MSc Business Information Technology

Master's Thesis:

Data Governance Structures in Data Mesh Architectures

Koen Hendriks

Supervisor:
Marten van Sinderen (University of Twente)
Sebastian Piest (University of Twente)
Esmee van Waalwijk van Doorn (Deloitte Netherlands)
Niko Vermeer (Deloitte Netherlands)

May 26, 2023

Department of Business Information Technology
Faculty of Electrical Engineering,
Mathematics and Computer Science

UNIVERSITY OF TWENTE.

Executive summary

The increasing complexity of data poses significant challenges for organizations, straining the capabilities of current centralized data teams. The migration to a data mesh architecture has emerged as a potential solution to the increased responsibilities faced by these data teams. The successful migration towards a data mesh requires a new Data Governance Structure (DGS) that effectively supports the principles of a data mesh architecture. Unfortunately, designing this new DGS is daunting, and the lack of guidance available in existing resources results in inefficiencies in the design process.

This research aims to develop a guidance tool for the design of DGSs for data mesh architectures. This tool presents important design decisions, possible organizational configurations, and the advantages and disadvantages of these configurations. These decisions affect the positioning of data domain owners and data product developers, the domain representation in the governance council and the role of the steering committee and enabling team. By providing guidance, the tool can support enterprise architects with structuring the roles and responsibilities in the federated governance team, data platform team, and domain teams in data mesh architectures.

The research follows a rigorous design science research methodology, encompassing four key phases:

The first phase provides a solid scientific background on data mesh architectures, challenges in data mesh migrations, and important data governance components of these architectures. The second phase investigates the DGS design process and the problems that occur during this process. The third phase includes the design and development of the guidance tool. This tool is created based on the theoretical DGS of a data mesh architecture specifically designed for this research, the practical DGSs of two case studies, and additions from subject matter experts. The fourth and final phase is the evaluation and refinement of the guidance tool through expert opinions based on requirements regarding its correctness, understandability, utility, efficiency, and effectiveness.

The research findings provide conclusive evidence that the guidance tool is correct, understandable, and useful, and that its use is expected to improve the efficiency of the DGS design process without sacrificing its effectiveness.

This research contributes to practice by providing a starting point for the design of DGSs and by explaining multiple important decisions that should be considered during the formulation of recommendations to clients or during discussions with stakeholders. These contributions could reduce the risks of data mesh migrations and create competitive advantages compared to other organizations for both the consulting

organizations and the consulted organizations.

The research also has scientific contributions as it presents the first exhaustive literature reviews on data mesh migration challenges and data governance components in data mesh architectures. Additionally, it presents the first theoretical DGS of a data mesh architecture. Lastly, it demonstrates the effectiveness of a new methodology that uses findings from theory and practice to develop the first guidance tool for DGS design in data mesh architectures.

Future research can improve the limitations of the current guidance tool. The tool could be further validated, preferably during technical action research at an organization engaged in the design of a new DGS. This enables further investigation and quantification of its influence on the efficiency and effectiveness of the design process. Furthermore, as the field of data mesh continues to evolve, regular revisions of the guidance tool are necessary to ensure its relevance and applicability.

Preface

Presented before you is the master's thesis titled "Developing a Guidance Tool for Data Governance Structure Design in Data Mesh Architectures." This thesis is written as part of the master's Business Information Technology at the University of Twente. The research is conducted in collaboration with Deloitte Netherlands from September 2022 to May 2023.

The thesis explores the emerging field of data mesh, which holds the potential to improve the scalability of large data-driven organizations and is expected to produce a needed paradigm shift in their approach to data management. Migrating to a data mesh approach has significant implications for organizations. However, there is limited knowledge of how organizations can fully leverage the potential benefits of this transformation. This contributes to the success of data mesh transformations by developing a guidance tool for data governance structure design in data mesh architectures. The idea of digitally representing information about the world and, with that, being able to explain the past, make decisions in the present, and predict the future has always been fascinating to me. I am thankful to be able to contribute to a data-driven future that aligns with my perspective on how data should be handled in organizations

To be honest, writing this thesis has proven to be one of the most challenging experiences I have encountered thus far. However, the support I received from the people who surrounded me made it truly rewarding.

I would like to express my gratitude to my supervisors from the University of Twente, Marten van Sinderen and Sebastian Piest. Their guidance and feedback have significantly contributed to the scientific quality of this research and their pleasant way of providing supervision has made the process very enjoyable.

I am also deeply grateful to Esmee van Waalwijk van Doorn and Niko Vermeer for their invaluable support during my time at Deloitte. They provided assistance whenever I encountered challenges and helped me structure my thoughts when needed. I would also like to extend my appreciation to the Tech Vision & Architecture team for creating a stimulating and pleasant working environment and to the interview and case study participants whose cooperation made this research possible.

Lastly, I would like to thank my family for their support. I express my deepest gratitude to my father for inspiring me, and to my mother, whose strength continues to empower me.

I invite you to delve into the pages of this thesis, and I hope that you find it enjoyable.

Koen Hendriks

Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Meaning
DAMA-DMBOK	Data Management Body Of Knowledge
DG	Data Governance
DGOM	Data Governance Operating Model
DGS	Data Governance Structure
DMA	Data Mesh Architecture
DODT	Design Option Decision Tree
DPA	Data Platform Architecture
EA	Enterprise Architecture
EDM	Enterprise Data Management
EIX	Evaluation Interview X
ETL	Extract Transform Load
FEDS	Framework for Evaluation in Design Science
MRQ	Main Research Question
OM	Operating Model
PIDG2	Problem investigation Interview with Data Governance expert 2
PIT1	Problem investigation Interview with Telecom Consultant 1
PIT2	Problem investigation Interview with Telecom Consultant 2
RQ	Research Question
SME	Subject Matter Expert
TAR	Technical Action Research
TDIDG1	Treatment Design Interview with Data Governance expert 1
TDIDG2	Treatment Design Interview with Data Governance expert 2
TDIDM1	Treatment Design Interview with Data Mesh expert 1
TDIE1	Treatment Design Interview with Energy Consultant 1
TDIE2	Treatment Design Interview with Energy Consultant 2
TDIOM1	Treatment Design Interview with Operating model expert 1
TDIT1	Treatment Design Interview with Telecom Consultant 3
TDIT2	Treatment Design Interview with Telecom Consultant 4
TOM	Technical Operating Model

