According to Hevner et al., design science is a problem-solving paradigm which aims to create an artifact that relies on existing kernel theories that are applied, modified, and extended (Hevner, March, Park, & Ram, 2004).

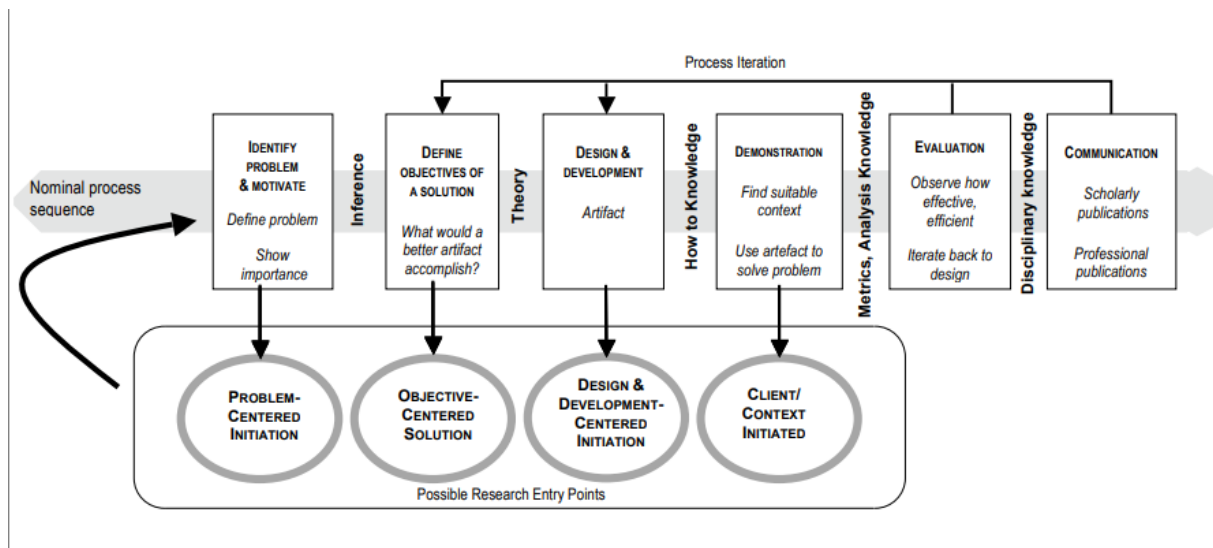


Figure 2.1 Design Science Research Methodology (Peffer, Tunnanen, Rothenberger, & Chatterjee, 2007)

Multiple approaches to structuring the DSRM process have been proposed by researchers. Problem investigation, solution design, and solution validation are the three processes proposed by Wieringa (Wieringa, 2014). We divide the dissertation into three parts using these three phases: As noted in [Section 1.3](#),

Part I is an introduction to the issue, Part II is a solution to the research challenge, and Part III is a validation and evaluation of the solution.

Design science, according to Peffers et al. (Peffers, Tunnanen, Rothenberger, & Chatterjee, 2007), takes a slightly different approach. Problem identification and motivation, defining the objectives for a solution, design and development, demonstration, evaluation, and communication are the six primary phases identified by the authors, which can be considered a further specification OF those presented by Wieringa (Wieringa, 2014). To format the dissertation chapters, we selected to use the six phases recommended by Peffers et al. (Peffers, Tunnanen, Rothenberger, & Chatterjee, 2007). Figure 2.1.1 depicts the activities associated with the design science research approach. Hevner et al. (Hevner, March, Park, & Ram, 2004) argue that behavioral science (which has the goal of uncovering the truth) and design science (which has the goal of creating utility artefacts) are inseparable and highly influential on each other. Furthermore, the authors introduced a number of guidelines to help researchers provide and perform high quality design science research. The description of those guidelines can be seen in Table 1.

Providing the general recommendations from Hevner et al. (Hevner, March, Park, & Ram, 2004) throughout the dissertation, which essentially produces an artefact in the form of a method for DG & DA alignment, to address the business needs for the alignment was accomplished. Furthermore, we demonstrate the benefits of our artefact by means of a case study. Additionally, conducted a Systematic Literature Review (SLR) on research topics to gather all the necessary knowledge in order to ensure that the designed artefacts are in accordance with all the requirements of the problem environment (DG & DA alignment). The following part of this thesis is focused on presenting the results in a way that is understandable and meaningful to both technology and management-oriented audiences. Using terminology specific to both fields alongside easily understood explanations will aid in that.

<i>Guidelines</i>	<i>Description</i>
<i>Guideline 1: Design as an Artefact</i>	DSRM must produce a viable artefact, such as a construct, model, method, or instantiation.
<i>Guideline 2: Problem Relevance</i>	The main objective of DSRM is to develop technology-based solutions to relevant business problems
<i>Guideline 3: Design Evaluation</i>	The utility, quality, and efficacy of an artefact must be rigorously demonstrated via evaluation methods
<i>Guideline 4: Research Contribution</i>	DSRM must provide contributions to the areas of the design artefact, design foundations, and/or design methodologies.
<i>Guideline 5: Research Rigour</i>	DSRM relies on the application of rigorous methods in both the construction and evaluation of the artefact.
<i>Guideline 6: Design as Research Process</i>	The search for an effective artefact requires the utilization of available means to reach desired ends while satisfying the rules of the problem environment
<i>Guideline 7: Research Communication</i>	DSRM must be presented effectively to both technology oriented and management-oriented audiences

Table 1 Guidelines for DSR research and application.

2.2 Research methods

This research has employed many research methods in order to comprehensively understand the problem, to examine and validate the proposed solution, and to support decisions. Table 2 below outlines the different research

methods and accompanying measurement instrument(s) that have been utilized to answer each research question. The measurement instruments are more thorough and specific in the section following this one.

<i>Research question</i>	<i>Research method</i>	<i>Data collection</i>
<i>SQ1: What is Data governance and Data architecture?</i>	Literature review	-
<i>SQ2: What are the areas of overlap in data governance and data architecture?</i>	Literature review	-
<i>SQ3: How to achieve the alignment between data governance and data architecture?</i>	Literature review Case study	- Observation Interview/ Questionnaire

Table 2 Overview of research methods & data collection methods used

2.2.1 Literature review

As described by Kitchenham and Charter (Kitchenham & Charters, 2007), this research was carried out as a systematic literature review (SLR). SLR is a means of identifying, evaluating, and interpreting all available research relevant to a particular research question, topic, or phenomenon of interest (Kitchenham & Charters, 2007). The main goal of SLR is to summarize existing data governance and data architecture knowledge by determining the true relations between the disciplines in order to find the alignment between them, as well as to gain a true Understanding and reflection on current data governance and data architecture research and practice, and to identify potential research directions based on the current literature.

Planning	
1	Define the main Research Question and its Sub-Questions
2	Select scientific databases
3	Formulate search query based on the main Research Question
4	Define inclusion and exclusion criteria
Selection	
5	Execution of formulated search query for each scientific database
6	Article selection for each query results from inclusion criteria
7	Remove duplicate studies across scientific databases
8	Exclusion of irrelevant articles based on title and abstract assessment
9	Exclusion based on full text availability and its assessment
Result Analysis	
10	Data extraction according to defined main RQ
11	Synthesis of the extracted data
12	Report synthesis results on defined main RQ

Table 3 SLR Activities

The aim of this study is to look at and review the most recent published studies on data governance and data architecture to find the alignment. This SLR process is divided into three process steps, which are started by Planning, Conducting, and Analysis of Results as shown in Table 3. However, because the underlying practices revolve primarily around the selection of previous studies, the term "conducting" will be referred to in this article as "selection.". One key research issue will be addressed in this article. Table 3 shows a more detailed list of activities, which will be explained in more detail in the following sub-sections.

2.2.2 Case study

The How? And why? questions are the most common research questions that can be answered via a case study. A case study is a research method that contrasts and complements survey research. It is a study of a population, which can also be an individual. Case studies, on the other hand, are utilized to gain a deeper understanding of real-life occurrences and relationships (Yin, 2003). Interviews, observations, and workshops are the most typical data collection devices utilized in case studies. We conducted a case study in our research, with one quite diverse organization. In [Chapter 5](#), you'll find more information about the case study.

2.3 Data collection method

During this research, questionnaire data collection method is used to gather relevant information for the research. The following section describes the characteristics of the collection method of the questionnaire.

2.3.1 Questionnaire

The questionnaire is a data gathering method that involves a series of questions or other types of items that are designed to gather useful information that can be analyzed. There are numerous sorts of questionnaires that can be utilized in a research, such as self-administered questionnaires and interviewer-administered questionnaires (Saunders, Thornhill, Lewis, & Bristow, 2015). Self-administered questionnaires, in which respondents complete the questionnaire without interaction with the researcher by filling the answers, on the other hand, interviewer administered questionnaires, in which the researcher asks questions and records the responses interacting with interviewee, (Kotzab, Seuring, Muller, & Reiner, 2005). Questionnaires can be categorized according to the kinds of delivery such as emails, personal interviews, or via mailing.

In this thesis, we have utilized both the self-administered questionnaire and interviewer administered questionnaire. Interviewers administered questionnaires used for semi-structured interviews to evaluate the model to determine the correctness, quality and understandability of the model. A self-

administered questionnaire to collect qualitative data from our respondents to validate the benefits of the alignment model. The findings of the questionnaires have served to outline the model and thesis conclusion. We will take these findings into consideration while formulating the design criteria. The full questionnaires are included in [Appendix C](#) & [Appendix D](#).

2.3.2 Observation

Observation, as the name suggests, is a method of gathering information through observation. Because the researcher must immerse herself in the setting where her respondents are while taking notes and/or recording, observational data collection is classed as a participatory study. Observation can be structured or unstructured as a data collection approach. Data is collected using certain variables and on a pre-determined timetable in structured or systematic observation. Unstructured observation, on the other hand, is carried out for an open and unstructured manner, with no pre-determined variables or goals. Direct access to study phenomena, high levels of application flexibility, and the creation of a permanent record of events to be referred to later are all advantages of observation data gathering. At the same time, the observation approach has drawbacks such as lengthier time requirements, high levels of observer bias, and observer impact on primary data, which means that the presence of an observer may influence the behavior of sample group parts. However, for the evaluation part of the thesis it is done by observing the case study and it can be seen in the later [chapter 5](#).

2.3.3 Interview

The interview is a data gathering strategy that focuses on the interviewer and interviewee's verbal engagement with the goal of developing knowledge in a certain area or topic. Data gathered through interviews is primarily reliant on respondents' ability and willingness to provide correct information. Structured, semi-structured, and unstructured/in-depth interviews are the three primary forms of interviews (Lussier, 2015).

Researchers favor semi-structured interviews because it allows them to ask both the pre-prepared questions as well as go further into areas that are important to interviewees (Lussier, 2015). In this dissertation, we conducted two rounds of semi-structured interviews to acquire essential information for the development of the model. An expert working in the field will be interviewed to test the usefulness, efficacy and understandability of the designed model. Their reactions to the model will be utilized to make the necessary adjustments to the theoretical model.

2.4 Summary and Conclusion

In this chapter, we present our selected research methodology, namely Design Science Research Methodology (DSRM), which not only directs our research but also defines the structure of this thesis. DSRM is a research methodology that is commonly employed in information systems studies due to its emphasis on building artifacts that are aimed at solving specific problems. We chose the research methodology because it aligns with the main goals of our research, which is to design an alignment model for DA and DG.

For qualitative research, we employ a case study, which is an empirical investigation that explores a current phenomenon in depth and within its real-life environment, particularly when the borders between the phenomenon and the context are unclear. We decided to conduct a semi structured interview to acquire the essential data. The term "interview" refers to a direct data gathering approach focused on conversational contact between the interviewer and interviewee with the goal of developing knowledge in a certain area or topic. Observations can be categorized as behavioral or non-behavioral, and they are used to research participants in their natural surroundings. They are sometimes the only way to obtain certain sorts of data. Finally, documentation refers to actual records containing information on a certain topic or organization that can be utilized as input for research methodologies such as for case study.

Furthermore, we performed a Systematic Literature Review (SLR) to gather information about the research topic, determine what has already been established to help solve the problem, and help build the research solution. The SLR is a three-step procedure that includes searching for information, evaluating the information received, and synthesizing the information assessment. The information acquired with the use of the SLR is used to develop the alignment model.

3 Literature review

In this chapter, the results of the Systematic Literature Review are presented. SLR will be carried out to explore the alignment between data architecture and data governance. Specifically, to find the potential synergies mentioned in the research and finally, to determine how the alignment can be achieved. The final goal of the research presented in this paper is to identify the alignment between these two disciplines from previous studies and construct a model in the next chapter derived from the findings that can be applied to a specific context.

Scientific Databases

This section defines the scientific databases chosen for this review in order to obtain relevant academic publications and answer the defined research questions. These databases were chosen because they are capable of providing comprehensive coverage of both the latest and earlier scholarly literature related to this topic. Furthermore, these databases are considered among the top five most reliable academic resource databases. In addition to those, other records were identified, including white papers. The scientific databases selected for this review consisted of:

NAME OF THE ELECTRONIC DATABASE	WEBSITE LINK
GOOGLE SCHOLAR	https://scholar.google.com
IEEE XPLORE	https://ieeexplore.ieee.org
SCIENCEDIRECT - ELSEVIER	https://www.sciencedirect.com/

Table 4 Database of sources

Search Query Formulation

The search query is formulated based on a set of keywords related to the research questions. The main keywords are obtained from the relevance towards answering the main question as well as the sub-questions. Furthermore, synonyms are also defined for each main keyword so as to widen the articles that can be gathered. The key words are:

“Data Governance”, “Data Architecture”, “Business Process”

Based on the keywords listed above, search queries for each scientific database are formulated by clustering the synonymous keywords together using the logical operator “OR” and further attached by the other clusters using the “AND” operator. In order to further control the relevance of the search result, the search query is applied the article’s title, abstract, and keywords. The resulting search queries after several iterations are as follows:

Data governance AND Data architecture AND (Business process OR Business performance OR processes OR artifacts)

Inclusion and Exclusion Criteria

Kitchenham and Charters (Kitchenham & Charters, 2007) stated that defining the selection criteria is essential in order to reduce the likelihood of bias in the search process and can help to identify the direct evidence towards the primary study. In this section, the inclusion and exclusion criteria are defined and listed in Table 5. Following this, articles that comply with the defined inclusion will be chosen as candidates and likewise, those which do not satisfy the exclusion criteria will be removed. The papers used in this paper are those written in English in order to ensure that the articles chosen were peer-reviewed globally. Since peer-review is taken into account, studies presented in conference proceedings and journal articles are chosen with the same care to ensure the publication's quality. Other articles related to the topic were identified, including white articles. Furthermore, as previously discussed, included research areas are used to keep

the search results relevant to the primary study. In terms of publication year, this study does not restrict the search criteria in order to capture the topic's overall development. Furthermore, since the same article is often found in different scientific databases, duplicates suggested by a similar title or material would be reduced. Finally, articles that are incomplete or too short, such as those that only show the first page of online search, will be excluded.

Inclusion Criteria	Exclusion Criteria
English based peer reviewed Studies	Studies that are not related to the main RQ from title, abstract and content
Studies published in Conference Proceedings and Journal Articles	Duplicate articles with title or content
Study areas focusing in the field of Computer Science, Engineering, Business Management & Accounting, Social Science	Articles that are not complete or too short
H-index higher than 4	Paper published before 2000

Table 5 Inclusion and Exclusion Criteria

Selection

The gathered papers must still be checked in order to increase the relevancy of this review to the primary study and to avoid wasting time reading irrelevant publications. The first step is to perform the specified search queries on each scientific database, followed by the second step, which is to run the defined search queries on each scientific database. Using the previously mentioned inclusion and exclusion criteria, the metadata from the search results is then exported to EndNote, where it can be further selected based on the title and abstract. The third step is to filter duplicate results by title and abstract. Fourth,

collect the full text of the selected articles and discard those that cannot be contained in its full text document or whose full text is incomplete. The fifth step is to evaluate the full text of the posts, and only those that include discussions that are similar to answering the main and sub questions are chosen. By the end of the operation, 16 papers have been chosen, and the flow of the entire procedure is depicted in Figure 3.1. An overview of the literature paper identified during this literature review is given in [Appendix A](#).

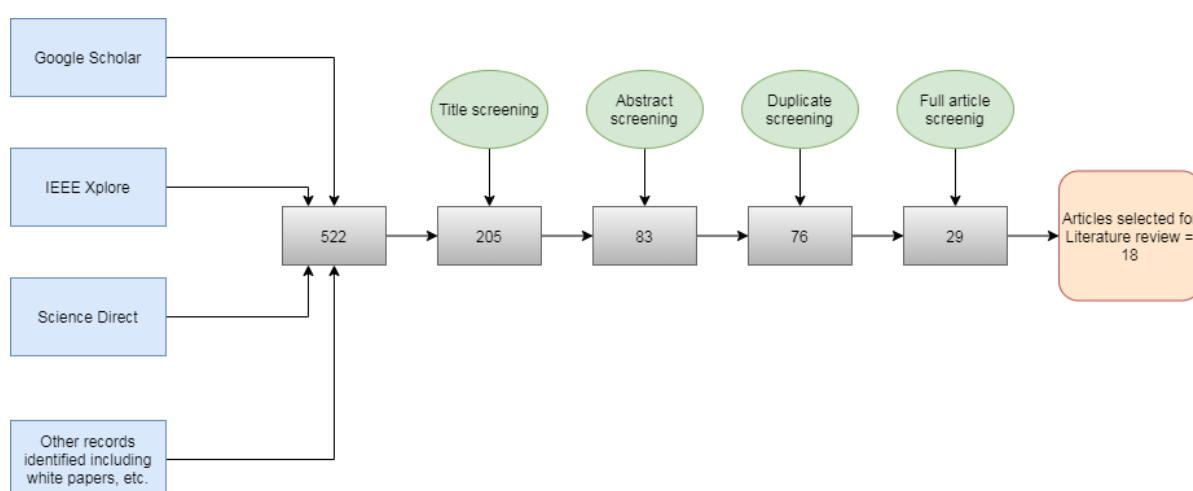


Figure 3.1 The process used to conduct the systematic literature review

3.1 Data governance and Data architecture

This section discusses the effort towards answering the question "What is Data Governance and Data Architecture?" There are so many unequivocal definitions for Data Governance and Data Architecture. Below are a number of definitions of both disciplines taken from across several scientific publications.

Data governance

Data governance is just one part of the overall discipline of data management, concerning the capability that enables an organization to ensure that high data

quality exists throughout the complete lifecycle of the data, and data controls are implemented that support business objectives. It encompasses the people, processes, and technologies required to manage and protect data assets. Researcher differ in defining data governance, the Data Governance Institute (DGI) defines it as follows “data governance is a system of decision rights and accountabilities for information-related processes, executed according to agree upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods” (Khatri & Brown, 2010). According to the DAMA Guide to the Data Management Body of Knowledge (DAMA, 2017) Data Governance is the exercise of authority and control (planning, monitoring, and enforcement) over the management of data assets.