List of Definitions

Concept	Definition used during this research
Data governance	The exercise of authority, control, and shared decision-making (planning, monitoring and enforcement) over the management of data assets
Data governance design	The creation of frameworks, principles, policies, and procedures for data governance by defining roles, responsibilities, and decision-making processes
Data governance operating framework	A description of the various arrangements and interactions of the organizational elements that will operate data governance is designed
Data governance structure	An organizational structure that describes the roles, responsibilities and accountabilities regarding data management tasks
Data mesh	A socio-technical decentralized data platform architecture that incorporates four key principles: 1) domain-driven ownership of data, 2) data as a product, 3) self-serve data platform and 4) a federated computational governance structure
Data mesh migration	A process of transitioning from a traditional data platform architecture to a data platform architecture that incorporates the data mesh principles
Data platform architectures	The technical and organizational structures and frameworks that are designed to support the storage, processing, and management of data within an organization.
Data platform team	A group of individuals that is responsible for managing the technical infrastructure and tools used to store, process, and analyze data within an organization.
Data product consuming team	A group of individuals that is responsible for the consumption of on or more data products in a data mesh architecture
Data product producing team	A group of individuals that is responsible for the production of on or more data products in a data mesh architecture
Data product team	A group of individuals that is responsible for one or more data products of a data mesh architecture
Design decision	A conscious choice made during the design process to determine the form, structure, or characteristics of a system, process, or artifact.
Domain team	A group of individuals that is responsible for the domain of a data mesh architecture

Federated governance	An approach that consists of both decentralized governance units and a centralized governance unit
Federated governance council	A forum in which the federated governance team coordinates, makes decisions and oversees governance activities
Federated governance team	A group of individuals that is responsible for or contributing to the execution of federated governance initiatives
Operating model	A representation of how a specific organisation's components are configured, and function together to execute on strategy
Organization design	The creation of the arrangement of an organization's roles, processes and resources to achieve its goals
Organizational configuration	A multidimensional constellation of conceptually distinct characteristics that commonly occur together

Chapter 1

Introduction

The past decade has seen the rapid development of data usage in organizations. The rise of unstructured data and the increase in the variety of datasets inside organizations create a demand for new data management strategies that can cope with the current scales of data environments [60]. Being able to effectively manage data and, with that, realize a greater portion of the potential value of data analytics, is now seen as a competitive advantage [19, 59]. Many organizations are transforming their businesses to realize the capability to extract business knowledge and insights for decision-making, in a cost- and time-efficient manner [19]. Large investments have been made by organizations in the design and development of data platform architectures that support this data-driven decision-making [21]. Although these architectures enable cost-effective storage and processing, some organizations still struggle with realizing the value of these data platforms [23]. One of the reasons for this is that organizations often have one central data team that is responsible for all data management activities, however, the new scale of data collection and analysis overloads the capacity of this team [23]. This results in problems in data ownership, responsibility allocation and lack of domain knowledge about the data [34].

Recent developments in the field of data platform architectures have led to a renewed idea of the organizational structure on top of these data platforms. Dehghani (2019) proposes a Data mesh Architecture (DMA) which is a socio-technical decentralized data platform architecture that incorporates four key principles: 1) domain-driven ownership of data, 2) data as a product, 3) self-serve data platform and 4) a federated computational governance structure [23]. In short, data mesh aims to bring the data responsibilities closer to the origin of the data and inside the business domains. These business domains are responsible for providing their data with product-like attributes to the rest of the organization. These domains are enabled by a self-service infrastructure that abstracts away some of the complexity of building and maintaining these data products and by a federated governance model that realizes standardization, interoperability and automation of decision-making while ensuring domain sovereignty [34]. This approach has shown its potential in organizations like Zalando and Netflix [59].

However, many organizations struggle with the migration towards an architecture that realizes the benefits of this approach [58]. This could be traced back to multiple causes: 1) the migrations towards data mesh architectures affect the organization both on technical and organizational levels which makes these projects large and complex, 2) the migrations often require a cultural change of the organizations and these changes are known to be

difficult in large organizations, 3) due to the novelty of data mesh, there is a lack of research on data mesh and a lack of experience about how to effectively operationalize its components in the organization [11, 58, 36, 21].

One of these components is the Data Governance Structure (DGS). This is the organizational structure that describes the roles, responsibilities and accountabilities regarding data management tasks [51]. An effective DGS can lead to improved data quality, increased trust in data, better compliance, improved decision-making, increased efficiency, and better risk management [51]. The effectiveness of these DGSs depends on organization-specific characteristics like the organization's structure, strategy, internal processes and external factors [78]. Therefore, there is not a universal approach to the DGS that is most effective for all organizations [78].

The DGSs currently implemented in organizations are constructed on an organizational model that does not follow the principles of a DMA. Migrating towards an organizational model that does follow these principles, impacts OSCs which means that retaining the traditional DGS is expected to be ineffective [23]. Therefore, organizations are required to design a new DGS based on the new OSCs.

Bode et al. found that the design of and migration towards a fitting DGS is seen as one of the main practical challenges in data mesh migrations [11]. Due to the complexity of organizational transformations and the complexity and novelty of the data mesh principles, designing a new DGS is a complicated process [21]. The DGS design process is time-consuming and involves input from multiple stakeholders, resulting in high costs for organizations [51]. To reduce these costs, organizations aim to establish an efficient design process.

However, Bode et al. (2023) found that there is a need for further guidelines that assist in the structuring and operation of data governance in a data mesh as available resources do not sufficiently support a data mesh perspective on the decision-making process regarding the DGS design process. This finding is supported by the practitioners interviewed during this study as they experienced a lack of guidance during this design process which resulted in inefficiencies in the design process and in discussions with stakeholders (PIT1, PIT2, TDIE1, TDIE2, TDIT1, TDIT2).

A thorough investigation of this problem is presented in **Chapter 4**. The results of this investigation are condensed and presented in the next section.

1.1   Problem statement

As stated by the literature, there is a lack of guidance in the DGS design process which according to enterprise architects hampers its efficiency [11]. How the problem addressed by Bode (2023) resonates in practice is investigated through interviews with case representatives from two organizations (Energy 1 & Telecom 1) which resulted in the formulation of two main problems: First, the theoretical resources on data mesh are regularly not directly applicable in practice because of large discrepancies between the DGS proposed in theory and the current implemented DGS in practice. Secondly, the practical resources do not yet take into account the OSCs that are required to adhere to the data mesh principles. Data mesh-specific roles, responsibilities, and team structures that are expected to contribute to achieving the goals of a data mesh architecture (PIT1, PIT2, TDIT1, TDIT2).