Complimenting this definition, Seiner (Seiner, 2014) has defined as it is the formal execution and enforcement of authority over the management of data and data-related assets. Author of (Panian, Some practical experiences in data governance., 2010) defined data governance as “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”. Another study has defined this as (Russom, 2008), it is usually manifested as an executive-level data governance board, committee, or other organizational structure that creates and enforces policies and procedures for the business use and technical management of data across the entire organization. From the various studies it is understood that data governance is not something that can be applied immediately. It generally requires planning and preparation because it entails various complex tasks that must be coordinated. Meanwhile, the authors of (Niemi, 2015) defined data governance as “specifies the framework for decision rights and accountabilities to encourage desirable behavior in the use of data. To promote desirable behavior, data governance develops and implements corporate-wide

data policies, guidelines, and standards that are consistent with the organization's mission, strategy, values, norms, and culture".

The above definitions stress the importance of the terminals from which data governance activities may be carried out on data-related assets that support the organization's strategy. All scholars further recognize that data governance encompasses both decision rights and responsibilities related to the management of data assets in organizations.

According to (DAMA, 2017) the scope and focus of a particular data governance program will depend on organizational needs, but most programs include:

Strategy: Defining, communicating, and driving execution of Data Strategy and Data Governance Strategy

Policy: Setting and enforcing policies related to data and metadata management, access, usage, security, and quality

Standards and quality: Setting and enforcing Data Quality and Data Architecture standards

Oversight: Providing hands-on observation, audit, and correction in key areas of quality, policy, and data management (often referred to as stewardship)

Compliance: Ensuring the organization meets data-related regulatory compliance requirements

Issue management: Identifying, defining, escalating, and resolving issues related to data security, data access, data quality, regulatory compliance, data ownership, policy, standards, terminology, and data governance procedures

Data Management Projects: Sponsoring efforts to improve data management practices

Data asset valuation: Setting standards and processes to consistently define the business value of data assets

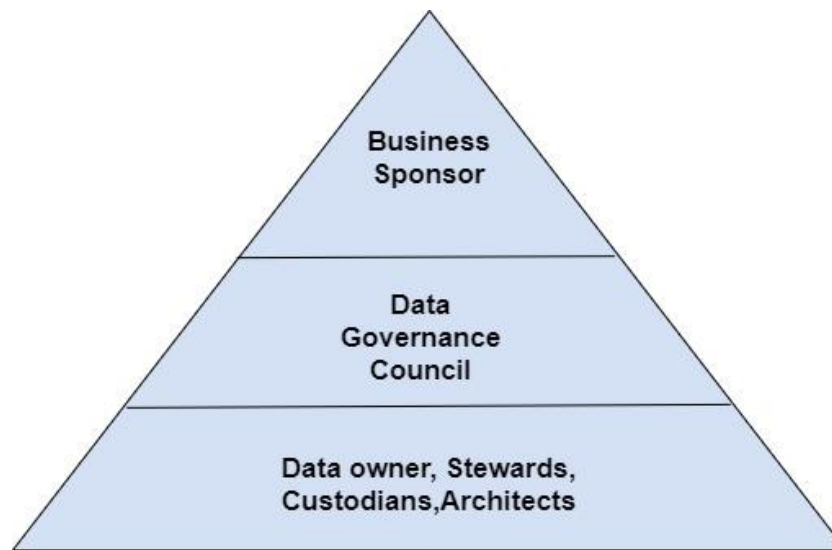


Figure 3.2 Data governance organization

To ensure that the same data standards and policies are defined and enforced across the entire organization, one needs to establish a Data Governance organization (Figure 2) that represents a generic data governance model and involves a multi-tiered combination of business and technology roles. At the top of the organization's business sponsors, business sponsors provide overall leadership and sponsorship to all data governance efforts. Data Governance initiatives require resources, funding, and sponsorship and are the key roles in providing that. The next layer in the data governance pyramid is the Data Governance council. It provides consistency and coordination for cross functional initiatives, while maintaining an enterprise perspective and strategic approach to data quality. The last layer is roles such as data owners, stewards, custodians and architects are responsible for operationalization of data standards, policies and procedures. Each of these layers can be associated with one term that indicates the role in the data governance capability, sponsor provides *sponsorship*, data governance council provides *directions* and data owners stewards custodians and architects, provides *execution* of data governance principles.

Since the word data policies and data standards are going to be mentioned several times, it is ideal to know the exact the difference between those two terms - data policies refer to general guidelines, usually related to the entire subject area, for instance, sales or finance. On other hand, data standards refer to particular data elements, like customer names. The core of data governance is ensuring the data assets are in accordance with business policies. Data Governance thus serves several purposes:

- ✓ Data identification, classification, and registration
- ✓ Identify the appropriate data quality standards for each data type (e.g., no outdated data)
- ✓ Identify compliance standards that apply to certain data sets (e.g., retention times for financial records)
- ✓ Implement concrete measures to establish compliance with applicable regulations for a specific set of data (e.g., automatic alerts if data reaches its retention period and must be deleted)
- ✓ Creating efficient methods to ensure that data management is carried out as efficiently and effectively as possible.

Data governance relies on creating a standardized data architecture plan that serves as the foundation for layering data policies to ensure usability, quality, and consistency. All authors also agree that Coordinating Data Architecture to support better understanding of the data and the systems. In order to support a better understanding of the data and the systems, coordinating with Data Architecture is a prioritized activity in Data Governance. So, what is Data Architecture?

Data architecture

The definition provided by the author (Gupta & Cannon, 2020) is “data exist to satisfy business requirements, and data architecture is the foundational element to link data with requirements”. Data Architecture is the way in which information flows around the organization. It is a well-designed framework to determine what data is required to move the company forward, where the data can be stored,

and how it can be distributed to deliver actionable information to decision makers. Other authors say it comprises the definition of enterprise data objects and the development of an enterprise data model on a conceptual, logical and physical level. (DAMA, 2017) According to the DAMA Guide, Data Architecture will be considered from the following perspectives and together these three forms the essential components of Data Architecture:

- *Data Architecture outcomes*, such models, definitions and data flows on various levels, are usually referred to as Data Architecture artifacts.
- *Data Architecture activities* to form, deploy and fulfill Data Architecture intentions
- *Data Architecture behavior*, such as collaborations, mindsets, and skills among the various roles that affect the enterprise's Data Architecture.

The authors mention that (Sherman, 2015), (DAMA, 2017) “a solid data architecture is a blueprint that helps align your company's data with its business strategies” as it governs how the data is collected, integrated, enhanced, stored, and delivered to business people who use it to do their jobs. It helps make data available, accurate, and complete so it can be used for business decision-making. The goal of architecture is to simplify as much as possible, create reusable standards and optimize efficiency, so that the practice can support the future growth of the business. Data Architecture breaks down by going through three traditional architectural processes:

- Conceptual - represents all business entities.
- Logical - represents the logic of how entities are related.
- Physical - the realization of the data mechanisms for a specific type of functionality.

Data architecture states how data is persisted, managed, and utilized within an organization. (Cristian, Anca, & Cerasela, 2008) Data architecture also describes the following:

- ✓ How is data stored in both a transient and permanent manner?
- ✓ What components, services, and other processes utilize and manipulate the data?
- ✓ How do legacy systems and external business partners access the data?
- ✓ How do common data operations (create, read, update, delete) occur in a consistent manner?

Data architecture is important for many reasons, including that it (Sherman, 2015):

- ✓ Helps you gain a better understanding of the data.
- ✓ Provides guidelines for managing data from initial capture in source systems to information consumption by business people.
- ✓ Provides a structure upon which to develop and implement data governance.
- ✓ Helps with enforcement of security and privacy.

Data architecture principles vary considerably from one enterprise to another, depending on an enterprise's business requirements and the importance of data to that enterprise (Hoven, 2006.) However, here are some common principles that form the foundation of data architecture:

- Data should be viewed and managed as a shared asset.
- Common and shared definitions to ensure common understanding.
- Users require adequate access to data.

3.2 Areas of overlap in data governance and data architecture

Let's see how specifically then we do governance and data architecture support and reinforce each other in this section. The figure below is the DAMA wheel of 2017 (DAMA, 2017) with 11 disciplines placing Data Governance right at the center of the data management activities. The implication is that you really cannot do anything without having a solid core around data governance in the middle as it is required for consistency within and balance between the functions. The other noticeable implication from the wheel is that if one needs to deliver a Data Architecture, it is almost difficult and impossible to do without having Data Governance in place to drive the business leadership of that Data Architecture.

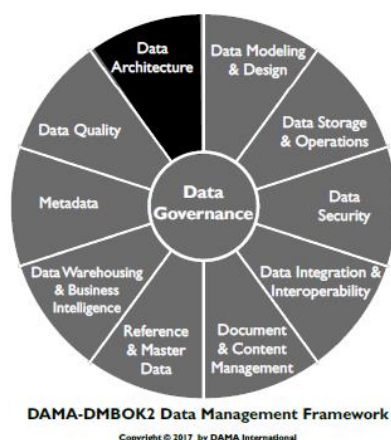
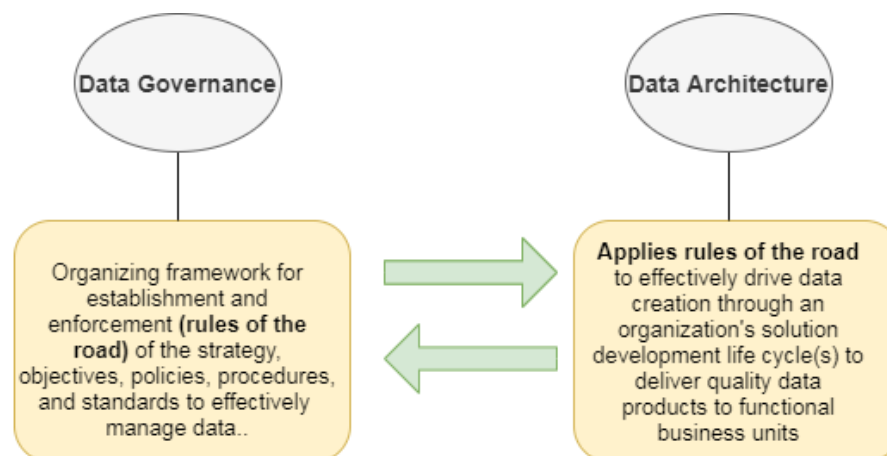


Figure 3.3 DAMA-DMBOK2 Data Management Framework

If we go back to the definition of data governance as the organizing framework for the strategy around data, then data governance is really creating the idea of the rules of the road. What are the standards and policies and what are the processes that we want to implement in order to make data consistent, appropriately available, trusted and consumable across the enterprise? Governance sets up these standards and data architecture also sets up their standards, but they most commonly apply these rules down road to effectively drive data creation, data storage, and developing the additional applications and data capabilities throughout the organization.



The paper (Burbank & Roe, 2017) is an analysis of a Survey on the latest trends in Data Architecture and the below Figure 3.4 is the graph of who is typically responsible for creating Data Architecture. From the graph it can be seen that the person most responsible for creating a Data Architecture is the Data Architect. But what is interesting about the survey is that to see the key areas where the collaboration between two different disciplines is needed to deliver architecture, the Data Governance officers are one of the top two. Some researchers and practitioners have mentioned the relationship and the need to align both the disciplines. Data architecture explains where data is stored and how it moves around the organization and its systems. It emphasizes changes and transitions that occur when data is transferred from one system to the next.

These data inventory and data flow diagrams provide the Data Governance Team (DGT) with the information and tools it requires to make effective data policy and standard decisions. When business people raise data issues, these artifacts assist the DGT in performing root cause analysis and resolving those issues. Data flow diagrams and data inventory can also assist in determining what can be measured, when, and how. They will assist in identifying the potential business impacts of enhancing data quality in systems by gaining a better understanding of who uses the systems and for what purposes, as well as facilitating the development of metrics and measurements. Depending on who

creates and updates the data and in which systems, these diagrams will aid in determining how to measure adherence to standards.

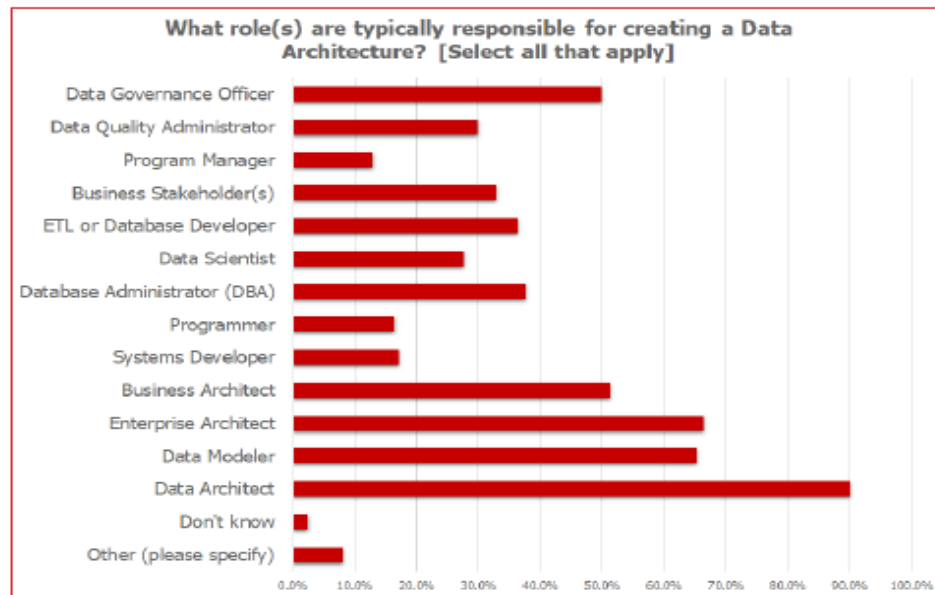


Figure 3.4 Analysis of a Survey on the latest trends in Data Architecture

Furthermore, it's also useful for identifying the right owners and stewards, as well as the main stakeholders with a direct interest in the data. Product Numbers, for example, may be owned by one or more Operations managers (possibly in various parts of the world), Product Description by Global Marketing, and Product Price by regional Finance teams. This information will also be used to ensure that the right people are in the room and that they involve cross-business collaborators in any work to improve the quality of product information. *Data inventory and data flow diagrams, with data accountability and ownership overlaid, are critical for finding any gaps in accountability and ownership.*

When the organization starts to determine what data types and sources should be within the scope of a formal data governance program, and which to exclude, at least at the outset is a challenge. This is the first challenge where data architecture can be used to good effect. A great place to start identifying key data types is using two of the core artefacts of data architecture, namely conceptual

and logical data models. This is the first task in which data architecture can be beneficial. Using two of the main artefacts of data architecture, conceptual and logical data models, as a starting point for identifying key data types is a great place to start. The author (Loshin, 2015) agrees that, mentioning that, data governance relies on developing a uniform data architecture plan that provides the foundation for layering data policies for ensuring usability, quality, and consistency. This data architecture plan must embrace the vision for a unified set of conceptual and logical models while integrating the details of the existing data artifacts in use across the organization.

The paper (Loshin, 2015) suggested that because of the difficulties posed by the lack of governance in legacy system designs, the increasing interest in repurposing data from around (and even outside) the enterprise indicates that modeling and metadata management cannot be done in a vacuum going forward. Rather, best practices for enterprise data design, modeling, sharing, and reuse must be developed at the organizational level. This indicates the need for clear data governance policies related to various aspects of data architecture aiming to minimize structural variation. If we look at this as one direction of alignment, where data governance can provide guidance on a data architecture practice that helps architecture practices stay aligned with their business constituents, and to ensure that the work is prioritized from a business perspective. Now, data architecture is also in a position where, as they're engaging with their peers in the IT organization that is, they can identify opportunities for the data governance organization, to ensure that data governance is linked in to the IT side of the house as well.

In practice, it is difficult to see a clear distinction between Data Architecture, Data Modeling, and Design. DAMA-DMBOK2 (DAMA, 2017) recognizes conceptual, logical and physical data models as the main deliverables of Data Modeling. At the same time, DCAM (DCAM, 2020) and TOGAF 9.1 recognize these models as outcomes of Data Architecture.

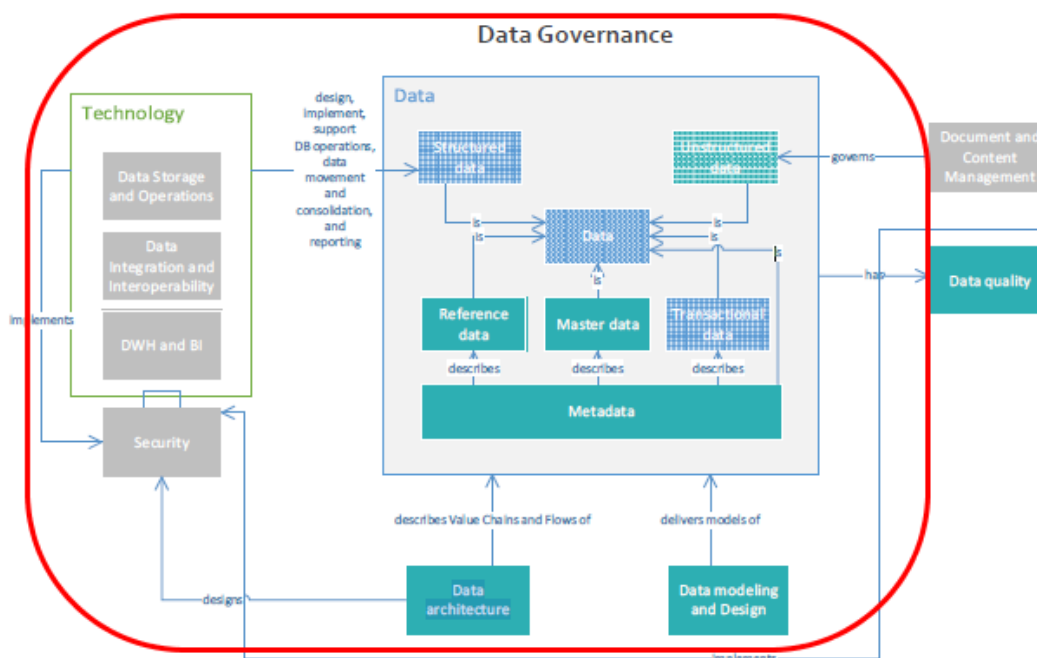


Figure 3.5 DM model based on DAMA Knowledge Areas (Steenbeek, 2019)

The above Figure 3.5 from the (Steenbeek, 2019) entails that Data Architecture details Information Value Chain and Data Flows. Data Modeling and Design develops data models to express data requirements. Both Data Architecture, Modeling, and Design are Knowledge Areas as well. Data security is designed by Data Architecture and then implemented in data processing systems. Data Governance handles organizational aspects of data management such as strategy, policies, processes and roles and is an overarching part of all other data management functional areas.

Let's look at the areas of overlap in data architecture and data governance:

DCAM (DCAM, 2020) considers data domains, critical data elements, data domain taxonomies, data classification and requirements as tasks of data governance, while from DAMA-DMBOK2 (DAMA, 2017) point of view these are elements of Data architecture and Data modeling which they will consider next.

Also (Cervo & Allen, 2015) mentioned that the data governance program should work with data architect teams to agree on how data domain definitions and structures that should align. Creating an aligned and commonly recognized data domain structure will greatly simplify your ability to focus on data management, data governance, and data-quality improvement initiatives.

- ❖ *Data structures - Data domains, critical data elements, data domain taxonomies, data classification*

DAMA-DMBOK2 (DAMA, 2017) says data architects often act as business liaisons for governance activities, setting rules, guidelines, and specifications for how data is used within the organization, as well as creating artifacts that enable compliance with governance directives.

- ❖ *Defining standards*
- ❖ *Creating data-related artifacts*

DAMA-DMBOK2 recognizes Business Glossary and Data Dictionary as a deliverable of Data Governance, while DCAM sees it as a deliverable of Data Architecture. Also agreeing to (DCAM, 2020) the article (Urso, Vince, 2020) claims that a standard business glossary is required to define business terms along with links or mappings to the various technical data dictionaries that define the production management of these items as data attributes and that is a part of data architecture.

- ❖ *Business Glossary and Data Dictionary*

Data inventory and data flow diagrams provide the Data Governance Team (DGT) with the information and tools it requires to make effective data policy and standard decisions.

- ❖ *Data inventory and data flow diagrams*

Security policies to be applied to each data entity. The standardization of policies and procedures in the data architecture reduces the complexity generated by

multivariate implementations of similar operations by preventing duplication of effort. The conceptual model of data architecture defines which data components are sensitive information, the logical design will have confidential information protected in a database with limited access, restricted data replication, particular data types, and secured data flows to protect the information. Data architecture must be fully compliant with privacy regulations and data protection laws such as GDPR. All data should be encrypted before ingestion and personally identifiable information (PII) should be anonymized.

❖ *Aligning policies, Requirements and controls*

Data governance focuses on planning, monitoring and enforcing activities related to data handling, maintaining the required level of quality of data, and specifying and describing data by means of data architecture and modeling (Steenbeek, 2019)

❖ *Description data*

Data architecture & modeling develop and maintain data lineage that helps to identify main critical data elements and in the Data governance the details tracked in data lineage are a good way to provide compliance auditing, improve risk management, and ensure data is stored and processed in line with organizational policies and regulatory standards.

❖ *Data lineage*

A well-structured data governance process aligned with the MDM program model and data architecture strategy will greatly aid in determining the master data priorities and control points where data steward roles can be most effective (Cervo & Allen, 2015).

❖ *data stewards can be most effectively engaged - in a multi-domain model that will be highly dependent on data architecture*

3.3 The impact of business process performance when there is no alignment

These capabilities are marked blue in Figure 3.6 and are limited to data governance, data architecture, and data modeling. In the data management toolkit book (Steenbeek, 2019), it describes how both DA and DG are connected to business processes, even though that is not under the scope of the DG framework. The data-filled green box is placed in the model's center to emphasize this point. To properly manage data, you'll need to develop a number of skills. Data governance, data architecture, and data modeling are the only capabilities marked in blue. Application architecture can be a part of the enterprise architecture function or a capability within the data management function. This is why the box exists. Governance of systems and applications as well as business processes are not within the scope of the data governance framework. This is why boxes with systems, applications, business processes, and documents are marked in orange. Context is added by mapping the relationships in a typical organization, creating a generalized high-level overview of the effects of data governance and data architecture alignment in an organization (Figure 3.6). The figure explains Data architecture and modeling describes data. Data architecture describes the following aspects of data: data definition, data location, and the path the data flows from its origin to its current destination.

Data modeling is a technique that specifies relations between data elements via constructing data models at different abstraction levels. Data governance focuses on planning, monitoring and enforcing activities related to data handling, maintaining the required level of quality of data, and specifying and describing data by means of data architecture and modeling. Data is usually located and processed in systems or applications as well as in some documents. systems and applications, as well as documents relating to particular functional business processes.

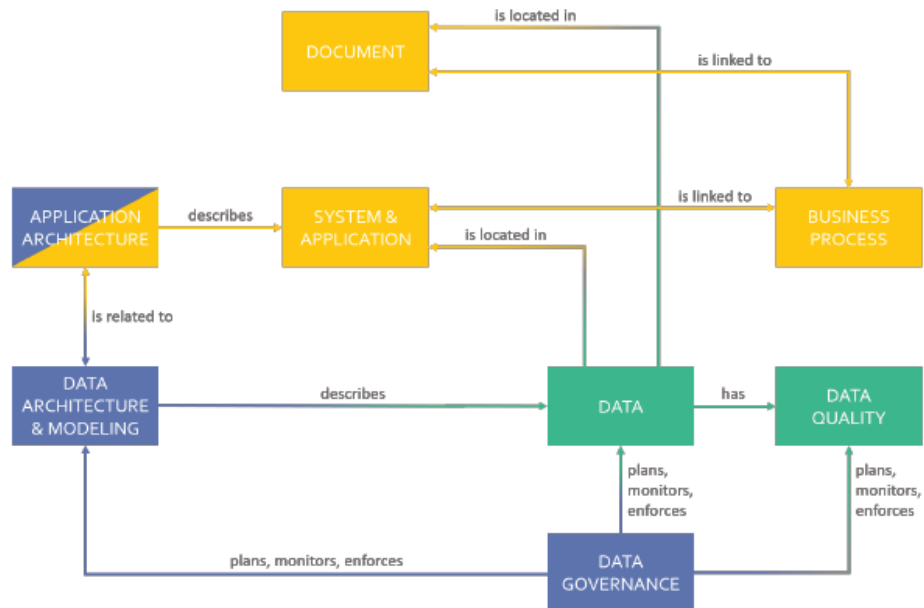


Figure 3.6 Relation between DA, DG & business process (Steenbeek, 2019)

The four basic processes conducted with data in databases are known as CRUD processes. (Nick, Joris, & Mark, 2015) Other important data processes are the acquisition, validation, exchange and archiving of data Figure 3.6. When a piece of data is used in the (supply chain) process, such as information about the shipment address, size, weight, and content of goods, the data is "Read" for use in the business. Since the organization would have more control over its data processes, data governance has an effect on CRUD and other data processes in the Data layer. This includes things like roles and responsibilities for enterprise data.

When no responsibilities are defined, data errors can go undetected or be discovered much later in the process. If a responsible employee is assigned, it is fair to expect that less errors will be brought into the system, or that errors will be resolved sooner, resulting in more effective processes. Clearly, there is an alignment happening between the two disciplines. Data layer processes produce a certain level of data quality. When a purchase order is made and processed, for example, the data quality of the order is determined by the efficiency of these

processes. This is due to the fact that data quality is a multifaceted concept that includes factors like timeliness.

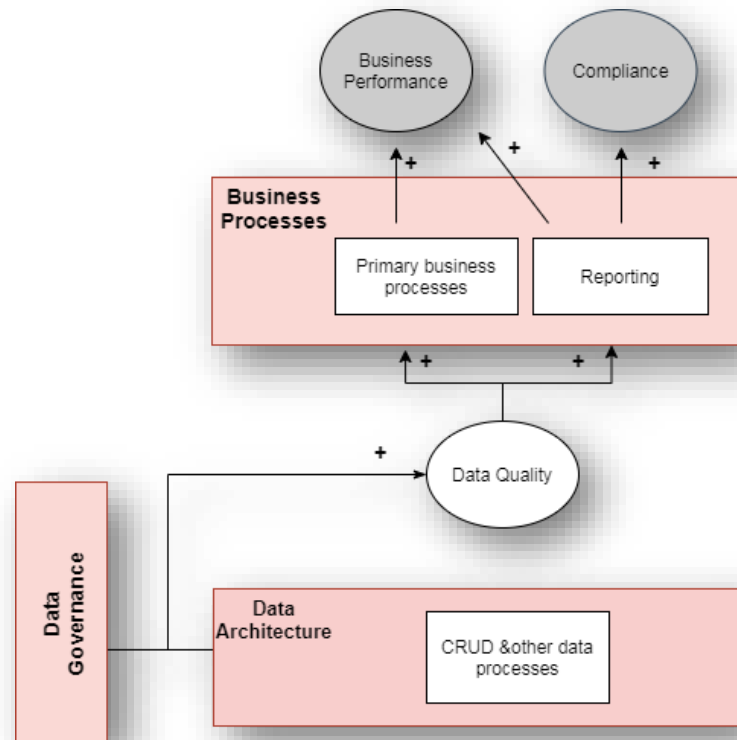


Figure 3.7 Conceptualization

Data quality has an impact on primary business processes and reporting, as adequate data quality is necessary to support efficient business processes and reporting (Nick, Joris, & Mark, 2015). The amount and composition of the planned products, for example, must be known during the production process of a specific product. If data quality is poor, as a result of failing CRUD processes, it may lead to errors in the production process as well as in business process reporting, such as error logs and financial data.

The causal relationship chain in Figure 3.7 ends with business performance and regulatory compliance. Initially, the quality of business processes largely reflects

the enterprise's business performance (Nick, Joris, & Mark, 2015). This relationship is fairly clear, as a lower throughput in the production process would result in lower output and, as a result, lower business performance (considering constant product quality and market price). Second, reporting has an effect on an organization's business performance and compliance. On the one hand, reporting is used for business process forecasting and performance overviews, allowing processes to be managed (monitoring). Reporting (audit trails), on the other hand, is used to demonstrate compliance with a variety of laws and regulations, depending on the country and sector. From the casual relations between the disciplines, it can be seen there is an alignment happening between data architecture and data governance. As seen in the diagram, it can be inferred that data governance and data architecture together form an overall part of the organization's structure. That this whole organization is influenced by alignment.

If there is no alignment, it affects the data itself. A database's data, like other resources, can be faulty, out of date, or flat-out inaccurate. Most businesses waste huge sums of money each year due to incorrect data. Quality data procedures are important in order to make sound judgments that are based on complete and accurate data eventually affecting the business process. For example, an organization may choose to implement a client data master hub to serve all business lines that produce or consume core client data in order to achieve more accurate client identification. For marketing purposes, a client is someone who has shown interest in the organization's programs, even though they have not signed up for them or have used them once but not in many years. In sales, a client is sometimes referred to as a "prospect" before a specific event occurs, such as the signing of a contract. When it comes to operations, the same person cannot be considered a customer until after services have been started. There may be several existing structures in the current environment where a client record is first developed. A client's record may be updated by several systems, each of which may require different mandatory data about the client. Operations that require that a new client's service status be set to "active"

before a record in their tracking system is made. This requirement can cause client data to be out of sync with sales, resulting in poor aggregate reporting and the failure to implement certain business processes.

The need for alignment between data governance and data architecture is emphasized in the second issue. Non-compliance with regulatory demands, uncertainty regarding data definitions, and who is responsible for what data can all result from poor alignment. It is self-evident that key assets cannot be managed if no one is held accountable for the information. This is why it's critical to have a strong data governance program in place to maintain the process of uniformity. First and foremost, this enables firms to adhere to regulatory requirements. Second, it aids businesses in having a clear understanding of data definitions, roles, and responsibilities. Finally, it allows businesses to be more transparent about who makes modifications to data.

As an additional example, post-processing and -implementation errors can also occur due to inadequately designed communication processes between the two disciplines. The lack of communication processes in data governance with data architecture can lead to data collection errors remaining undetected and unresolved for prolonged periods of time. Changes in local law, such as GDPR regulations in Europe, which went into force in May 2018, have had an impact on data governance. Pre-collection and data processing had to change as a result of GDPR regulations. These changes necessitated optimization of current tag management methods for many businesses. If required changes are not conveyed effectively, they can result in post-processing privacy issues if they are not addressed quickly and efficiently.

Process efficiency and performance as they relate to internal quality – As stated above, data governance affects the CRUD and other data processes in the *Data* layer, as the organization will be more in control of its data processes as a consequence of data governance. (Nick, Joris, & Mark, 2015) It entails, for instance, roles and responsibilities regarding enterprise data. When no

responsibilities are set, errors in the data might not be rectified, or only recognized much later in the process. If a responsible employee has been appointed, it can be assumed that fewer errors will enter the system, or that errors will be rectified earlier, leading to more efficient processes. Data governance is attributed to improving data quality due to increased accuracy, availability, completeness, consistency, and timeliness of data, and the limitation of errors due to data inconsistencies (Abraham, Schneider, & Brocke, 2019). For example, due to failing CRUD processes, it could induce errors in the production process, as well as in the reporting of the enterprise processes, such as error logs or financial data.

3.4 Aligning Data Architecture and Data Governance

By performing the SLR, it was able to find several aspects relating the to the state-of-the-art literature available for Data governance and Data architecture. Both Data governance and Data architecture are two different domains which are still developing and validating research. The literature regarding Data governance has been relatively abundant in the last few years while Data architecture is also becoming prominent. Data architecture has recently been contributing positively to the achievement of governances' business goals in many industry domains. Since the purpose of this study used systematic literature review, the results of the study are the current state of knowledge in the area of data architecture and data governance alignment.

Based on the findings, if we use a full range of data architecture artifacts such as Data flow diagram, CRUD Matrix, Data Model, Process Model in the data governance effort, the data architecture or data model can provide answers for the following questions such as,

- What exactly do I mean by “customer”?
- What exactly do I mean by a “claim”?
- Do I understand the process that it's a part of?

- May I know where the data was used?
- What are my business rules and policies?
- Where is the PII?

While data architecture obviously helps with data governance, it's still crucial to remember that it's not a one-way path. Most data architecture has been shelfware in many organizations, never being deployed in the world today. There are many explanations for this, but one of the most common is that several data architecture initiatives lack business support and participation. Every single organization has different levels of maturity, both in terms of data architecture and in terms of data governance. So, there will be some organizations that may have a well-defined data governance program, but no really clear data architecture. There will be others who have a clear data architecture, but they don't have governance in place. Some of them have both. It is important to see that when we align the disciplines of data architecture and data governance, business performance will improve and evolve.

The alignment can be achieved from the findings of the overlap in the areas of both data governance and data architecture. Highlighting the areas of overlap presented in [section 3.2](#) above is listed below:

- Data standards and procedures,
- Data structures - Data domains, critical data elements, data domain taxonomies, data classification
- Aligning policies
- Requirements and controls
- Data lineage
- Business Glossary and Data Dictionary
- Data inventory, data flow diagrams, Description data
- Identify and appoint stewards - data stewards can most effectively engage
 - in a multi-domain model that will be highly dependent on data

architecture, reviewing and approving DA, managing and resolving data related issues.

3.5 Summary and Conclusion

The delivery of the proposed framework in this paper has shown a set of common elements found in a typical data architecture and data governance initiatives from academic articles. However, this result cannot be said as verified yet, since, compared to the considerable number of publications on data architecture and data governance in general, the number of articles presenting research findings on the alignment is rather low, so some limitations can still be identified from this framework. First, the domain or organization where this alignment is being applied is still limited by the defined keywords being used to construct the search query. This limitation has been taken into account to limit the scope that this systematic literature review should cover. In order to grasp a more comprehensive domain coverage that is closer to real industry implementations, more exhaustive research that considers industrial reports and an expanded industrial sector should be considered.

By performing the SLR, it was able to find several aspects relating to the state-of-the-art literature available for Data governance and Data architecture. Both Data governance and Data architecture are two different domains which are still developing and validating research. The literature regarding data governance has become relatively abundant in the last few years, while the data architecture is also becoming more prominent. Data architecture has recently been contributing positively to the achievement of governance goals in many industry domains. Since the purpose of this study used a systematic literature review, the results of the study represent the current state of knowledge in the area of data architecture and data governance alignment.

One aspect that is common to all the selected literature is that most of these studies are still very specialized and focused individually while not paying too much attention to further connections between both the disciplines. Identifying

the search results found that there was certainly a limit to conducting the study because of the articles available relating to both, there were very few disciplines together, leading to very few published papers talking about the relationship between data architecture and data governance or the synergies between them. An analysis of the governance processes where architectural alignment is happening has been identified. Even though data architecture and data governance are the disciplines of data management, their studies are still lacking in both DA and DG. More studies are therefore required to expand, develop and implement the alignment. And so, this is where I need to conduct more research and case studies are needed to validate the effectiveness of alignment to achieve a better result and to understand how alignment can bring benefits to the organization

Therefore, I would like to conclude that there are several research gaps that can be identified from the literature. Firstly, there are few solutions that consider combining knowledge from multiple domains to propose a way to improve business process performance, so the future work and the challenge is to find the method to achieve the alignment. Second, what could be the potential synergies of this alignment method? And finally, to find the overlaps in the disciplines, such as in responsibilities and accountabilities of the data, establishing the right decisions, identifying stakeholders, and so on.

PART II

Solution to the research problem

4 Design and development

In the previous chapter, the literature review has been studied which is the foundation to design and develop the model. This chapter will answer the sub-research question SQ2 – “What are the areas of overlap in data governance and data architecture?” and SQ3 – “How the alignment is achieved between data governance and data architecture?” It leads to the design and development of the DG & DA alignment model in [Section 4.3](#).

Different components could be included in the framework for implementing data governance and data architecture inside an organization. In fact, there are so many different organizations, each with its own culture. There are many alternative ways Data governance and Data Architecture can be applied. Different attributes and components have been presented by various authors and industry experts and it is included in a DG and DA framework. Depending on the demands of the company, different frameworks would be used.

The design of this model is mainly based on the contents of the following papers:

(DAMA, 2017) – [DAMA-DMBOK2 by DAMA International] In the data management world, the handbook is the most well-known and widely utilized. This is the product of the work of a large number of specialists from all over the world. The book also gives an outline of how the Knowledge Areas are related to one another.

(DCAM, 2020) - The Enterprise Data Management Council created the Data Management Capability Model (DCA MTM). It's also based on industry standards. Because it focuses primarily on the demands of financial institutions, the handbook is not well-known. The guide develops the concept of an organizational ecosystem in which various business functions collaborate on data-related issues.

4.1 ArchiMate

For this research, ArchiMate is the enterprise architecture language of choice as it fits with the modeling goals. The ArchiMate architecture modeling language consists of a generic set of core concepts. A brief description follows; for further details regarding each concept and relationship, please refer to the ArchiMate standard specification (TheOpenGroup, 2013).