As a result, enterprise architects lack understandable and practically useful resources that present and explain important design decisions that have to be made during the design process (PIT1, PIT2, TDIT1, TDIT2). This leads to a lack of structure in discussions with stakeholders and/or the need for enterprise architects to investigate and formulate design decisions (PIT1, PIT2, TDIT1, TDIT2). This may result in inefficient discussions and duplication of work (PIT1, PIT2, TDIT1, TDIT2).

Additionally, practitioners lack understandable and practically usable resources that present and explain possible configurations of data governance components, requiring enterprise architects to identify and formulate their own configurations with evaluation opportunities to decide which configuration fits the organization the best (PIT1, PIT2, TDIT1, TDIT2) and they lacked insight in the dependencies between configurations (PIT1, PIT2). This may result in additional time spent on the creation of configurations and the evaluation of these configurations.

These problems are visualized in the problem cluster presented in **Figure 1.1**. This figure shows the relations between the problems, their reference numbers (P1-P9) and their presence in the case studies Energy 1 and Telecom 1 (E1 & T1).
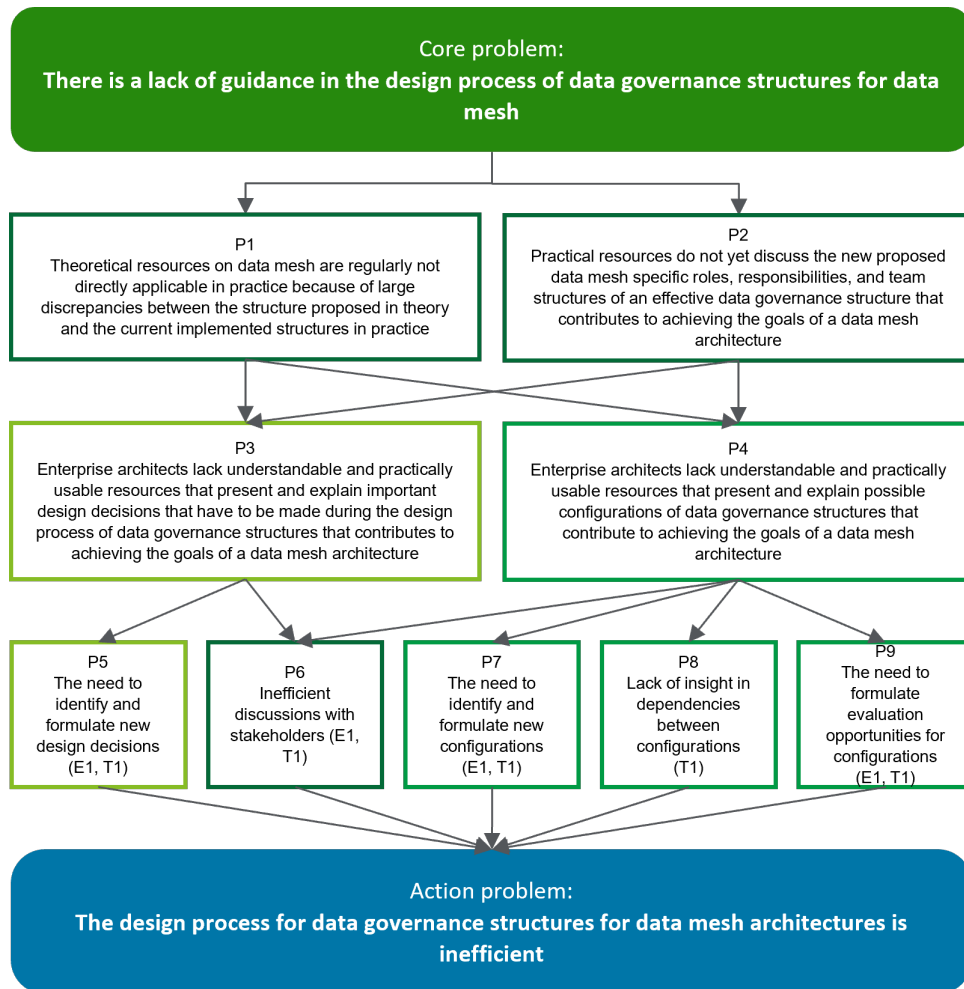
FIGURE 1.1: Research problem cluster

1.2 Research goal and approach

This study tries to improve migrations to data mesh architectures by providing guidance during the DGS design process to make this process more efficient without reducing its effectiveness.

This is done by designing and developing a guidance tool for DGS design for data mesh. For this, the design science research methodology proposed by Wieringa (2014) is followed [82]. During this study, the difference between a theoretical and a practical data mesh is identified through case studies and expert opinions. The findings are used to create a guidance tool which is validated by experts' opinions. Additionally, this study aims to contribute to the growing area of data mesh research by conducting a literature study on data mesh, a scoping study on data mesh challenges and a systematic literature on data governance components in a data mesh.

1.3 Relevance and demand

The practical relevance of this study is confirmed both in literature and in practice. Designing data governance is one of the main challenges in data mesh migrations as described in the scoping study in **Section 3.3** and this challenge is confirmed in practice during the problem investigation studies presented in **Chapter 4** all interviewees indicated that it is difficult to design the DGSs and that the current resources lack applicability in organizations that try to migrate towards a data mesh architecture. The tool is expected to serve as an instrument that can be used to improve the decision-making quality during the design of the DGS as it helps with generating alternatives [72]. Additionally, this study answers the future research section of de Boer (2022) by following a design science research methodology into the design and development of data governance in a data mesh [21].

1.4 Research questions

The main research question to support the research goal is formulated as follows:

How to improve migrations to data mesh architectures by designing a guidance tool that supports deciding on data governance options by fulfilling criteria on correctness, understandability, utility, efficiency and effectiveness in order to make the design of data governance structures more efficient without reducing its effectiveness?

To find an answer on this question, the following sub-questions are formulated:

RQ1 What does the migration to a data mesh architecture look like?

  (a) What differentiates data mesh architectures from traditional data platform architectures?

  (b) What comprises a data mesh migration?

  (c) What challenges occur in data mesh migrations?

RQ2 How is the data governance in a data mesh structured?

  (a) What is data governance?

(b) How is data governance structured?

(c) What are important data governance components in data mesh architectures?

RQ3 How are data governance structures designed?

(a) What are the key processes for designing the structure of organizational data governance?

(b) What influences the design process of data governance structures?