The core language consists of three main types of elements: active structure elements, behavior elements, and passive structure elements (objects). Furthermore, ArchiMate defines three main layers based on specializations of the core concepts: The Business Layer, the Application Layer and the Technology Layer offers infrastructure services (e.g., processing, storage, and communication services) needed to run applications, realized by computer and communication hardware and system software. The layers, concepts and relationships used in the following model are explained below:

- *The Business Layer offers products and services to external customers, which are realized in the organization by business processes performed by business actors.*
- *A business process - represents a series of interrelated processes that achieve a specific result (Figure 4.1).*



Figure 4.1 Business Process Notation

- *Triggering relationship - A temporal or causal relationship between elements is described using it. (Figure 4.2)*



Figure 4.2 Triggering Notation

- *Serving Relationship* - This represents how one element offers functionality to another (Figure 4.3).



Figure 4.3 Serving Notation

- *Flow relationship* - Flow from one to another (Figure 4.4)



Figure 4.4 Flow Notation

- *Specialization relationship* - Indicates that one element is a subset of another element (Figure 4.5).



Figure 4.5 Specialization Notation

- *Junction* - Relationships of the same sort are linked together using this AND junction (Figure 4.6)

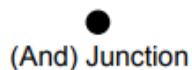


Figure 4.6 AND Junction

4.2 Defining activities of Data Governance and Data Architecture

4.2.1 Data Architecture Activities

To execute operations across the data control environment, DA defines data models, taxonomies, ontologies, data domains, metadata, and business-critical data. The DA function guarantees that data content is controlled, that data meaning is exact and unambiguous, and that data usage is consistent and transparent.

The main activities of DA are:

- ✓ Identify data domains, authoritative sources, and provisioning points, and set them up.
- ✓ Identify and inventory the data required to meet business needs, including all relevant information such as a glossary, dictionary, classification, and lineage.
- ✓ Define and allocate data inventory-related business definitions.
- ✓ Ascertain that DA is integrated into Data Governance (DG).

In order to decrease the complexity of the model and for easy understanding purposes, the DA activities in the model are divided into two categories. The categories are given and explained in the below section.

4.2.1.1 Relationship between identifying and defining the data

i. Identify the Data (Figure 4.8)

Identifying the data includes defining the logical data domains, mapping physical data repositories to the logical data domains, and cataloging the physical data in the repositories.

ii. Define the Data (Figure 4.9)

The following steps involve defining the data, developing and documenting the conceptual and logical models; defining the definitions of data and defining taxonomies to construct relationships between the data.

In short, the ultimate goal of identifying the data is to guarantee that data is properly utilized and everything should be based on how business works in reality. In order to relate logical data domains to physical locations, it is necessary

to map them with the appropriate authorities. Yet, the first action to take is to pinpoint each of the domains. A clear map is needed between logical data domains and their corresponding physical locations. The initial stage is inventorying all of the databases. Basic information (such as source, term name, term definition, field name, and field position) must be recorded for each field in a data domain. The necessary information to open the data will be obtained via basic metadata. So, these are catalogued.

The data that are identified in the first phase and defined in the second phase as models and processes. The first phase plays a critical role in the creation of the second phase and it's the sequence of the first. Once the data domains are identified and catalogued, models need to be designated and defined for all the enterprise data domains. Also, all of these models rely on glossaries as input components. The goal is to reach agreement on the meaning of data terms that are defined in the context of their use. In the same way, defining taxonomies to construct relationships between the data is important. The relationship between both the phases is shown as a model (Figure 4.7).

4.2.2 Data Governance Activities

The Data Governance (DG) component is a set of capabilities that codifies the structure, authority, roles and duties, escalation procedures, policies and standards, compliance, and routines for executing operations across the data control environment. This ensures that all levels of the company make authoritative decisions.

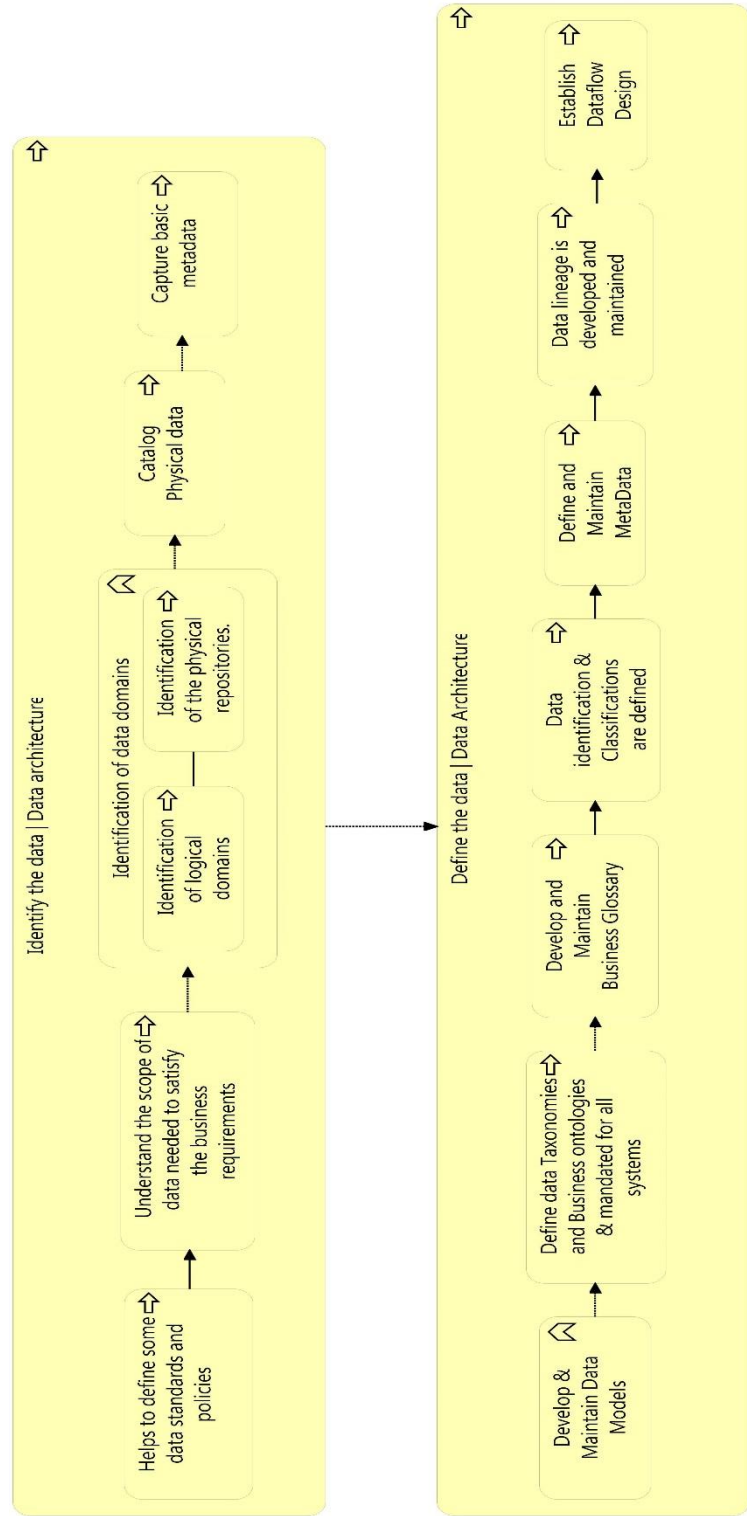


Figure 4.7 Relation between identify and define the data

The main activities of DG are:

- ✓ Within the Office of Data Management, create a data governance function (ODM).
- ✓ Clearly define roles, duties, and accountability for DM resources, including those required by DM policy.
- ✓ Develop and monitor DM policies, standards, and procedures that are comprehensive and attainable.
- ✓ Ensure that data governance policies, procedures, standards, and governance requirements are in line with other departments.

In order to stick to the scope of the thesis, only the activities relevant and aligning to architecture are considered in the model. Other programs such as strategy, issue management, and data management projects, data asset valuations are not taken into account.

The following tables are the descriptions of the main components included in the model. Table 6 describes the component of DA and the Table 7 is about DG.

DA PROCESS	DESCRIPTION
Defining data vision	The first step is to <i>understand the scope of data needed to satisfy the business requirements.</i> (DCAM, 2020)
Identify subject areas	Each subject area is a high-level classification of data that represents a collection of concepts related to a main area of interest to a company. <i>Subject areas can be used to illustrate broad business concepts (customer, product, employee and finance)</i>
Identification of data domains	<i>The data domain is a logical grouping of data.</i>
Develop & maintain data models	A conceptual data model is a logical representation of database concepts and their connections. <i>The goal of</i>

<p>Deliver issues and outstanding questions</p>	<p><i>developing a conceptual data model is to define entities, their attributes, and their connections.</i> The Logical Data Model is used to specify the structure of data elements as well as their relationships. The elements of the conceptual data model are complemented by the logical data model. The benefit of using a logical data model is that it serves as a basis for the Physical model</p> <p>The data modeling process frequently produces concerns and questions that aren't always addressed within the data modeling phase. Furthermore, the persons or organizations in charge of resolving these challenges or answering these questions are frequently located outside of the group that is developing the data model. As a result, a document including the current collection of concerns and open questions is frequently given.</p>
<p>Define data taxonomies and business ontologies</p>	<p>The classification of data into hierarchical groups in order to provide structure, standardize terminology, and popularize a dataset inside an organization is referred to as data taxonomy. This hierarchy is depicted using boxes and lines in the closely related data taxonomy chart, which restricts the data displayed to observation names and available attributes.</p>

<p>Develop and maintain business glossary</p>	<p>Taxonomies are necessary for providing a standard definition and vocabulary for data throughout an organization.</p> <p>A business glossary is a list of terms linked to data. <i>"Business relevant" or "human readable" data is articulated in the glossary.</i></p>
<p>Data identification is defined.</p>	<p>Unique identification can take the form of a customer ID, a legal entity ID, or a product ID. For data aggregation, classification, and analysis, establishing ID methodologies is crucial. Unique identity is a fundamental concept that is quickly becoming a prerequisite for regulatory reporting and risk analysis</p>
<p>Define and maintain metadata</p>	<p>Semantics, taxonomies and ontologies define and relate the content of data in order to enable the organization to realize its maximum value in a consistent and controlled manner. Once the content is defined, it needs to be precisely described as metadata. (DCAM, 2020)</p>
<p>Data lineage is developed and maintained</p>	<p>The capturing of data flow from a source through intermediary systems and data transformations to a final destination or consumer is known as data lineage. Data lineage provides a way to certify that data utilized by customers comes from reliable, authoritative sources, with proper controls in place to regulate system hand-offs.</p>

Establish dataflow design	Data flow is a sort of data lineage documentation that shows how information flows across corporate processes and systems. End-to-end data flows show where data comes from, how it's stored and used, and how it's modified as it goes through various processes and systems. Data lineage analysis can aid in the understanding of the state of data at a certain point in the data flow.
----------------------------------	---

Table 6 DA processes and description

DG PROCESS	DESCRIPTION
Establish DG function	To adopt governance, the organization requires a clear deployment strategy to ensure that the governance structure, processes, and procedures are all in place. The structure is established by creating a Data Management Office first then collaborating with relevant stakeholders to create a comprehensive governance plan. Once defined and documented, the governance structure will be communicated to all relevant parties.

Data principles are established.

At the heart of effective governance are organizational principles. DMBOK defines them as: (DAMA, 2017)

1. fundamental law, doctrine,
2. Any good data governance structure is built on principles.

Data policies and standards are developed, documented, shared and approved by stakeholders.

Data policies are applied globally. They support data standards as well as expected behaviors in major data management and usage areas. (DAMA, 2017) Policies define who is responsible in different situations and what procedures should be followed to manage it. Data Governance policies are evolving documents that must adapt to changing circumstances. While data policies outline what should be done and what should not be done with data, data standards provide more specific instructions on how to do it. Naming standards, data modeling standards etc.,

Business requirements are captured, reviewed and prioritized.

Business requirements are Data management requirements. Any project with a significant data component should capture data

	<p>management requirements. These include architecture, regulatory compliance, system-of-record identification and analysis, and data quality inspection and remediation. (DAMA, 2017)</p>
<p>Govern data domains that are identified and inventoried.</p>	<p>It is imperative that the data domains are identified and inventoried to ensure their proper use in critical applications. (DCAM, 2020)</p>
<p>Apply data governance</p>	<p>Governance is required to enforce in the activities for better decision making. (DCAM, 2020)</p>
<p>Maintain DM</p>	<p>DGC sponsors and approves data architecture artifacts, such as business-oriented enterprise data models (DAMA, 2017)</p>
<p>Document the data in the data dictionary</p>	<p>A Data Dictionary is a collection of structured data items and metadata extracted from the scope of a Data Model or Data Architecture.</p> <p>A dictionary's purpose is to supplement the glossary.</p>
<p>Critical data elements (CDES) have been identified and inventoried</p>	<p>CDEs refer to the individual data attributes that are used to support critical business functions. CDEs must be identified and catalogued to</p>

	ensure evidence of proper sourcing, lineage and usage. (DCAM, 2020)
Implement concrete measures to establish compliance	Establish the ability to respond efficiently and consistently to regulatory requirements.
Define data sharing agreements	Data sharing agreements are corporate documents that specify the data, where it is stored, who is responsible for its protection, and who has access to it.
Review and approve data architecture	It is DG who reviews approved most of the Data architecture activities and it is one of the primary processes of DG

Table 7 DG processes and description

4.3 Alignment Model

Let's see how the process of DA & DG are aligned with each other.

Figure 4.7 and Figure 4.8 demonstrate the activities aligning and the relationships between each other. Note that the activities selected in the model are not organized by sequence. But try to order in the manner in which it makes sense.

Note: For the understanding purposes of DA, these **are elaborated below in red** and **DG in green color**. The areas where the activities are aligned are given below in Purple color.

i. Identify the Data

Establish DG Function: Stakeholders must establish and approve the DG function strategy and methodology, which includes the design of governance organization structure, and DG processes in alignment with DM Policy and Standards.

- The Data Governance Body is responsible for creating the structure. The Data Governance Body includes

Data Principles are established: The design of new data services, updates to current datasets, and impact analyses of bigger internal and external changes should all be guided by principles.

Data Policies & standards: The policy and standards are developed, Documented, Shared and Approved. The policies and standards must be auditable and align with cross-control function policies and standards. It should reflect the basic Data principle.

- Data management experts, business policy staff, or a combination of both draft these. Principles, Policy and Standards are reviewed and refined by data stewards and management. Final evaluations, amendments, and adoptions are then carried out by the Data Governance Council.

Helps define some data standards and policies: The DGC may interact with the DA steering committee to develop some of the policies and standards.

Understanding the scope of data: In order to satisfy the business requirements, understanding the scope data is the first step in DA which will eventually push to the creation of two following activities.

- Data Manager, Data Architect and Enterprise Architect are responsible for this activity and together they are called Data Management Professionals.

Identify subject areas: The subject area is the high-level classification of data which helps in the creation of the other models as they are built up by the combination of Subject area models.

Identification of data domains: The business must lead the definition of logical data domains by determining what data is required to accomplish the relevant business functions and must be identified, documented, inventoried and authorized. Physical data repositories of data to be mapped to the logical data and it should be documented and inventoried.

Catalog Physical data: The next stage is to catalog the physical data in the repositories after the physical repositories of data aligned to the data domains. Create a data element catalog that is aligned with the data domain.

Capture basic metadata: Basic metadata to be captured for data aligned to a data domain. This metadata will be required by any data consumer, particularly data analytics consumers, as part of their discovery process prior to defining data for production use.

Govern data domains: Governance is enforced to ensure the data consumer gets the ultimate use of the organization-wide authoritative data domains.

Governance to control the access & use of the data: Applying governance is essential to restrict data access and usage once authoritative data domains have been established. An understanding of who is utilizing the data and for what purpose is necessary for appropriate usage.

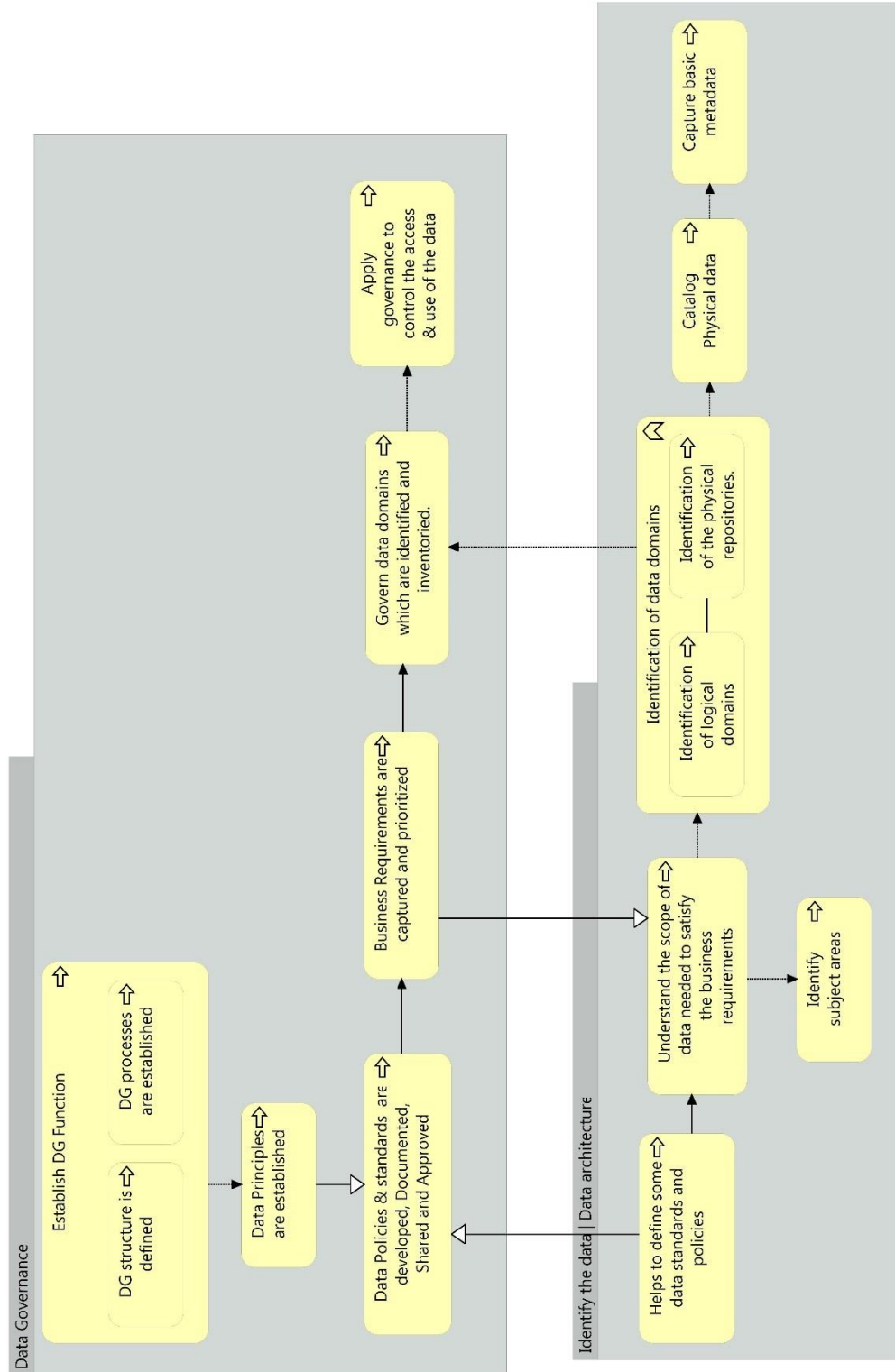


Figure 4.8 Identify the Data

ii. Define the Data

Develop & Maintain Data Models: All enterprise data domains must have conceptual and logical models created and documented. Model alignment must be mandated by policy and included into enterprise change management policies and needed to consult with relevant stakeholders to confirm conceptual and logical data models.

Maintain DM: Governance is required to enforce the publishing and use of standard data models. DG is responsible for creating standards for data modeling, maintaining the quality of data models, version controls and reviewing it.

Deliver Issues and Outstanding questions: A document is delivered that contains the current set of issues and outstanding questions. The data modeling process frequently produces concerns and questions that aren't always addressed within the data modeling phase.

Manage and resolve data related issues; coordinating resolution of data-related issues raised by the DA Team.

Define data taxonomies and business ontologies: Define data taxonomies and business ontologies as the next stage in addressing data architecture. Hierarchical relationships are defined by taxonomies. Taxonomies are essential for providing a standard definition and terminology for data across the company, as well as ensuring that the data is used correctly. Taxonomy must be created and maintained for each officially designated data domain that is identified, inventoried, and deemed critical.

Develop and Maintain Business Glossary: Non-technical descriptions of data attributes realities must be produced as business definitions. Need to assign authorized business definitions to defined taxonomies.

Govern and ensure semantic definitions and taxonomies: Governance is required to enforce definitions and taxonomy schemes and needs to be ensured they are properly assigned and maintained.

Define data identification & Classifications: After the taxonomies to achieve a correct data organization, data identification, classification schemes and procedures must be applied. For critical business elements identifiers must be defined.

Governance glossaries, identifiers, classifications: Governance is required to enforce the definition, approval and publishing of identification and classification and needs to be ensured they are properly assigned and maintained.

Define and Maintain Metadata: Taxonomies and ontologies define and relate data content so that an organization may get the most out of its data in a consistent and controlled way. Once the content has been defined, metadata must be defined to correctly describe it.

Implementing concrete measures to establish compliance: Monitoring and enforcing contractual restrictions on third-party data entering the business requires governance. Apply governance to the record of data use limitations as part of the data's metadata to monitor the restrictions and the metadata should indicate the restrictions on its use, ideally linked to a description of the restrictions.

Data lineage is developed and maintained: Data lineage is developed and maintained. Lineage is important to capture during data modeling as it shows where the data comes from.

Establish Dataflow Design: Data flow map and document relationships between data and business roles, depicting which roles are responsible for creating, updating, using, and deleting data (CRUD).

Data lineage has been documented and validated - Lineage tracking: Even though data lineage is developed and maintained by DA, after the lineage tracking Data lineage is documented and validated by DG as it can help explain the state of data at a given point in the data flow. It helps to find the Critical Data elements.

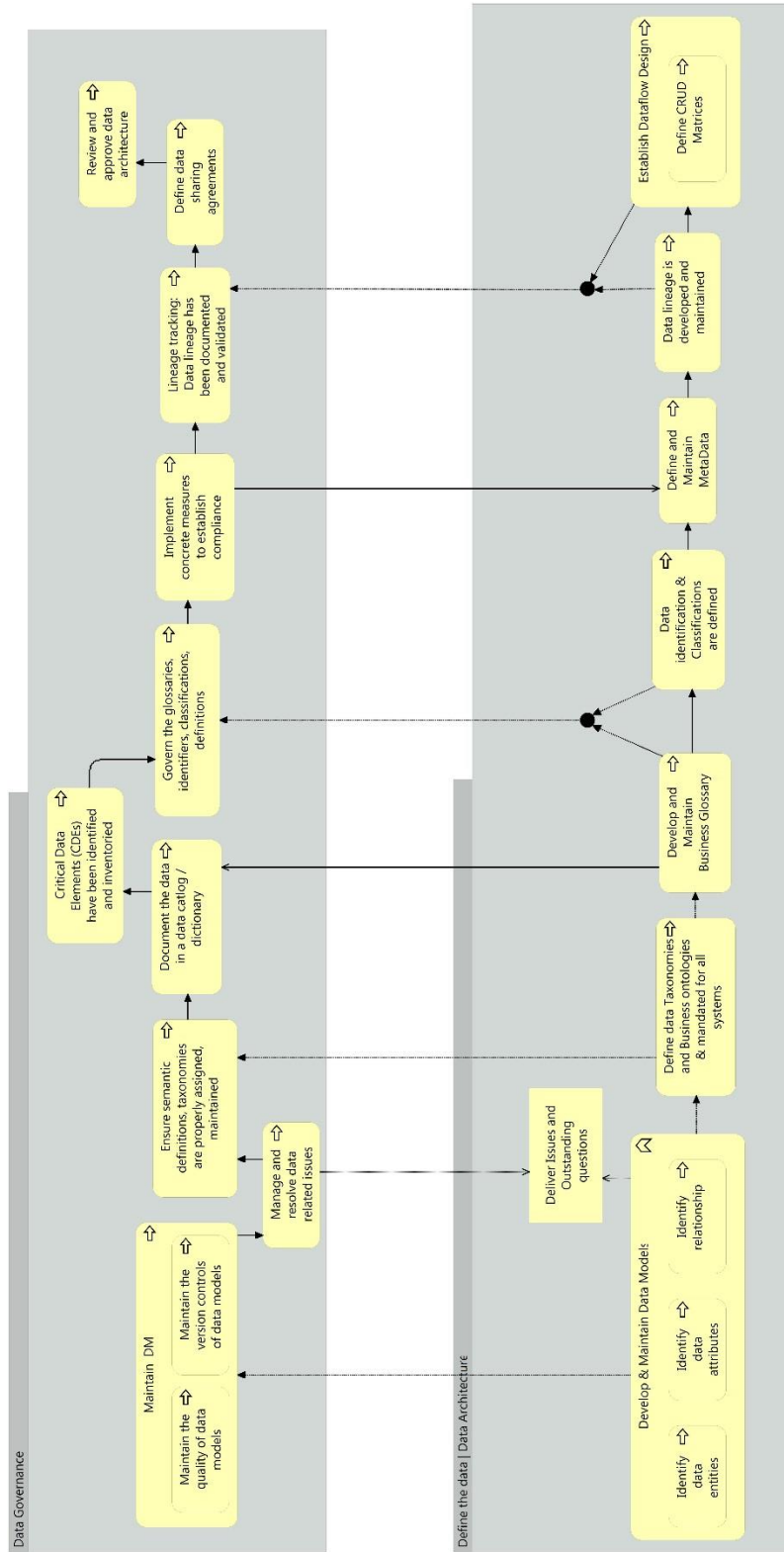


Figure 4.9 Define the Data

Define data sharing agreements: Data governance teams have a significant role to play in setting up data sharing agreements. They must ensure that legal and compliance teams sign off prior to any movement of personal data

Review and approve data architecture: In almost all DA activities, DG will review it and then approve architecture.

4.3.1 DA and DG processes with relevant tasks and roles.

In the table below (Table 8), I have linked DA and DG processes with relevant tasks and roles. The RACI model has been used to highlight possible accountabilities and responsibilities of the relevant roles. Please note that the accountabilities and the responsibilities change according to the company's situation.

Note: For understanding purposes, DAs are elaborated below in red and DG in green. The areas where the activities are aligned are given below in Purple color.

DAMA-DMBOK2 recommends separating data management specialists from business professionals. *Data managers, data analysts, data architects, and other data management professionals fall into Data management professionals. Business professionals, business stakeholders and IT stakeholders are included.*

Which governmental bodies you require will be determined by the type of governance structure you use. *The Data Governance Steering Committee, Data Governance Council (DGC), Data Governance Office (DGO), Data Stewardship Teams, and Local Data Governance Committee are all common Data Governance bodies (DAMA, 2017)*

Task	Roles
------	-------

	Business Professionals	Data Management Professionals	Governmental Body
<i>Establish DG Function</i>			R
<i>Establish Data Principles</i>	A	R	R
<i>Define Data Policies & standards</i>	A	R	C
<i>Capture Business Requirements</i>	A	R	C
<i>Understanding Data Requirements</i>		R	
<i>Identification of data domain</i>	A	R & A	
<i>Governing data domains</i>			C
<i>Apply governance to control the access & use of the data</i>	R	A	C
<i>Catalog Physical data</i>	R	R & A	
<i>Develop & Maintain Data Models</i>	A	R & A	
<i>Maintain DM</i>	A		C
<i>Manage and resolve data related issues</i>	R	A	C
<i>Define data taxonomies, business ontologies, Data identification &</i>	A	R & A	

<i>Classifications</i>			
<i>Develop and Maintain Business Glossary</i>		R & A	
<i>Governing the glossaries, identifiers and taxonomies Classifications and definition</i>			C
<i>Document the data in a Data Catalog / Dictionary</i>		R & A	
<i>Define and Maintain MetaData</i>		R & A	
<i>Implement concrete measures establish compliance</i>	R	R	C
<i>Develop and maintain Data Lineage</i>	R	R & A	
<i>Document and Validate Data Lineage / Identify Critical Data Elements - CDEs</i>	R	A	C
<i>Establish Dataflow Design</i>	R	R&A	
<i>Define data sharing agreements</i>			R & A
<i>Review and approve Data architecture</i>			R & A

Table 8 Responsible people for the task

4.4 Summary and Conclusion

The aim being to construct another model that allows it to be combined and complimented with both data architecture and data governance is achieved. The answer to two sub-questions SQ2 & SQ3, the areas that overlap and how the achievement is made can be extracted from the model. The alignment model was designed with the use of literature review and mainly based on the contents of the two books that are widely used for Data management DAMA (DAMA, 2017) and DCAM (DCAM, 2020). The model gives clarity on how the areas are aligned and why these are the areas that are aligning are described in the design section. Also, the people who are responsible for the tasks are also given in

Table 8. However, it is best to take into account that the description of the alignment model was provided with adjustments incorporated from the next Evaluation and Validation phase. In order to determine the correctness, quality, and understandability of the alignment model it is evaluated through a semi structure interview with an expert and it can be found in the next following chapter.

PART III

Validation of the research

5 Model Evaluation and Validation

This chapter will cover the evaluation and validation phase. The developed alignment model is validated and evaluated. The evaluation of the alignment model is performed using the validation method – ‘expert opinion method’ mentioned by (Wieringa, 2014). Section 5.1 will discuss the evaluation approach. The model has been adjusted as a result of the feedback received during the evaluation phase and the final version of the model is provided in this section. Opinions are collected and documented in this chapter. After that, the model is validated using a case study to observe the usefulness of model in a practical setting and it can be seen in the section 5.2.

5.1 Evaluation approach

In line with (Hevner, March, Park, & Ram, 2004) this section is **concerned with determining the correctness, quality, utility, efficacy and understandability of the designed model. This is accomplished by qualitatively evaluating what an expert thinks of the model** (as described in Chapter 4). It is also important to understand what aspects of the created model need to be focused on, modified, or removed. In order to have a more accurate result, the correctness of the model should come from an additional assessment. In section 2.3.3 it was made clear that the alignment model was evaluated by conducting a semi-structured interview. This open and flexible approach benefits both the interviewer and the interviewee because they can talk about the alignment even more freely by continuing to share new ideas. The evaluation interview serves to obtain perception and interpretation regarding the developed model. The model has been adjusted as a result of the feedback received during the evaluation phase.

5.1.1 Selection of the Interview Questionnaire and the material used

Due to time limitations, one interview was conducted to evaluate the developed model. The interviewee was chosen based on their experience in management or governance in any organization. One candidate was chosen and contacted, who

works for an external company as a Principal Information Management Consultant having a lot of experience and knowledge about this subject. Since it is a semi-structured interview questionnaire, it is not the best compilation because that word connotes a fixed instrument to be read verbatim, rather than the flexible, interactive approach of SSI questions (Adams, 2015). Instead, an agenda for the interview guide, an outline of the topics, models and questions to be addressed was created. The interview questions asked were more focused on the correctness, quality, understandability and efficacy of the proposed alignment model. Interviews were conducted for about 62 minutes, one-on-one sessions, in which models were given to the interviewee, explained and where multiple open questions were posed regarding the problem relevance, alignment, practicality and implementation relevance. The agenda for the interview guide is mentioned in [Appendix B](#). The questions asked are attached in the [Appendix C](#).

5.1.2 Profile of the interviewees

The overview of the evaluation interviewee is shown below in Table 9. The interviewee Mr. Nigel Turner specializes in Information Strategy, Data Quality, Data Governance, and Master Data Management and has over 20 years of experience in Information Management (IM). He built and led large IM and CRM consultancy and delivery practices in numerous consulting businesses, including the British Telecommunications Group (BT), IPL, and FHO. Nigel has also worked as the Vice President of Information Management Strategy at Harte Hanks Trillium Software, a major global provider of Data Quality and Data Governance solutions and consulting, where he worked with over 150 customer businesses from all over the world. Nigel is a well-known thought leader in the field of information management, having spoken at multiple international conferences and written numerous white papers and blogs on the subject. Nigel is an active part of DAMA International's mentorship program, which helped to establish. He has lectured on Data Governance at Cardiff University and is an active member of DAMA International's mentoring program, which he helped to establish.

Person	Organization	Function & knowledge	Date	Type of contact
Interviewee Mr. Nigel Tuner	Global Data Strategy Ltd	Principal Information Management Consultant, Vice-Chair of the Data Management Association (UK), an active member of DAMA International's mentoring program.	28/09/2021	Zoom Video call

Table 9 Overview of evaluation interviews

5.1.3 Evaluation results and final version of the alignment model

The evaluation interview yielded very positive results, which met the model's requirements. According to the evaluation questions, the generated alignment model is presented in a logical and clear manner. It satisfies the standards for coherence and comprehensibility. The generated model gives a broad overview of the Governance and architecture implementation especially the areas they have to align in the company. The developed model, according to the interviewee, makes sense and it can be used in the organizations for better decision making. It may be claimed that the model could be customized based on development approaches or the extension of key elements from an organizational point of view.

However, the main limitation discussed during the interview is that there that is only limited academic literature is present about data governance and alignment. The interviewee agreed to it, adding that data governance is not a well-researched topic, he is also researching and building a new data governance framework for the organization. Also, the interviewee mentioned that it would

be good to do a piece of research in the area that needs it. Interviewee provided three suggestions for improving the established model, which are outlined below.

5.1.4 Suggestions and feedback gathered from the interview

Regarding the first step in the data governance of the first phase of the model Interviewee outlined that data governance structure is defined and the processes are established and then the business requirements capturing are not in the correct order. In his experience, he said establishing the DG function is not the first thing one would normally do in a Governance Program. The first thing you need to do is really understand the business's need for data. Then establish the data principles, then you then might develop your data policies. Once the principles are established, the link between the business and the data is made and how it's going to be governed, so capturing the business requirement would come before establishing principles.

Next, in the second phase define the data Figure 4.9 the activities such as defining the taxonomies, identification and classification, Interviewee think that some would argue that those all are really metadata activities. Defining the metadata activity encompasses all those things. He suggested to group those as Metadata activity.

Finally, the interviewee made one last comment to the model as to the way it was laid out. He mentioned that the model makes logical sense, but as a point to remember, these are continuous cycles of improvement. The activities in the model are not done once and moved on. In organizations we don't develop or maintain neglect of a Business Glossary. The glossary is a living document. when you review and approve the data architecture, that's not the end of it. If a business decides to go into a new line, the organization thinks that they need to change the diagram significantly, but somewhere in there, they need to recognize at least anyway this isn't a linear process. So, every time a new this comes, they need to update everything, like updating the model, definitions and also the

dictionary. The interviewee illustrated with a real example why the processes are continuous cycle of improvement which can be found in the [APPENDIX C](#).

Final version of the alignment model

The model has been adjusted based on the results of the feedback received during the evaluation phase. Figure 5.1 and Figure 5.2 present the final version of the alignment model. The following are the changes that were made:

- ✓ In the first phase of identifying the model, the business requirement is moved as the first activity in data governance because businesses need to understand and capture the data before defining the policies. Based on the captured data, the data principles and policy will be established aligning it along with the business needs.
- ✓ The activities in the data architecture of the second phase such as defining the taxonomies, business ontologies, identification and classifications and delivering glossaries are all grouped into metadata activity.

The point to remember is to note that the processes in the model are not linear, it is a continuous cycle of improvement.