(c) Which tools are currently used as guidance during the data governance design process?

RQ4 How to develop a design guidance tool for data governance structure design in a data mesh?

(a) What are the goals and requirements of the guidance tool?

(b) Which method can be used to design and develop the guidance tool?

(c) Which method can be used to validate the guidance tool?

RQ5 What constitutes a guidance tool for data governance structure design?

(a) What high-level components does the guidance tool need to possess?

(b) Which design decisions for data governance structures in data mesh architectures can be distinguished?

(c) Which organizational configurations for data governance structures in data mesh architectures can be distinguished?

(d) How can the guidance tool be used during the design process?

1.5   Thesis structure

To address the research questions, the research framework, as presented in **Figure 1.2** was devised. The research approach consists of four phases: 1) theoretical background, 2) problem investigation, 3) treatment design, and 4) treatment validation.

Chapter 2 covers a detailed description of the methodologies used during this research and presents a methodology for the literature studies and for the development of the guidance tool based on theoretical and practical organizational DGSs.

Chapter 3 covers the theoretical background of this research and consists of three complementary parts. The first part is a literature review on the data mesh architecture and the challenges of migrating towards this architecture. The second part is a literature review on data governance and important data governance components in data mesh architectures. The third part is a literature review on the design of DGSs.

Chapter 4 covers the findings of the problem investigation interviews and case studies by introducing the social context and presenting the constructed conceptual problem framework.

Chapter 5 covers the design and development process in which a conceptual data governance framework of a data mesh is created and two case studies are conducted to formulate a practical point of view on the operationalization of this theoretical structure.

These findings are analysed to identify design decisions and organizational configurations.

Chapter 6 covers the evaluation and refinement of the guidance model which consists of a validation process through expert opinions and the presentation of the refined guidance tool.

Chapter 7 concludes the research by answering the research questions, presenting the contributions to theory and practice and discussing the research.
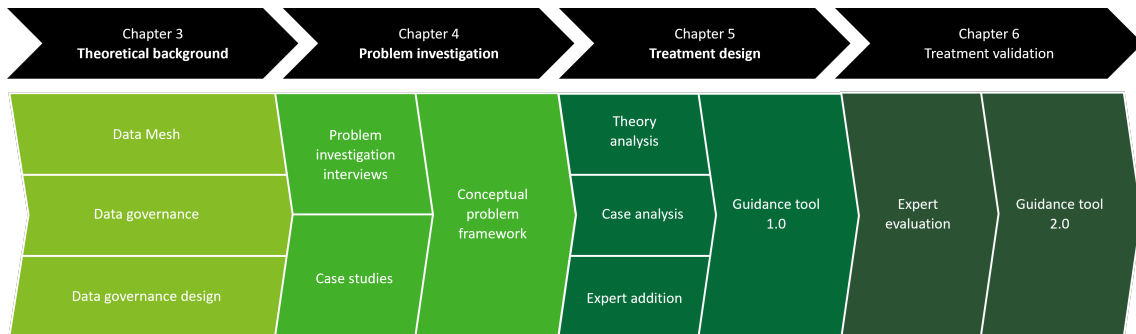
FIGURE 1.2: Research framework

Chapter 2

Research methodology

This section describes this methodology and introduces the methods used to answer the research questions. This research follows a methodology to answer the research questions presented in **Section 1.4**. The methodology consists of a problem investigation method, a treatment design method and a treatment validation method.

The theoretical background needed for the problem investigation is acquired through three literature studies. **Section 2.1** presents the method for the simple literature search regarding data platform architectures, data mesh and data governance. **Section 2.2** presents the method for the preliminary scoping study conducted on available literature regarding the challenges in data mesh migration. **Section 2.3** presents the method used for the systematic literature review on data governance components in a data mesh.

Section 2.4.2 introduces the method used for the investigation of the problem, in **Section 2.4.3** used for the design and development of the guidance tool is introduced, and **Section 2.4.4** introduces the method used for the validation of the guidance tool.

Table 2.3 presents an overview of the research questions with the used method for formulating the answers and the chapters in which the results of the methods are discussed.

2.1 Literature review

RQ1a, RQ1b, RQ2a, RQ2b are answered by following a narrative-style literature review method introduced by Ferrari (2015). Articles from online databases are selected based on search terms and perceived relevance and the information in the selected articles is used to answer the research questions. The implementation of this method in the context of this research and the research questions is presented in **Table 2.1**.

2.2 Preliminary scoping study methodology

To answer RQ1c, a scoping study is conducted. This extensive methodology is chosen because, at the time of writing, literature reviews on mesh migration challenges are nonexistent or undiscoverable. This scoping study tries to create an overview of challenges that

TABLE 2.1: Search terms, selection criteria and online databases used to find answers on Research questions RQ1a, RQ1b, RQ2a, RQ2b

Research question	Search terms	Selection criteria	Database sources
1a: How do data mesh architectures differentiate from traditional data architectures?	"Data platform" AND architecture OR "Data Mesh"	Peer-reviewed & pertinence	Google scholar, Science Direct & Semantic Scholar
1b: What comprises a data mesh migration?	"Data Mesh" AND Migration OR Transformation	Peer-reviewed & pertinence	Google scholar, Science Direct & Semantic Scholar
2a: What is data governance?	"Data governance" OR "Information governance"	Peer-reviewed & pertinence	Google scholar, Science Direct & Semantic Scholar
2b: How is data governance structured?	"Data governance" OR "Information governance"	Peer-reviewed & pertinence	Google scholar, Science Direct & Semantic Scholar

arise during data mesh migrations by following a method based on the methodology proposed by Arksey & O'Malley [6].

This method consists of five steps: 1) Identifying the research question, 2) Identifying relevant research studies, 3) selecting studies, 4) charting the data, and 5) collating, summarizing and reporting the results. A comprehensive explanation of how these steps are operationalized during this study is presented in **Appendix B**. The method is visualized in **Figure 2.1**.

2.3   Systematic literature study methodology

The systematic literature review (SLR) methodology of Okoli [61] is followed to answer RQ2c. This methodology is an SLR methodology specially designed for information systems research. It draws key concepts from multiple reviewed literature research methodologies, but primarily from Kitchenham's guide to SLRs in software engineering [45], Petticrew and Roberts' book on SLRs in the social sciences [64], Arlene Fink's guide on SLRs in health sciences [30]; Rousseau, Manning and Denyer's article on SLRs in management and organization science [69]; Levy and Ellis' article on conducting literature reviews in information systems [55]; and Webster and Watson's article [79] on writing up literature reviews in information systems [61]. The method of Okoli is used because this multi-sided methodology is expected to capture the socio-technical nature of data governance in a data mesh better than a methodology that is focused on one single research domain. The systematic literature review methodology used in this research is presented in **Figure 2.2**.
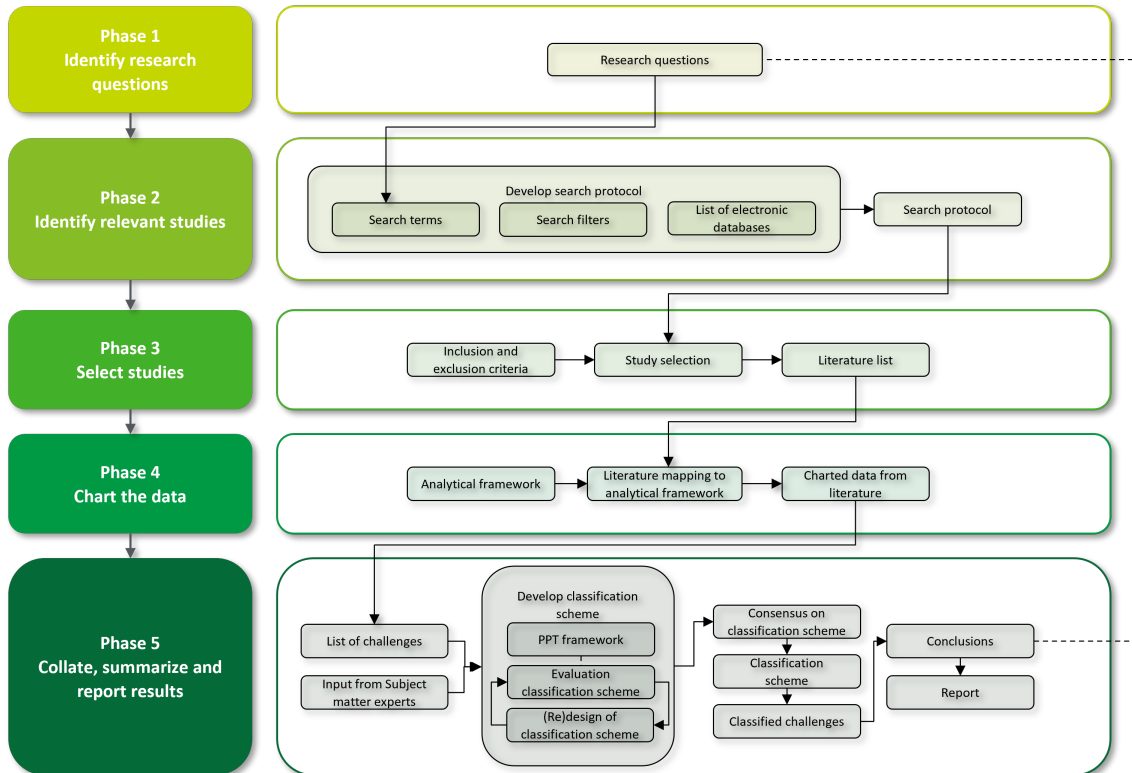
FIGURE 2.1: Methodology of scoping study based on the methodology of Arksey & O'Malley [6]

2.3.1 The purpose of the literature review

Multiple researchers state that defining and describing the data governance structure within the data mesh is one of the challenges in data mesh implementation [43, 59, 26, 21, 24].

One of the reasons for this challenge is the lack of a reference model [59, 26, 21]. A reference model describes types of entities or domains and their relationships, including a description of the problem that it solves, and the concerns of the stakeholders who need the problem to be solved [71].

This literature review tries to contribute to the design and development of a guidance tool for DGS design in a data mesh by defining the entities and relationships within this governance structure that are discussed in the literature. This is done by answering the following question:

RQ2c: What are important data governance components in data mesh architectures?

2.3.2 Research protocol and training

This research uses a research protocol to improve its replicability. This protocol is a plan that describes the conduct of the proposed systematic literature review. This protocol is presented in **Figure 2.3**

Okoli advises changing the protocol on the basis of the use of trainees if possible, this is not
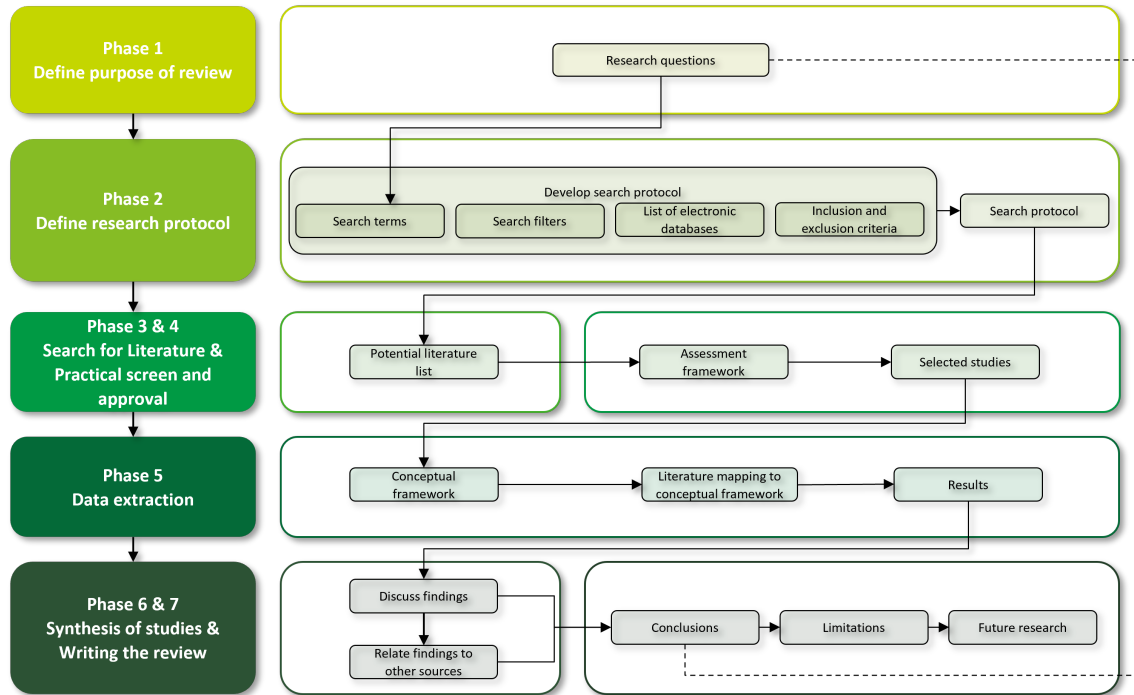
FIGURE 2.2: SLR Methodology based on the methodology of Okoli et al. (2010)

applicable in this literature research because of the individual execution. The formulated research questions have gone through multiple iteration steps. The last iteration resulted in the research question stated in **Section 2.3.1**. The main components of this research question are data governance and data mesh. The peer-reviewed papers on data mesh are very limited. Therefore, published articles that are not yet cited by others, are also included to increase the amount of used sources and possibly increase the generalizability of the results. Grey literature is excluded, however, published articles reviewing grey literature, are included.