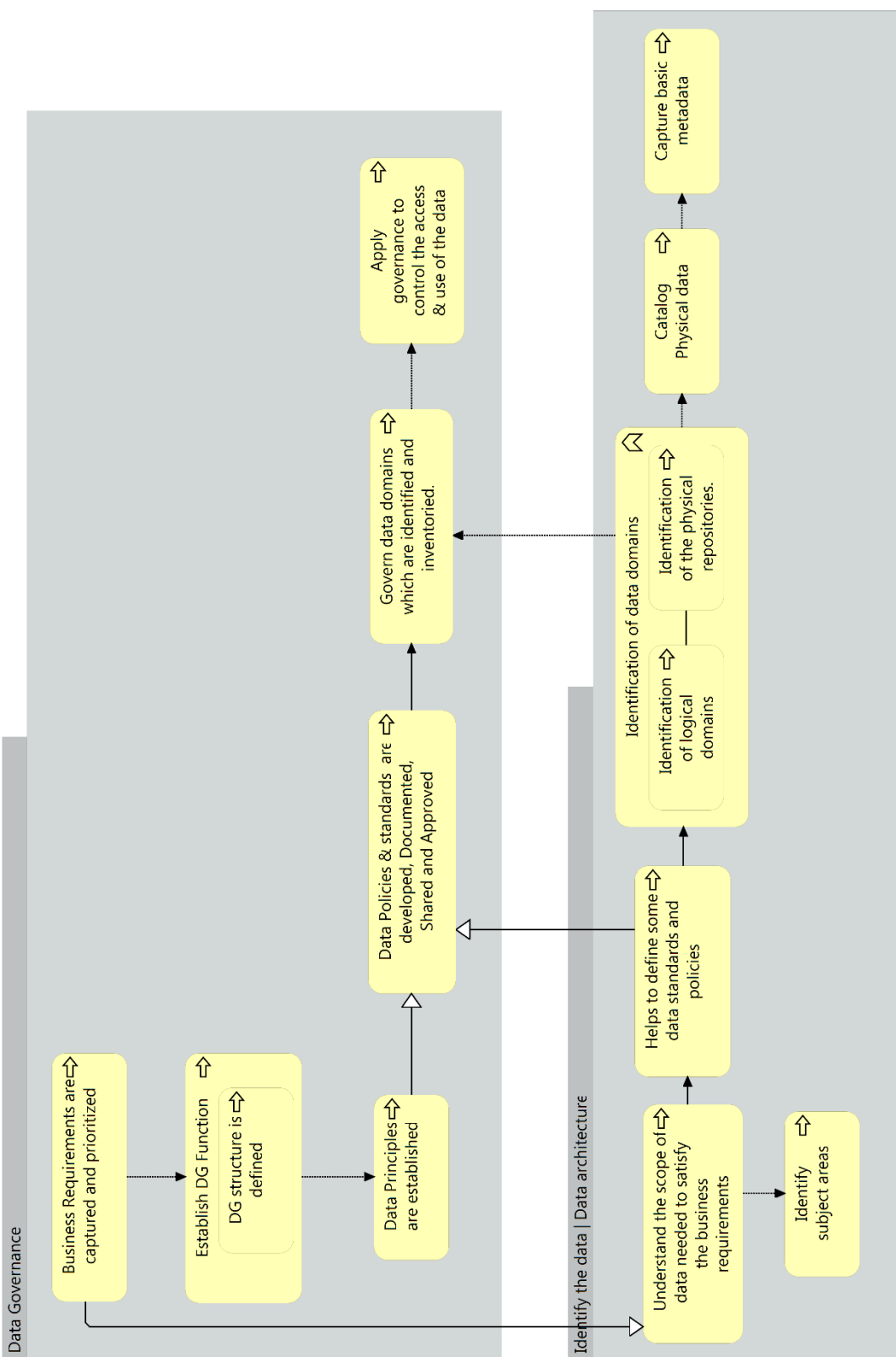


Figure 5.1 The final version of Alignment model - Identify the data

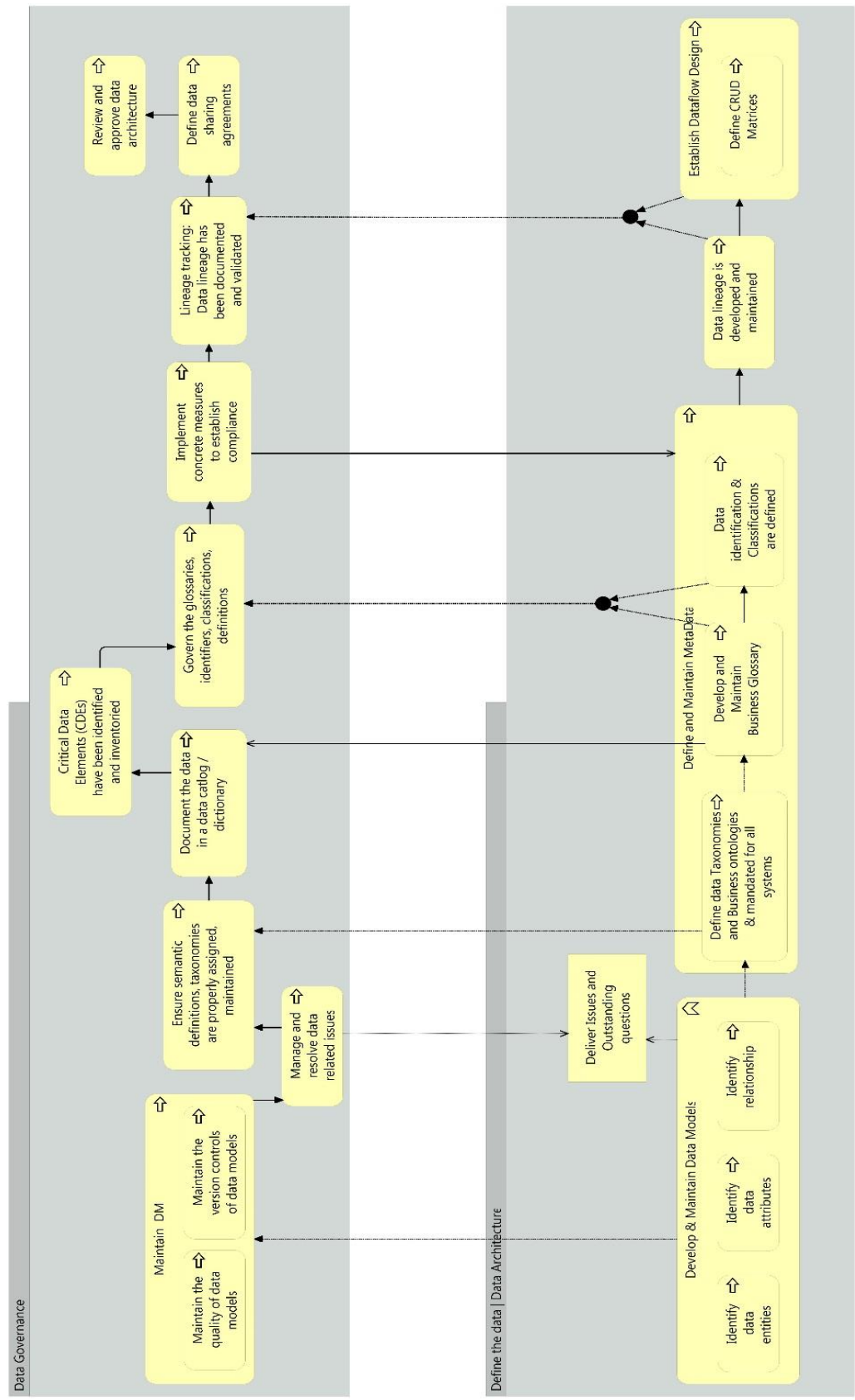


Figure 5.2 The final version of Alignment model – Define the data

5.1.5 Evaluation Conclusion

The alignment model of data governance and data architecture have been evaluated via interview with an expert. With the gathered feedback, the model has been adjusted and corrected accordingly and the final version of the developed alignment model is presented (See Figure 5.1, Figure 5.2). According to the expert, the model has been proven to be presented in a clear and logical manner for both technology-oriented and business-oriented audiences. The essential components stated in the model can be used as a template and guide for new governance and architecture team adaptors. Moreover, the evaluation interview yielded very positive results, which met the model's quality and correctness.

5.2 Case study

A case study was performed to observe the usefulness of the developed alignment model in practice. For the case study, the company Aurelius Enterprise provided internal documents of their ongoing project of company Y. The documentation contains data governance and data architecture processes of the company Y. The details are confidential and so was not briefed.

Case studies are observed using documentation from DA and DG of the organization to identify how much of the developed alignment model is followed in of the real organization and how to know what they are missing to align. The documentation of DG contained the processes of how organizations are created and changed in the HR system called XYZ in a structured and standardized manner. These processes describe the regular creation or change of organizations a part of XYZ business requirements. It provides a clear description of the governance processes of the operating model. The other DG document is their data dictionary, which contained the following information of version control, meta model & glossary, conceptual data model, data domains, data entities, data attributes, data quality rulebook, data quality glossary. In the DA document, the description of the High level solution for the processes around the AB data set on the data management platform was given along with the document describing

the business requirements and business scenarios as input for the High Level solution document. The benefits of using the model are given. Then usefulness was measured by quantitatively validating by a panel of experts.

5.2.1 Case Description

The department X wants to publish data regarding the workforce of Y on the Data Management Platform. The dataset is called the AB and it is provided by X from their HR system called XYZ. The dataset which is shared is of about every employee. The data's such as employee number, job name, contract type, location, payroll and so forth. AB datasets are published under data governance and synchronization with the Data Management Platform is a continuous process that is triggered daily i.e. The data synchronization process between XYZ and Data Management Platform is triggered daily. And the flow is from XYZ, then the data goes into confluent Kafka and from there into Data Management Platform which is based on Elasticsearch.

There are three teams involved, the team doing governance, the team doing the interpretation of the data flow and the team who are owning the XYZ, the HR system.

For Data architecture, the technical description of how the data flows from this point to that point and how it was implemented was given. The business requirements for the creation of dataset AB and business scenarios document were also provided for the study. Data dictionaries containing version control, data domains, conceptual data models, meta models, entities and attributes and their descriptions were given by the governance team. Also, a clear description of the governance process of the Operating Model in order to maintain and safeguard the organization structure and building blocks of the operating model was provided. It is to be noted that the Data Management Platform is still in a project stage and development; governance, policies and processes are under construction.

5.2.2 Case Observation

In order to know how much the approach of the model is followed in the case, following are the observations made and the benefits of using the model: what would they gain in a practical setting if they followed the approach as per the alignment model given below.

Number of aligned areas identified from the case:

Data Principles: Data principles are established by the governance team for the Operating model. The most important principles for the registration of the HR system are also identified by Governance. They check whether the data is in line with the Personnel system principles. Any deviations between guidelines and principals will be investigated and discussed with EA, approach or proposal and recommendations will be offered and discussed. The AB dataset should be managed under data management principles, such as data governance and architecture.

Data standards: For setting data standards, data governance is responsible and we regularly maintain it. A standard check is performed by the data owner of AB data between the personnel system and the planning system to ensure completeness and quality. This is checked jointly (a department and all personnel administrators) to determine how the quality process of the AB data can be safeguarded.

Business Glossary & Data dictionary: Glossaries are created in the creation of AB data and it is published. As we know, Data dictionary is a collection of structured data items and metadata extracted from the scope of a Data Model or Data Architecture. A dictionary's purpose is to supplement the glossary and it is maintained and delivered by the data governance.

Data identifiers: Unique identification can take the form of a customer ID, result ID, or a Business rule ID. For data aggregation, classification, and analysis, establishing ID methodologies is crucial. Unique identity is a fundamental concept

that is quickly becoming a prerequisite for regulatory reporting and risk analysis and it is delivered by DA in this case.

Mis-aligned areas identified from the case:

Data Domains: There is a domain expert but it is unclear in which department the domain expert is part of. Meaning whether he belongs to the DA team or an individual is unclear. However, data domains are governed by the governance teams as the data domain is found in the data dictionary which is delivered by the governance.

Data Model: Since the case is not at the starting point of the creation, it is unclear who created the model, but it could be predicted that it must be from the DA team based on the facts mentioned in the case document that knowledge of the data and data model lies within the Project Overhead team. However, the conceptual model, data entities, attributes and version controls are governed and maintained by the governance team.

Data Classification: Classification cannot be found on who is delivered but it can be found in the data dictionary so it can be stated that it is governed by the Data Governance.

Data lineage & Data flow: Data lineage provides a way to certify that data utilized by customers comes from reliable, authoritative sources, with proper controls in place to regulate system hand-offs. Data flow is a sort of data lineage documentation that shows how information flows across corporate processes and systems. End-to-end data flows show where data comes from, how it's stored and used, and how it's modified as it goes through various processes and systems. The lineage in this case is not created manually, but it is created during infrastructure code and the information goes into their governance tool which is Apache Atlas. This was demonstrated during the discussion of the case study description. According to the designed alignment model, DA creates the data lineage and the lineage tracking is maintained and validated by DG.

After observing the case, it can be seen that some of the areas are not aligned in the case. Data architecture is driving to some extent how you structure data governance, but data governance is also driving how you structure the data architecture that can be observed from the case study. It also observed that these two things have to be done in parallel.

Usually, the alignment is implicit and it can be captured that if you implement data governance and try to implement a data architecture, then the two things support and reinforce each other. However, using this alignment model will bring the two parties together and make it more aligned.

Benefits of using the alignment model in this case:

As mentioned by the author (Sherman, 2015) about the benefits of data architecture that it helps to gain a better understanding of the data, provides guidelines for managing data from initial capture in source systems to information consumption by business people, provides a structure upon which to develop and implement data governance and helps with enforcement of security and privacy.

However, from the case study, it is observed and understood that data governance relies on creating a standardized data architecture plan that serves as the foundation for layering data policies to ensure usability, quality, and consistency.

In this case we have DA groups of people and DG groups of people do not necessarily align themselves. The model shows the relationship between the two and how working closely together will help them both achieve their goals. to show how this model will be beneficial. So together the proposed aligned model will provide the following benefits in the case. The benefits do not necessarily only apply to in this case, they can be considered as general benefits of using this alignment model. These benefits are then qualitatively evaluated measuring the extent to which the alignment model meets the benefits.

The benefits are:

Maximize and increase in speed to make better decisions (B1) - As a result, having team members who understand and are aligned with the company's main objective is becoming increasingly vital. This model will increase speed in decision making. The alignment model clarifies who in the organization has decision-making authority over each work activity. Organizations that are aligned make better judgments and take less time to execute. Employees can also focus on doing their jobs rather than figuring out who is required to do what and who has to be consulted when making decisions with the help of an alignment model.

Increase accountability (B2) – In the proposed model, the architecture can help governance decide what the critical data is by delivering the issues and questions. There is then governance that will help architecture by providing accountable people who can own the architecture with its development.

Make data accurate, complete and consistent across the organization (B3)- If you build a DA, from a technical perspective who is accountable for data? The answer frequently is no one. Whereas, if you have governance in place and you have clear definitions and stewards then the architecture is owned by somebody or it's accountable for its maintenance.

Reducing the likelihood of errors being introduced (B4), and ensuring that data is established, implemented, and governed (B5) - A lot of the problems that are seen in organizations with data architecture is where you don't have an alignment with governance. Technical people produce models and all the rest of it, but there is really no relationship to the business. That's why these models never get implemented properly. Data architecture is a discipline with a lot of deliverables, sitting on a shelf somewhere, and the key thing about data governance is that it gives data architecture, a channel into the business and ownership of the data, in order to make sure that the architecture is implemented, rather than it sits on the shelf as those two things have to come together. The model has captured that pretty well.

It also helps in the visibility of resources. Resources in an unaligned organization are frequently buried in shadow organizations. However, if it is aligned using the model, the business knows what resources are available and can redeploy redundant resources to challenges. Finally, the organization has less duplication and more resources to boost the company's performance and productivity.

To measure the extent to which the proposed alignment model meets the benefits, it was validated by a panel of experts. The validation approach, measurement design and the results are given in the following sections.

5.2.3 Validation approach

Following that, the above benefits are to be validated in order to provide a concrete conclusion that this alignment model is beneficial. The evaluation of the foresaid benefits of alignment model is performed using the validation method mentioned by (Wieringa, 2014) – 'expert opinion method'. A panel of experts was chosen related to the field and introduced to the alignment model as described in Chapter 4. A questionnaire based on the Unified Theory of Acceptance and Use of Technology (UTAUT) was created to collect opinion from experts. The UTAUT model is a unified theory that consists of four determinants that seeks to predict usage behavior and user acceptance: Performance Expectancy, Effort Expectancy, Social Influence, and Facilitating Conditions are all factors that influence performance (Venkatesh, Morris, Davis, & Davis, 2003). The questions are constructed based on Performance Expectancy and Effort Expectancy determinants Figure 5.3 . The key reason why other determinants are not used is because others are not related to the foresaid benefits. Furthermore, the main moderating factors (gender, age, experience, and willingness to use) will not be included because they are not determinant variables of user acceptance and behavior in this study. Due to the short duration of validation, none of the participants participated in the 'hands on' activity, therefore it was based on observation rather than actual use.

The experts that agreed to take part in the validation process are:

- Technical Designer and Architect at the Case Company, who is responsible for Industrial Digital Workflow Solutions.
- Business Engineer at BNR Quality
- Lead Data Architect at Miraclon company
- Principal Consultant Enterprise Architecture at Sogeti
- Senior Consultant at Deloitte.de
- Senior Analyst at BKI Production company.

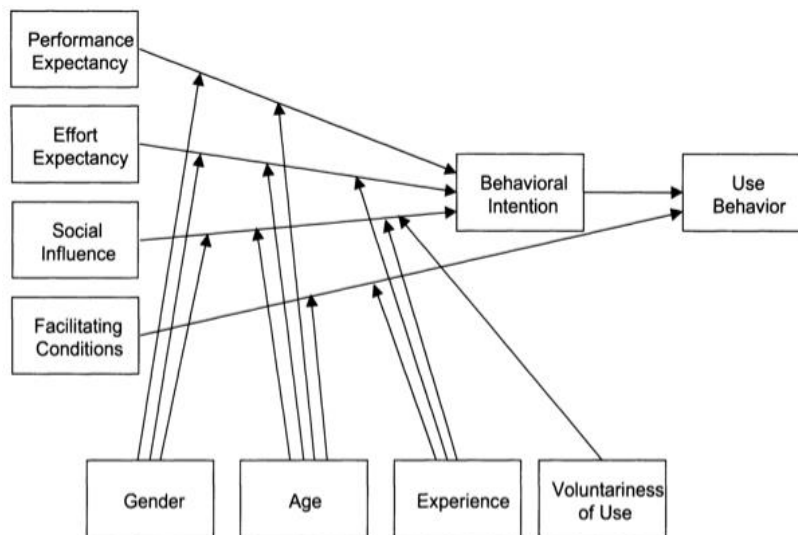


Figure 5.3 UTAUT Research Models (Venkatesh, Morris, Davis, & Davis, 2003)

As previously said, the expert panel is diverse. The main reason for choosing a diverse panel is that the benefits mentioned do not necessarily only apply to this case study of the research, they can be considered as general benefits of using this alignment model. However, selected professionals are generally familiar with data architecture and governance. We recruited a varied panel of experts to conduct an extensive evaluation of the designed alignment model, ensuring that the model is validated both within and outside the case study.

The constructs from the UTAUT model that were chosen to be answered in the questionnaire and the measurement design are given in the next section.

5.2.4 Measurement Design

The questions corresponding to the benefits and along with the constructs are given below in the given to the participants as a form of a questionnaire and listed out under the intention to measure the response from the participants in relevance by measuring the benefits of the alignment model. The questionnaire was distributed through an online form. The responses were then exported as a CSV file and processed in a spreadsheet. Furthermore, in regards to allowing the participants to elaborate on their opinion, for each question, the participants could also give their positive and negative opinion and as much detail as they desire. The outcome of the evaluation along with the opinions are attached in the [appendix D](#).