Multiple electronic databases are queried under which: Google Scholar, Science Direct, the thesis repository of the University of Twente, the thesis repository of TU Delft and the data mesh academic knowledge platform on Slack.

The search strategy for Google Scholar and Science Direct differs from the search strategy for the data mesh academic knowledge platform. For Google Scholar and Science Direct, an iterative search term development process is followed in which the narrow and broad search terms are used and evaluated, the final iteration uses the following search term and filters:

Search term:

"Data mesh" AND "Data Governance" AND NOT "2d" AND NOT "3d" "3d mesh" AND NOT "EEG-based"

Filters:

- Publication period = 1-1-2019 - 1-10-2022
- Language = English OR Dutch
- Type = Article

For the data mesh academic knowledge platform on Slack, all resources published on this channel until 1-10-2022 are queried.
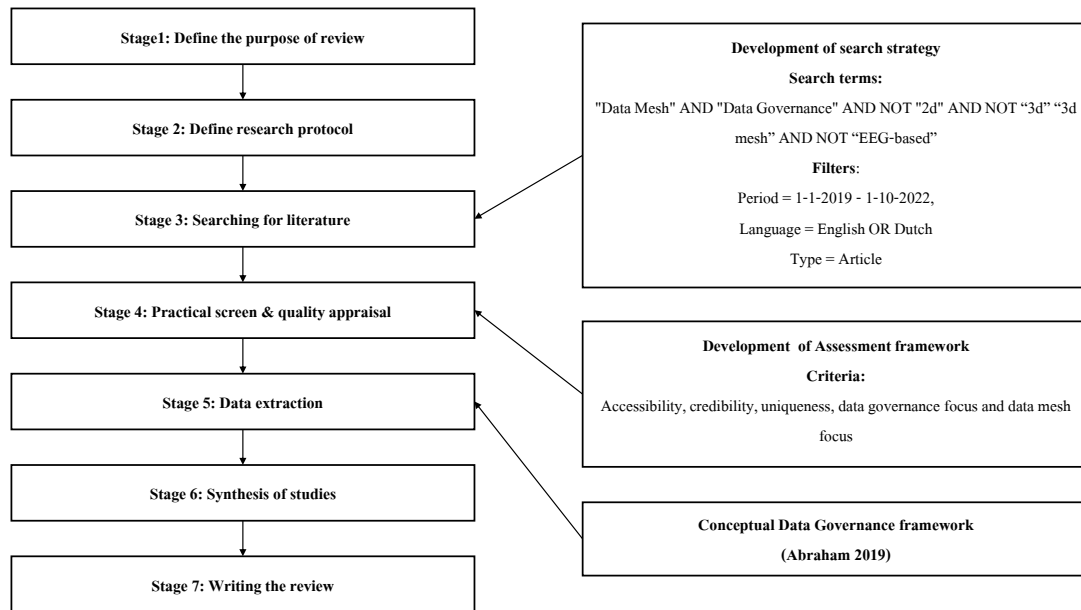
FIGURE 2.3: Research protocol of systematic literature review

2.3.3   Searching for the literature

With the search term, and filters described in the previous section, Google Scholar presented 55 articles, Science Direct presented 1 article, the Thesis repositories presented no results and the data mesh academic knowledge platform on Slack presented 20 sources.

2.3.4   Practical screen & quality appraisal

The practical screening and quality appraisal of the presented articles is done according to an assessment framework and inclusion and exclusion criteria.   The simplified visualization of this matrix is presented in **Figure 2.4**. The result of the use of the assessment framework is presented in **Appendix C**.

The article is considered "available" if the content of the article can be accessed for free through the publishers' website, via semantic scholar or with the license of the University of Twente. The quality and credibility of the article are assessed by checking for spelling mistakes, contradicting statements, use of references and possible vendor bias. The credible articles are read and the context in which data governance and data mesh is assessed. The process of the practical screening and quality assessment is visualised in **Figure 2.5** and its results are presented in **Table 2.2**. None of the sources originating from Slack is both peer-reviewed and scientifically relevant to this research. Therefore, all the sources queried from Slack are excluded.
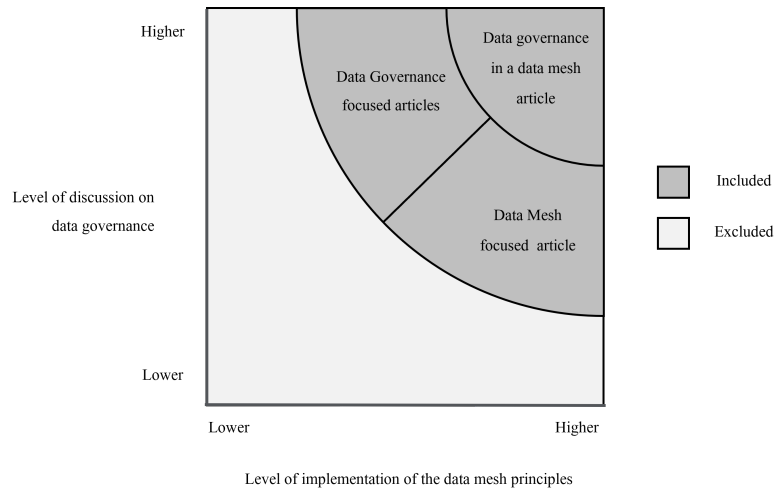
FIGURE 2.4: Inclusion and exclusion matrix used for literature screening
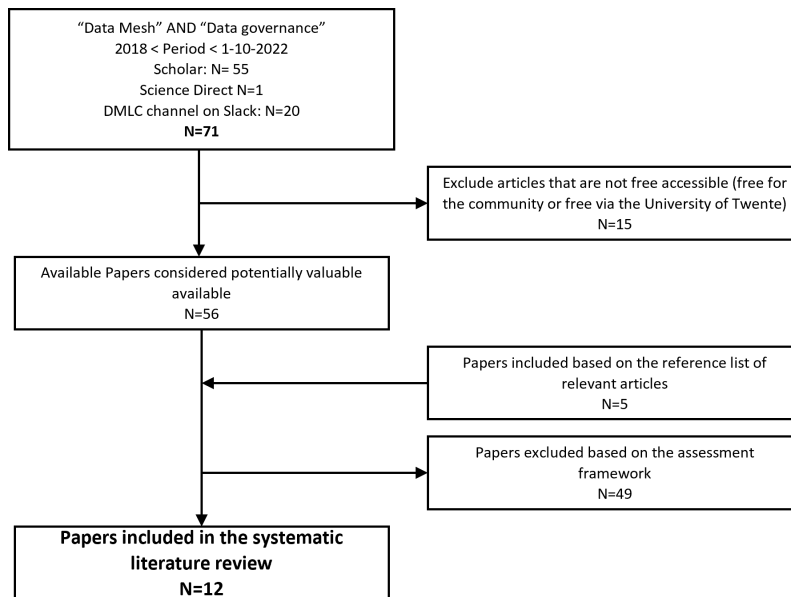
FIGURE 2.5: Visualization of the process of practical screening and quality appraisal

TABLE 2.2: Classification of articles for SLR

Article Classification	Amount	Source
Data governance in a data mesh articles	1	[43]
Data mesh focused articles	10	[58, 59, 37, 36, 34, 85, 57, 21, 29, 35]
Data Governance focused articles	1	[83]
Excluded	49	
Unavailable	15	
Total	**76**	

2.3.5 Data extraction

The data extraction is done through the use of an analytical framework based on the data governance reference framework proposed by Abraham (2019) [3]. This framework presents the concepts of data governance in organizations. Given that this research aims at defining the important entities and relations of data governance in a data mesh, this framework is expected to be able to function as a structural guideline to explain the concepts of data governance in a data mesh.

2.3.6 Synthesis of studies

The facts are extracted from the study using qualitative analysis techniques to find an answer to RQ2c. The statements on DGS elements are combined to form a conceptual description of the data governance entities and relationships of a data mesh. The results are presented in **Section 7.2**

2.4 Design science research methodology

The methods used to answer RQ3, RQ4 and RQ5 are based on the design science methodology proposed by Wieringa (2014) [82]

2.4.1 Motivation for DSR

As earlier mentioned, this research attempts to improve the guidance during the design process of DGSs during data mesh migrations. Normally in this type of organization structure design, models, tools or other artifacts are used to provide this guidance. However, according to literature [11] and interviews with experts (**Appendix D**), currently available artifacts do not provide sufficient guidance during the DGS design process.

Design science Research (DSR) is the design and investigation of artifacts that interact with a problem context in order to improve attributes of that context and accomplish stakeholders' goals [82]. The goals of this research methodology align with the goals of this research. In this research, an artifact is a guidance tool that supports design decision-making and that interacts with the process of designing the DGS for organizations that try to implement the data mesh principles in their organization. Wieringa (2014) created the following template to structure a design problem like this:

> **Improve**
> *a problem context*
> **by**
> (*re*)designing an artefact
> **that satisfies**
> *some requirements*
> **in order to**
> *help stakeholders achieve some goals*

When applied in the context of this study, the following design problem is introduced:

Improve
migrations to data mesh architectures
by
designing a guidance tool that supports deciding on data governance options
that
fulfils criteria on correctness, understandability, utility, efficiency and effectiveness
in order to
make the design of data governance structures more efficient without reducing its effectiveness

To solve this design problem, the DSR methodology describes three phases: Problem investigation, Treatment design and Treatment validation. How these phases are addressed during this study is presented in the next sections.

2.4.2 Problem investigation method

The knowledge from the preliminary literature research and conversations with practitioners are used to create the social context with its stakeholders, their goals and the requirements of the artifact that need to be met to achieve these goals.

Additionally, the theoretical background is used for the formulation of a semi-structured interview protocol for the investigation of the problem. This interview protocol is presented in **Appendix D**. During the interviews, the topic is introduced, a list of definitions used during the interview is presented, the goals of the interview are communicated and a list of topics with corresponding questions are discussed. The questions asked during the interview are predominately open questions. However, to increase the efficiency of the interview and improve the analysis, the list of questions also contained closed questions. Pre-designed follow-up questions are introduced in the interview to increase the consistency of the subjects. The interview protocol is first discussed with two enterprise architects to validate the interview structure and formulation of questions. The feedback resulting from this assessment is used to improve the interview protocol. This protocol is field-tested to further validate coverage, relevance, workflow and timing.

EA consultants from Energy 1 (N=2), EA consultants from Telecom 1 (N=2) and a data governance expert (N=1) are interviewed to investigate what problems occur during the design process of the DGSs in a data mesh. During these interviews, the causes and mechanisms of these problems are investigated and the effects of these problems on the stakeholders' goals are discussed. This knowledge is later used to design the treatment of the design problem. The results of these interviews are used to formulate the problem statement presented in **Chapter 1** and further elaborated in **Chapter 4**. The interviews with the EA consultants from Telecom 1 and the interview with the data governance expert are transcribed and presented in **Appendix D**. The interviews with the EA consultants from Energy 1 are conducted and important findings are noted. These findings are during the treatment design and development interviews validated and these discussions are used to further enhance the problem investigation phase. These interviews are presented in **Appendix F**.

2.4.3 Treatment design and development method

There is currently no method available that can directly be used to design and develop a guidance tool that is applicable in the context of this research. Therefore, multiple methodologies are combined to develop a new method for the design and development of this tool. The first process of the treatment design and development is the extraction of requirements for the tool. This is done based on the problem investigation interviews and the interviews conducted during the treatment design and development phase.