NO	Benefits	Construct	Definition	Questions	Score	opinion
1	Maximizing the use of data to make decisions (B1)	Performance Expectancy (PE)	The degree to which an individual feel that by utilizing the system, he or she would improve job performance	To what extent does the proposed alignment model contribute to better decision making?		
2	Increase accountability (B2)			To what extent does the alignment model contribute to increasing accountability?		
3	Create data that is accurate, complete and			To what extent is the alignment model		

	consistent across the organization (B3)			bring standardization and consistent data across the organization?		
4	Reduce the likelihood of errors being introduced (B4)	Effort Expectancy (EE)	degree of ease associated with the use of the system.	To what extent does the proposed alignment model help in the duplication of data and resources?		
				To what extent does the proposed alignment model help in reducing the likelihood of errors being introduced?		
5	Ensure that data is established, implemented, and governed (B5)			To what extent can the proposed alignment model contribute to creating a standardized data architecture plan that serves as the foundation for		

				layering data policies to ensure usability, quality, and consistency.		
--	--	--	--	---	--	--

Table 10 UTAUT constructs used in the questionnaire

Additionally, to quantify the amount of participant approval, a Likert Scale from 1 to 5 is utilized, with a score of 1 indicating "benefit is not agreed" and a score of 5 indicating "completely agreeing to the benefit". The following values were used: 1 – strongly disagree, 2 – disagree, 3 – neutral, 4 – agree, and 5 – strongly agree. The results table color-codes the various values. Moreover, we calculate the median and standard deviation for each question. The median value is calculated in order to determine the most often chosen option. Standard deviation is used to quantify the dispersion of values within a set of data values. The questionnaire responses are included in [Appendix D](#) of this report.

5.2.5 Analysis and Results

Table 11 summarizes the results of the questionnaire. The alignment model is evaluated positively with an average score of 3.71 which can be considered relatively good. The mean depicts the overall trend and standard deviation depicts the dispersion between the participants' responses. The mean for each question ranges from 3 to 4.16, with the lowest mean value being noted for Q4 and the greatest mean value being noted for Q5. The lowest and the highest mean are the questions from the same benefit (B4) - *Reduce the likelihood of errors being introduced*. Two questions were constructed for that particular benefit as you can see in Table 10.

If a standard deviation of 0 indicates strong agreement among respondents, then more than one indicates that there is a significant variety in their responses. So, in this case, the obtained standard deviation ranges from 0 to 1.1 then the acquired range of standard deviation values below show that there is a relatively

good agreement among the participants in approving the benefits of the alignment model except one benefit 1 (B1). However, we noticed that the highest dispersion among the answers exists for benefit 5 (B5).

NO	Construct	Questions	A	B	C	D	E	F	AVG	STD DEV
1	Performance Expectancy (PE)	To what extent does the proposed alignment model contribute to better decision making?	4	4	4	1	4	4	3.5	1.118
2		To what extent does the alignment model contribute to increasing accountability?	4	5	4	5	4	4	4.3	0.471
3	Effort Expectancy (EE)	To what extent is the alignment model bring standardization and consistent data across the organization?	3	4	4	2	4	3	3.3	0.745
4		To what extent does the proposed alignment model help in the duplication of data and resources?	4	3	3	2	3	3	3	0.534
5		To what extent does the proposed alignment model help in reducing the likelihood of errors being introduced?	4	5	5	5	4	2	4.16	1.067
6		To what extent can the proposed alignment model contribute to creating a standardized data architecture plan	4	4	4	4	4	4	4	0

		that serves as the foundation for layering data policies to ensure usability, quality, and consistency.								
AVG score									3.71	0.65

Table 11 Questionnaire results

Performance expectancy (PE)

The determinant of 'performance expectancy' reveals whether an individual would increase his performance in his workplace if he used the offered strategy (Venkatesh, Morris, Davis, & Davis, 2003). Each expert's evaluation for the Performance Expectancy is given in Table 12. Overall, all of the experts agree on Benefit 1 (B1) except for one expert. He strongly disagrees with the opinion that states that in order to get better decision making, these two disciplines are not enough. They need inputs from strategy, portfolio management, etc., For the second benefit there is a relatively good average that is to be noted.

NO	Construct	Questions	A	B	C	D	E	F	AVG
1	Performance Expectancy (PE)	To what extent does the proposed alignment model contribute to better decision making?	4	4	4	1	4	4	3.5
2		To what extent does the alignment model contribute to increasing accountability?	4	5	4	5	4	4	4.3

AVG PE														4.9
-------------------	--	--	--	--	--	--	--	--	--	--	--	--	--	-----

Table 12 Questionnaire Results of PE

Effort expectancy (EE)

The factor 'effort expectancy' reflects the ease with which the approach can be used (Venkatesh, Morris, Davis, & Davis, 2003). Each expert's evaluation of the Effort Expectancy is given in Table 13. This determinant has mixed reactions according to the foresaid benefits. For the third benefit (B3) (Q3) there's is relatively less average with only 3 people agreeing out of 6. The most noted opinion for this benefit is that the model itself does not make data better. It is the use of the processes that could. And only if this process is embedded in complete governance and aligned with the change process. If a company decides to implement this model, autonomous and separate from the rest of the organization, it will not be effective and failing its purpose is very likely.' Next, for the fourth benefit- Reduce the likelihood of errors being introduced (B4), two questions were constructed for that particular benefit as you can see in Table 10. The first question supporting that model would help in the duplication of data and resources has the lowest mean of 3. Whereas the question supporting that helps in reducing the likelihood of errors being introduced has the highest meaning. So, for the benefit (B4), we can consider that it only helps in the likelihood of errors. Lastly, for Benefit 5 (B5) has the greatest mean value and there is a relatively good agreement among the participants in approving this benefit.

NO	Construct	Questions	A	B	C	D	E	F	AVG
3	Effort Expectancy (EE)	To what extent is the alignment model bring standardization and consistent data across the organization?	3	4	4	2	4	3	3.3
4		To what extent does the proposed alignment model help in the duplication of data and resources?	4	3	3	2	3	3	3
5		To what extent does the proposed alignment model help in reducing the likelihood of errors being introduced?	4	5	5	5	4	2	4.16
6		To what extent can the proposed alignment model contribute to creating a standardized data architecture plan that serves as the foundation for layering data policies to ensure usability, quality, and consistency.	4	4	4	4	4	4	4
AVG EE									3.615

Table 13 Questionnaire results of EE

After analyzing the benefits, they are ranked according to the score.

1. Increase accountability (B2)
2. Reduce the likelihood of errors being introduced (B4)
3. Ensure that data is established, implemented, and governed (B5)
4. Maximizing the use of data to make decisions (B1)
5. Create data that is accurate, complete and consistent across the organization (B3)

5.2.6 Case study conclusion

The validation phase of this study was performed in order to answer SQ2 “What are the areas of overlap in data governance and data architecture?” and SQ3 “How to achieve the alignment between data governance and data architecture?”. The purpose of the case study evaluation was to observe and identify how much of the developed alignment model is followed in the real organization. DA group of people and DG group of people align in most of the areas mentioned in the model. These are the areas and observed and mentioned in the case observation. The areas like data standards, data policies, taxonomies, and data identification are not very clear in the case study who creates and maintains those. Maybe those that do not necessarily align themselves in this case. The usefulness was measured by observing and validating quantitatively.

5.3 Conclusion: Evaluation and Validation

The alignment model of data governance and data architecture have been demonstrated and evaluated via case study and an interview. The purpose of the case study evaluation was to observe and identify how much of the developed alignment model is followed in the real organization. The usefulness was measured by observing how the model can bring these two aspects together: data architecture and governance. It is also observed how much of the proposed model is followed in the case study and the benefits of the proposed model are provided. With the gathered feedback from the interview, the model has been

adjusted accordingly and the final version of the developed alignment model is presented.

In terms of the evaluation approach's strengths and weaknesses, it can be concluded that the strength of the evaluation approach via semi-structured interviews has provided the opportunity to gain deeper insights by receiving feedback and perspectives on how to enhance the model. On the other hand, the small number of participants and evidence of only one interviewee from a single organization give the evaluator a weak picture. This has implications for the final alignment model, as some components of the model were modified in response to interviewee feedback. It is advised that another iteration with more interviewees be performed to address the key implications of the demonstration and evaluation step.

There are several strengths and weaknesses that need to be evaluated when it comes to the proposed alignment model. First, the model has been proven to be presented in a clear and logical manner for both technology-oriented and business-oriented audiences. The model has been considered to be useful because it follows a standardized manner of functioning in the company. The essential components stated in the model can be used as a template and guide for new governance and architecture team adaptors, as this list can be expanded and adjusted as needed. The proposed model can help with a variety of issues, such as making better decisions. On the other hand, the developed model's weaknesses include: which individual is responsible for which activity is not included in the model, so the developed model does not indicate roles and interactions amongst stakeholders that have not been considered.

6 Conclusion

This chapter presents the conclusions of the main research objective and the sub-research questions. The scientific and practical contributions will be highlighted in this chapter. Lastly, this chapter ends with the limitations of the research which is presented in [section 6.3](#) and recommendations for further work in [section 6.4](#).

6.1 Summary & Conclusion

The main objective of this thesis was to "*Develop a model that covers and compliments with both the data architecture and the data governance*". From the literature review, it is found that little attention is given to the topic "the relationship between data architecture and data governance" in the academic papers. In order to fulfill the goal of this dissertation, we have decided to apply the design science research methodology adopted from (Peffer, Tunnanen, Rothenberger, & Chatterjee, 2007) since it is aligned with the overall objectives of the thesis. By executing all steps of the DSRM, the main objective of the research is accomplished by presenting the developed alignment model. The alignment model is based on the synthesis of the literature and the analysis from the evaluation.

The main research question is "How to align data governance and data architecture and how it can be achieved?" To achieve the research objective of this thesis, the sub-research questions were formulated which supported the development and evaluation of the final alignment model. Three sub-research questions were introduced in [section 1.2](#) and answered during the research. We'll now go through a quick recap of the answers to the sub-research questions.

SQ 1: What is Data governance and Data architecture? (K)

I was able to find several aspects relating the to the state-of-the-art literature available for Data governance and Data architecture. Both Data governance and Data architecture are two different domains which are still developing and validating research. The literature regarding Data governance has been relatively

abundant in the last few years while Data architecture is also becoming prominent. Since the purpose of this study used systematic literature review, the results of the study are the current state of knowledge in the area of data architecture and data governance alignment and it is described in the Literature Review chapter.

SQ 2: What are the areas of overlap in data governance and data architecture? (K)&(D)

We went deeper into the literature articles for this research question. One aspect that is common to all the selected literature is that most of these studies are still specialized and focused individually while not paying too much attention to further connections between both the disciplines. The search results found that there was certainly a limit to conducting the study because the articles available relating to both there were very few disciplines together, leading to very few published papers talking about the relationship between data architecture and data governance or the synergies between them. So, it was very difficult and took longer to find the areas where they overlap. However, managed to find the areas and it is briefly explained in chapter 4 Design and Development and the reasons why these are areas that are aligning are also described. In order to get the clarity and the correctness it was also evaluated by an expert.

SQ 3: How to achieve the alignment between data governance and data architecture? (D)

The aim is to achieve alignment data architecture and the data governance is constructed through a model and how achievement is made can be extracted from the model. The alignment model was designed with the use of literature review and mainly based on the contents of the two books that are widely used for Data management DAMA (DAMA, 2017) and DCAM (DCAM, 2020). The model gives clarity on how to achieve alignment and the benefits of using the model is also provided and validated quantitatively.

Based on the research, it can be concluded that the areas of data architecture and data governance overlap can be extracted and alignment can be achieved using the model. This model will increase speed in decision making. The alignment model clarifies who in the organization has decision-making authority over each work activity. Employees can also focus on doing their jobs rather than figuring out who is required to do what and who has to be consulted when making decisions with the help of an alignment model. Also, it helps with the visibility of resources, so the organization can have fewer duplications of resources by helping to look at the same information from different angles, allows to identify early errors and misalignments and more resources to boost the company's performance and productivity.

6.2 Scientific and practical contribution

This research contributes to scientific literature and provides practical contribution.

At the moment, combined data governance and data architecture research are a relatively new field of study. It was discovered throughout the literature study that there is very little scholarly literature on the topic of alignment models. In terms of scientific contribution, this thesis can be considered one of the first studies to create an alignment model for data architecture and data governance, which will add to the current governance & architecture literature, mainly to the data management literature. The scientific contribution of this thesis is mainly the developed alignment model, which was based on a combination of governance and architecture elements and insights gained from empirical research. Based on the findings, the current alignment model evaluation did meet the key benefits of using it. This conclusion, however, needs to be confirmed with follow-up studies based on this research so that more inferences may be reached. The conducted interviews were an important part of collecting current knowledge in the industry. Because data governance and data architecture are becoming more widely used, the thesis also raises awareness of the importance of undertaking research on data architecture and its governance among scientists. Furthermore,

empirical study analysis contributes to the enriched literature on data governance and data architecture. Furthermore, for other academics, the thesis may serve as a starting point for understanding and developing new ideas. Within the alignment topic, there is still a lot of room for future scholars to examine in order to improve this research area.

For practical contributions, the thesis provides an overview and establishes areas and processes that align with data architecture and data governance. Another outcome of the thesis is that in large organizations data architecture or any of the discipline may not be done in the same organization, it could be done by another external company. So, the research provides the areas where both the organizations need to be interacted and responsible for the tasks. The organization can use and enhance the developed alignment model in order to provide a better model and new insights in the future.

6.3 Limitations

As with any other research, this thesis has limitations which should be considered. In this section, the limitations of this research are explained.

The research findings are drawn based on the evaluation by conducting an interview and by investigating a case study. This could be further investigated by applying the model to more case studies. Even though the developed alignment model consists of generic processes and elements, taken from literature, it has not been validated by different organizations and different industries for its usability and applicability. It is also important to point out that there is no single ideal governance and architecture model which would always suit all organizations. This research does not prove that the developed model is best suited for all the organizations. Due to that, the appropriate level of detail that needs to be conveyed to design and develop the model is differentiated. However, the key components and implementation methodologies that are commonly used in the business and literature were applied.

The second group of limitations refer to the literature review, one aspect that is common to all the selected literature is that most of these studies are still very specialized and focused on one or two domains while not paying too much attention to further connections between both the disciplines. Identifying the search results found that there was certainly a limit to conducting the study because of the articles available relating to both, there were very few disciplines together, leading to very few published papers talking about the relationship between data architecture and data governance or the synergies between them. Analysis of governance processes where architectural alignment is happening. Even though Data architecture and Data governance are disciplines of Data management, their studies are still lacking in both DA and DG. More studies are therefore required to expand, develop and implement the alignment.

In terms of the limitation of evaluation, it can be concluded that the strength of the evaluation approach via semi-structured interviews has provided the opportunity to gain deeper insights by receiving feedback and perspectives on how to enhance the model. On the other hand, the small number of participants and evidence of only one interviewee from a single organization give the evaluator a weak picture. This has implications for the final alignment model, as several components of the model were modified in response to interviewee feedback. The evidence is presented by displaying the interviewees' responsibilities and the feedback they provided. It is advised that another iteration with more interviewees be performed by hand on practice to address the key implications of the demonstration and evaluation step.

6.4 Recommendations

In this section, the recommendations will be given for potential future research and then for the analyzed case study. These recommendations are extracted from limitations and analysis from the evaluation.

For further research, more extensive research is needed to fully design the governance and architecture model for managing it in large organizations and it could be evaluated using a more real case.

Firstly, the designed alignment model can be applied and tested in more real-life cases for its effectiveness and usability. The theoretical knowledge gained from the developed model can be investigated further.

Secondly, the impact on business process performance could be identified if there is no alignment between data governance and data architecture. This could be measured with some metrics.

Thirdly, the alignment can be further investigated through the tools of DA and DG used by the organization to analyze more.

More broadly, scientists could not just focus on alignment, but in general on the process automation landscape at an organizational level when there is alignment that can also be researched. It has been observed that integrating several automation tools can create a multiplier effect and lead to end-to-end automated services, so for further research focusing on the bigger picture of automation can be suggested.

References

- Abraham, R., Schneider, J., & Brocke, J. v. (2019). Data governance: A conceptual framework, structured review, and research. *elsevier*. doi:<https://doi.org/10.1016/j.ijinfomgt.2019.07.008>
- Adams, W. C. (2015). Conducting Semi-structured interviews. In *Handbook of Practical Program Evaluation* (p. 15). Jossey-Bass. doi:10.1002/9781119171386.ch19
- Burbank, D., & Roe, C. (2017). *Trends in Data Architecture*. Global Data Strategy. Retrieved from www.datadiversity.net
- Cervo, D., & Allen, M. (2015). *Multi-Domain Master Data : Advanced MDM and Data Governance in practice*. Elsevier Inc.
- Cheong, L. K., & Chang, V. (2007). The Need for Data Governance: A Case Study. *18th Australasian Conference on Information System*. Retrieved from <https://www.researchgate.net/publication/228966685>
- Cristian, E., Anca, M., & Cerasela, P. (2008). THE ROLE OF DATA ARCHITECTURE AS A PART OF ENTERPRISE ARCHITECTURE. doi:10.1007/978-3-642-20279-7_2
- DAMA. (2017). *DAMA-DMBOK: Data Management Body of Knowledge* (2nd ed.). Technics Publications.
- DCAM. (2020). *Data Management Capability Assessment Model (DCAM)*. Enterprise Data Management Council.
- EDM council. (2014). *Data Management Capability Assessment Model*. Enterprise Data Management Council.

- Gupta, U., & Cannon, S. (2020). Data Governance Maturity Models. In *A Practitioner's Guide to Data Governance* (pp. 143-165.). Emerald Publishing Limited. doi:10.1108/9781789735673
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 75–105. Retrieved from <https://www.researchgate.net/publication/201168946>
- Hoven, J. V. (2006.). Data Architecture: Principles for Data. *Information systems Management*, 93-96. doi:10.1201/1078/43205.20.3.20030601/43078.11
- Johnson, B. (2015). Big Data: Architecture and Its Enablement. (K. M. Lehmann, Ed.)
- Khatri, V., & Brown, C. V. (2010). Designing Data Governace. *Communication of the ACM*, 148-152. doi:10.1145/1629175.1629210
- Kitechenham, B., & Charters, S. M. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering. *Journal of Software Engineering and Application*, 65. Retrieved from www.elsevier.com
- Kotzab, H., Seuring, S., Muller, M., & Reiner, G. (2005). *Research methodologies in supply chain management*. Physica.
- Loshin, D. (2015). Enterprise Data Architecture and Data Governance: Use Metadata to Get to the starting gate. *Knowledge Integrity Inc*. Retrieved from www.knowledge-integrity.com
- Lussier, R. N. (2015). Management fundamentals: Concepts, applications, and skill development. *Sage Publications*, 7th.
- Nick, M., Joris, H., & Mark, B. d. (2015). Determining the Effects of Data Governance on the Performance and Compliance of Enterprises in the Logistics and Retail Sector. doi:10.1007/978-3-319-25013-7_37

- Niemi, E. (2015). Designing a Data Governance Framework. 14. Retrieved from https://www.researchgate.net/publication/257194534_Working_Paper_Designing_a_Data_Governance_Framework
- Panian, Z. (2010). Some practical experiences in data governance. 939–946.
- Panian, Z. (2010). Some practical experiences in data governance. *World Academy of Science, Engineering and Technology*. Retrieved from <https://citeseerx.ist.psu.edu>
- Peffer, K., Tunnanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 45–77. doi:10.2753/MIS0742-122240302
- Russom, P. (2008). *Data Governance strategies: Helping Your Organization Comply, Transform, and Integrate*. Retrieved from TDWI.org: <https://tdwi.org/articles/2008/05/27/data-governance-strategies-helping-your-organization-comply-transform-and-integrate.aspx?m=1>
- Saunders, M. N., Thornhill, A., Lewis, P., & Bristow, A. (2015). *Research methods for business students*. Pearson Education.
- Seiner, R. S. (2014). *Non-Invasive Data Governance*. Technics Publications.
- Sherman, R. (2015). Data Architecture. In R. Sherman, *Business Intelligence Guidebook* (pp. 107-142). Elsevier. Retrieved from <https://doi.org/10.1016/B978-0-12-411461-6.00006-X>
- Steenbeek, I. (2019). *The Data management Tool kit* (1st ed.). Data Cross roads. Retrieved from <https://datacrossroads.nl>
- TheOpenGroup. (2013). Archimate 2.1 Specification. *Van Haren Publishing*, 1-30.

- Thomas, G. (2020). *How to use the DGI Data governance framework to configure Your program*. Retrieved from The Data Governance Institute: www.DataGovernance.com
- Urso, Vince. (2020). *Leveraging Data governance in financial service*. Retrieved from perficient.com: www.perficient.com
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology:.. *MIS quarterly*, 425-478.
- Wieringa, R. (2014, May). Design science as nested problem solving. *DESRIST '09: Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology*, Pages 1–12.
doi:<https://doi.org/10.1145/1555619.1555630>
- Yin, R. K. (2003). Case study research : Design and Methods. *Thousand Oaks*, 5, 93-95. Retrieved from <https://www.jstor.org/stable/23279888>

Appendix A: Identified Literature paper

The overview of the Literature paper identified during the literature review is given in the quality assessment form.

No.	Reference	Research Purpose	Research Method			Output			
			LR	O	E	T	CM	A	IA
P1	(Abraham, Schneider, & Brocke, 2019)	“This paper provides an overview of the state-of-the-art of data governance, and identified a research agenda of the lack in knowledge about and the building blocks of data governance.”	X				X		
P2	(Burbank & Roe, 2017)	“This paper is an analysis of a Data diversity 2017 Survey on the latest trends in Data Architecture and analyzed the results in the remainder of the paper and investigated other key findings that affect today’s data-driven organization”		X					X
P3	(DAMA, 2017)	“This book provides data management and IT professionals, executives, knowledge workers, educators, and researchers with a framework to manage their data		X		X	X		

		and mature their information infrastructure”							
P4	(Cristian, Anca, & Cerasela, 2008)	“This paper proposes the structure of Data architecture of all corporate data and its relationships to itself and external systems.”		X		X			
P5	(Gupta & Cannon, 2020)	“This paper presents the Data governance maturity models”	X					X	
P6	(Hoven, 2006.)	“This paper data architecture principles guiding the decision-making process about data and its related technologies through all levels in an organization, stating an organization’s values as well as guide product selection and implementation.”		X		X			
P8	(Loshin, 2015)	“This paper talks about the relation between both enterprise data architecture and data governance using meta data as a starting date also explores some challenges in bootstrapping a data governance program, and then considers key methods for using metadata to establish the starting point for data governance.”		X		X			

P9	(Niemi, 2015)	<p>“This working paper describes a research outline aiming at designing a data governance framework for globally operating companies. In addition, it introduces a doctoral dissertation research on data governance and positions it in the context of existing research.”</p>			X		X		
P10	(Panian, Some practical experiences in data governance., 2010)	<p>“This paper presents the Some Practical Experiences in Data Governance examining two of the four data governance components – organization and technology – in more depth.”</p>		X			X		
P11	(Russom, 2008)	<p>“This report is designed for business and technical executives responsible for planning and implementing a program for data governance (DG). This report helps organizations worldwide successfully navigate the unknown waters of DG by presenting the best practice techniques and common pitfalls involved in starting and sustaining a DG program. It identifies and evaluates common starting points</p>			X				X

		and strategies, with an emphasis on the cross-functional nature of DG”							
P12	(Seiner, 2014)	“This paper provides a complete set of tools to help deliver a successful data governance program also presents a practical and non-threatening approach that can be applied to governing information and promoting stewardship of data as a cross-organization asset.”		X		X			
P13	(Sherman, 2015)	“This book helps to design the overall architecture for functioning business intelligence systems with the supporting data warehousing and data-integration applications”		X		X			
P14	(Thomas, 2020)	“This paper presents the how to configure the framework of Data governance for the program explaining all the components briefly”		X			X		
P15	(EDM council, 2014)	“This paper provides the Data Management Capability Model solely for peer review purposes.”			X			X	

P16	(Nick, Joris, & Mark, 2015)	<p>“This research paper has literature about data governance, both from theory and practice, reviewed dossiers of data governance projects conducted at a consultancy firm, and have held interviews with experts and with clients of this firm”</p>	X				X		
P17	(Cheong & Chang, 2007)	<p>“This paper presents a research into data governance and enterprise data management. The literature review on enterprise information management highlights that a high percentage of organizations across the world are engaged in the management of data as an enterprise asset”.</p>	X				X		
P18	(Johnson, 2015)	<p>“This paper articulates the need for and advantages of enterprise-wide data.</p> <p>The architecture encompassing the data life cycle also describes the main database designs for big data, the relationship between</p>		X					X

		data architecture and data governance.”								
--	--	---	--	--	--	--	--	--	--	--

Table 14 Quality Assessment Form

Appendix B: Interview Agenda for Model Evaluation

The following agenda was used to ask interview questions. Because the interview is semi-structured, the actual question posed by respondents may change based on the conversation's flow, but the procedure will still be followed.

1. Introduction

- a. The interviewer introduces herself;
- b. The research background and purpose of the interview session are explained;
- c. Ask for the consent to record the interview; and begin recording the interview.

2. Personal

- a. Role of interviewee in the organization? Could you brief me on your position?
- b. How long have you been involved with the Data governance and Data architecture activities?

3. Presented the alignment model

- a. Explained the challenges and limitations while designing the model first;
- b. Explained the model activities briefly;
- c. Explained why the areas proposed are the areas they align;

4. Questions

- a. Questions asked during the interview were to determine the correctness, quality, and understandability of the alignment model.
- b. The opinions gathered are highlighted and transcribed below briefly (see [Appendix C](#)).

5. Closing the interview

Appendix C: Evaluation Questions

The semi-structured interview for the evaluation of the proposed alignment model was conducted on 28/09/2021 via zoom call. Following the interview from the expert, a questionnaire was shared along with the discussion. The answers and his opinions are transcribed and tabulated in the table below:

Question	Opinion
<p>Too often in org DA and DG are done separately and they don't realize they are mutually reinforcing, is that the reason why there is less literature or blogs that talks about both the disciplines?</p>	<p><i>As far as I know, data governance is one of the newer disciplines of data discovery, which is why I decided to have a crack in a master's degree in data governance framework. So, when you look at the wheel you are familiar with the DAMA, in the middle. I think one of the segments or one of them spokes at the wheel because data architecture is a discipline, always treated traditionally as a technical discipline. Take the model for example that's quite a technical data management discipline. My colleague Donna Burbank, you know who I work with, she's an expert Data Modeler doing data modeling. So, it's quite a skilled thing to produce the data models. The problem with architecture as I said was, because it was a technical discipline, implementing it in the real world, was really difficult. So, when data governance came along, which was a later discipline, then people began to recognize that data governance, actually, can be a very good route of bridge if you like, between the creation of a data architecture, and its implementation in the real world. That's why I think there's a mismatch, because the other thing is a lot of organizations have a data governance function but no</i></p>

	<p><i>architecture function, or they have an architecture function and all data governance. You know my argument in the session, I did crack, I think DAMA would say the same is you've got to have both to optimize both. Defining your governance problem is really hard. If you have architecture but no governance, implementing your architecture makes it very hard to say things that support each other. So that's why I think there's that sort of lack of research because governance is, is a newer discipline. I discovered that as well. I mean when I did some recent literature reviews. There's not an awful lot out there that I find very useful about data governance. A couple of books, do people like John Landley and the rest of it. Yes, but they didn't talk about the alignment between the data architecture, it's very less.</i></p>
<p>Do you think now that experts are realizing the concepts are interrelated, that processes that may appear to be wholly related to DA can play a key part in DG?</p>	<p><i>Yeah, I would say that data governance, if you look into my presentation I think from your track I have a slide in there where I took the DAMA wheel to show you like to go to governance is in the middle of the DAMA and the data architecture, right side. I agree with the DAMAs view I agree with everything they say but data governance but that I agree with their view so that without data governance, you cannot optimize all the other, it's like the heart of the wheel. The only example I always use, which is that you know if you just need technical people to sort out data quality issues. It's an it's a massive failure because they will keep on doing that and keep on doing that, because the business will continue to create bad data. Okay, governance is the root to say and do, we got to fix our data quality problems, and we got to make sure we sustain the gains that if the business is leading efforts to improve data quality if you leave it at the technical people, they can never get to the root causes of the</i></p>

	<p><i>problems, which could be training or it could be you know, bad validation of data or inputs it could be posted things,, bad training of people. All these things are only things that the business can solve IT.</i></p>
<p>Do you think the alignment between DA and DG is important?</p>	<p><i>This is important as I said, you know, you get better architecture help to scope and prioritize governance activities, you have to do that. Because if you try and govern all the data in an organization, it's impossible to do time, you have to, you know, you have to put your hands in the air and say, this data is a mess. But we it's too complicated and too difficult to solve. What we need to fix is the real data that really matters to us. So, you know, that's, that's one reason why I think alignment is really important. So, architecture can help governance decide what the critical data is. There is then governance that will help architecture by providing accountable people who can own the architecture with its development. Those are the key things.</i></p>
<p>Do you think these are areas that both align?</p>	<p><i>Yes indeed, you have captured everything pretty well. The areas that are shown in the model are really the areas they have to work together. The model is quite clear.</i></p>
<p>Would you like to add more or want to improve on the model?</p>	<p><i>I made the three key points for me just not to make it linear and put the business requirements in the beginning and to group the metadata activities. Everything else is great.</i></p>
<p>Is the proposed model easy to use in practice?</p>	<p><i>I definitely think so. I mean you know it's useful to like develop similar things myself. As I said, my concern is that don't make it look too linearly covered already because it gives the impression that you did this once. You need to continue the continuous process. So, yes, it will be practically useful. Yes, it will be practically useful. But you</i></p>

	<p><i>may need to make it clear to people that you've never read up to the data architecture and data governance structure that that's the final output. Say data governance is not a project you may start with. But it's a continuous process. Data Governance established doesn't go away, and don't fix the problem or work on it all the time.</i></p>
<p>Since we know the areas that align with DA and DG from the proposed model, if we use this model will it be more efficient to collaborate or to get the people aligned?</p>	<p><i>Yeah, I definitely think so. You know, I had a lot of contempt. I didn't do it when I did my presentation. I think I did one of my slides to remind myself. I think for the continuous improvement, we made a slide, Governance and architecture, how to make it happen. You'll notice that when I do that, it's a simple lagain ittle diagram, but it shows this continuous alignment and improvement activity. You know, you go around and round this circle again and again and again, you never stop doing runs. Just to identify if your critical data and the example I gave you, the manufacturing company, they weren't collecting any personal data until the business decided he was going to sell directly to the public. Rest of it. So yeah, these areas are known now and it will benefit the organization to collaborate efficiently.</i></p>
<p>What do you think are the benefits of using this model? Do you think it will benefit your organization?</p>	<p><i>I mean, at the moment, one of the benefits is to me is that even in organizations where they have a data architecture function, and they have a data governance function, then those two groups of people don't necessarily align themselves demonstrate how aligning those two things would benefit both. There is a sort of benefit that of showing the relationships between the two, working more closely together and collaborating more closely together, would help them both achieve their goals.</i></p>

<p>Do you think this model can be used for a project that is half way constructed or it could only be used in projects that start from scratch?</p>	<p><i>I mean, it would be in my world of data management, that any project would start with some of these activities. You start, you know, creating new data sources, or whatever you would do, you want to get the governance in place, and you want to understand the architecture of where that data is going to be mastered, etc. That's the ideal world. In practice, that often doesn't happen. And it's often argued as I think it could be useful both at the beginning and also when a project is very often it becomes important when things get into trouble.</i></p>
<p>Other important points during the discussion</p>	<p><i>Data governance is not a well-researched topic, so it might surprise you to know that I am also a master student at the moment, I'm looking at building a new data governance framework. So yeah, so the studies have not been going that well, for me, because I'm not full time, I'm still working part time as well. But you're right. I mean, I'm finding it hard to find good stuff around governance, generally. Our governance and architecture, as you say, is pretty thin on the ground. So yeah, it's good, it's good that you're doing a piece of research in the area that needs it.</i></p> <p><i>I think I mean, the data governance committee may interact with the data architecture and people to develop some of the policies and standards so that is alignment here. For the rest of the development of policies, and where you show that is, you've got you've got governance and architecture, I think it's in one of the slides that you saw from the presentation I gave a new truth, I think, is that there is a there is a mutual relationship between the two. You know, when you're doing if you implement data governance and trying to implement data architecture,</i></p>

then the two things support and reinforce each other. So, you might do something like a, as you say, a conceptual or a logical data model, that would then help you to define ownership of that data. And that would be would be implemented within your Governance Program. So, as you know, architecture can is driving to some extent how you structure data governance, but data governance then is also driving how you structure your data architecture, to things you do not have to do in parallel. I think that's what I would always say, you know, you can't do one or the other first, and then it must be done as a mutual setback. Activities which then reinforce each other, that's the key thing for me was always. Yeah. So, in other words, it's not something you get right from day one, you get to read it. And as you go to read, you develop more subtlety, more sophistication, and the rest of it as you move along. So, you might start with something quite rough and ready. All the time, the architecture evolves as the governance function evolves into something more subtle, and more appropriate for the organization. I always say things like data governance, the same is true of architecture, you never get it right the first time. It's something that you need to constantly revisit, because of course, as the business changes, then, you know, your architecture might need to change as well.

I feel very positive about the research. Yes, I think it helps to address the gaps as you can see in the literature. If you let me know when it will be published, I will refer to my research.

Appendix D: Evaluation of Benefits

Response Q1: To what extent does the proposed alignment model contribute to better decision making?

Participant	Q1: Score	Positive opinion	Negative opinion
A	4	<i>It helps to separate the concerns; who is responsible for what.</i>	
B	4	<i>It can save my time and I can use my resources (budget & employees) more accurately.</i>	
C	4	<i>The alignment model helps the organization to recognize the stakeholder of the data and quicken the process of decision making. process</i>	
D	1	<i>This model identifies the difference in responsibilities of DM and DA</i>	<i>Better decision making cannot be done from a Data perspective only. In my opinion Architecture is a holistic approach, Data architecture is a single-perspective approach. To make better decisions on the use of data, for example, you also need input from other areas such as strategy, portfolio management, business cases etc.</i>
E	4	--	
F	4	--	

Table 15 Response Q1

Response Q2: To what extent does the alignment model contribute in increasing accountability?

Participant	Q2: Score	Positive opinion	Negative opinion
A	4	<i>The model assigns responsibilities and makes individuals accountable for the results; then you can follow up on it.</i>	
B	5	<i>It shall improve the quality of the data, business planning and customer satisfaction which could help in maximizing the profit of the company</i>	
C	4	<i>Data governance will help the organization to recognize authority and control over the data management.</i>	
D	5	<i>The model itself does not provide insights into accountability on the data itself.</i>	
E	4		
F	4		

Table 16 Response Q2

Response Q3: To what extent does the alignment model bring standardization and consistent data across the organization?

Participant	Q3: Score	Positive opinion	Negative opinion
A	3	<i>Standardization and consistent data are the domain of master data management, it is a result of a consensus process, which is not the focus of this model in my opinion</i>	
B	4	<i>As it allows for collaborative research, large-scale analytics, and can exchange clear definitions of data among the organizations it will make data more consistent and accurate</i>	
C	4		
D	2		<i>The model itself does not make data better. It is the use of the processes that could. And only if this process is embedded in complete governance and aligned with the change process. If a company decides to implement this model, autonomously and separately from the rest of the organization, it</i>

Table 17 Response Q3

			<i>will not be effective and failing its purpose is very likely.</i>
E	4		
F	3		

Response Q4: To what extent does the proposed alignment model help in the duplication of data and resources?

Participant	Q4: Score	Positive opinion	Negative opinion
A	4	<i>The model brings the two parties together and requires their cooperation. In the end the results have to fit together to make it more aligned.</i>	
B	3	<i>As per the model, it might help in avoiding data duplication</i>	

C	3	<i>The proper model should prevent errors to be created and by doing the exercise of data governance should build the bridge between IT and business.</i>	
D	2		If Governance and Architecture would have been implemented in the general sense, Data should have been covered already and we wouldn't need this model. This model suggests a completely new approach where methodologies like TOGAF and DMBOK already cover this.
E	3		
F	3		

Table 18 Response Q4

Response Q5: To what extent does the proposed alignment model help in reducing the likelihood of errors being introduced?

Participant	Q5: Score	Positive opinion	Negative opinion
-------------	-----------	------------------	------------------

A	4	<i>By having two different parts of the organization looking at the same information from different angles allows us to identify early errors and mis-alignments</i>	
B	5	<i>Document Management, Data Governance and Review of Data and Resources Could Reduce Human Errors</i>	
C	5		
D	5		
E	4	<i>When organizations are in alignment, the likelihood of errors is low. Better alignment leads to better communication and reduction in errors.</i>	
F	2		

Table 19 Response Q5

Response Q6: To what extent can the proposed alignment model contribute to creating a standardized data architecture plan that serves as the foundation for layering data policies to ensure usability, quality, and consistency?

Participant	Q6: Score	Positive opinion	Negative opinion
A	4		
B	4	<i>Data audits should comply with the company's internal</i>	

		<i>policies and processes. It could help in improving and understanding internal business processes and detect and analyze data breaches</i>	
C	4		
D	4	<i>It helps to clarify the responsibilities of Data in Governance and Architecture. I see it as a subset of more detail, part of Governance and Architecture that should be in place.</i>	
E	4	<i>Standardization is what's desired and consistency will follow once it is achieved. A closer examination would be beneficial and can certainly enrich the model quality.</i>	
F	4		

Table 20 Response Q6