This method combines the methodology of Yim et al (2004) to develop a knowledge-based decision support system, the methodology of Eisenhardt (1989) to perform observational case studies and the methodology of Askren (1974) to develop the design options tree [28, 8, 84]. The method of Yim et al (2004) is chosen because of its focus on knowledge base development for decision-making which is perceived as useful in the context of this study. The methodology of Eisenhardt regarding case studies is followed because this describes an approach to theory building in case study research that emphasizes open questions, multiple case studies, and integrating conflicting results from the literature [82]. In addition, this methodology is used because this methodology has shown its applicability in inter-organizational DGS research by Benfeldt (2017) [9]. The methodology of Askren (1974) comprises a literature study, a practical analysis of design options and an extension with external experts [8]. The resulting methodology consists of 6 phases: 1) Define problems and identify sources of knowledge, 2) conceptualization of theoretical knowledge, 3) conceptualization of case-specific knowledge, 4) extraction of design decisions and configurations, 5) conceptualization and integration of expert knowledge, 6) formulation of knowledge model.

1. *Define problems and identify sources of knowledge*

   In this phase, the strategic problems are defined. As proposed by Yim et al (2004), this is done during the problem investigation method described in the previous section.

2. *Conceptualization of theoretical knowledge* In this phase, the book of Dehghani (2019) introducing data mesh is analysed through open and axial coding. Roles, responsibilities, team structures and interaction between teams are identified and with this information, a data governance operating framework is created.

3. *Conceptualization of case specific knowledge* According to Yim et al, "the partial knowledge from different sources needs to be reorganized and combined into an integrated knowledge model for conceptualizing the target management problem" [84]. Important during this phase is that according to the requirement of the created tool, the tool must only present the different design decisions and their configurations and does not aim at calculating the best decisions for the organization.

   To initiate practical design choices and organizational configurations, observational case studies are employed to examine real-world scenarios without implementing any interventions [82].

   The steps of The methodology of Eisenhardt are followed and the details of the implementation of these steps for this research are presented in Appendix . These case studies compose interviews with case representatives. Two experts from each case are interviewed according to the interview protocol presented in **Appendix F**. Following these steps resulted in the data governance operating

frameworks of organizations of the case studies.

4. *Extraction of design decisions and configurations*

   Differences between the theoretical data governance framework and the practical data governance frameworks are identified and considered to be design decisions made during the DGS design process. The design decisions and the resulting configurations from both case studies are compared and generalized if possible.

5. *Conceptualization and integration of expert knowledge*

   The results of the observational case studies are discussed with subject matter experts in the area of data governance (N=2), operating models (N=1) and data mesh (N=2). These interviews followed the protocol presented in **Appendix F**. Additional design decisions brought up by the experts are gathered and assessed for inclusion in the guidance tool.

6. *Formulation of knowledge model:*

   According to Yim et al, the formulation of the knowledge model is done through finding additional quantitative variables to support mathematical formulation. This study approaches this step differently because it takes a contingency approach to data governance which is advised by Weber (2007) [78]. Because of this approach, it is outside of the scope of this study to quantify variables and predict effectiveness mathematically accurate. The knowledge model created for this study is formulated as a qualitative knowledge model that entails the design decisions, the possible configurations and the factors influencing the decisions. This model has similarities with a Design Option Decision Tree (DODT)[8] as it presents the design options, the resulting configurations and the areas of interest to support the design options.

2.4.4 Treatment validation method

The validation of a treatment is the justification that it would contribute to stakeholder goals when implemented in the problem context [82]. The method used during this research follows the Framework for Evaluation in Design Science (FEDS) which consists of four steps: 1) explicate the goals of the evaluation, 2) choose the evaluation strategy or strategies, 3) determine the properties to evaluate, and 4) design the individual evaluation episode(s) [77]. The following section describes these steps in detail.

The goal of the validation study is to predict the effects of applying the artifact in a real-world problem context [82]. In this study, the goal is to validate whether the tool fulfils the requirements captured during the design and development phase (**Section 5.1**). These requirements address the correctness, understandability, and utility of the tool and the expected influence on the efficiency and effectiveness of the DGS design process. For this evaluation, the "human risk & effectiveness" evaluation is chosen because the critical goal of the evaluation is to rigorously establish that the utility will continue in real situations and because it is relatively cheap to evaluate with real users. The exact properties that have to be evaluated are requirements described in **Section 5.1**.

There are multiple ways to design the evaluation episodes. Due to the number of people that need to be involved in the process of designing a DGS, the length of this process and the need for a thorough investigation of all the organization-specific characteristics like goals, strategies and current structures, performing a real-life case study in which the

artifacts can be tested during the whole DGS design process is resource heavy. To limit the required resources needed for this research, there is chosen to conduct interviews with experts for both evaluation episodes.

In the first evaluation episode, interviews are conducted with EA consultants and subject matter experts with experience in DGS design. During these interviews, expert opinions are extracted about the satisfaction of the correctness requirements of the knowledge presented in the tool.

This group is chosen because these stakeholders have experience in designing DGSs for different organizations and can assess the correctness of the design decisions and configurations. **Appendix G.1.1** presents the interview protocol followed for these interviews.

During these interviews, the interviewees are asked whether they recognize the design decisions, configurations, influential factors, and advantages and disadvantages from their experience. The responses of the interviewees are captured in the evaluation document and additional comments are written down for analysis. The tool is refined based on the findings from this evaluation episode.

The second evaluation episode tries to capture whether the tool satisfies the understandably, utility, expected influence on efficiency and expected influence on effectivity requirements through expert interviews. These interviews are conducted with experts from the previous evaluation episode and with additional enterprise architects less experienced in DGS design. The reason for this is that the guidance tool should be understandable and usable by people that are not as experienced with data mesh migrations. The opinions of the intended users of the tool (enterprise architects EA consultants) regarding the understandability, utility, influence on efficiency and influence on effectiveness are essential. Additionally, the inclusion of experts from the first evaluation episode is important, as they can validate that the guidance tool is refined correctly through member checking [17]. The interview protocol and the results of the interview are presented in **Appendix G**.

2.5   Methods summary

As described in this chapter, different methods are used to answer the research questions formulated in the introduction chapter. An overview of the research questions, the methods used to answer the research questions and the chapters in which the answers are formulated is presented in **Table 2.3**. The methodologies presented in this chapter are followed and the findings are presented in the next chapters.

TABLE 2.3: Overview of RQ's with methods and chapters presenting the results

Research questions	Method	Chapter
1. What does the migration to a data mesh architecture look like?		
a. What differentiates data mesh architectures from traditional data platform architectures?	Literature review	3.1
b. What comprises a data mesh migration?	Literature review	3.3
c. What challenges occur in data mesh migrations?	Scoping study	3.3.2
2. How is the data governance in a data mesh structured?		
a. What is data governance?	Literature review	3.4
b. What conceptual components constitute data governance?	Literature review	3.4.1
c. What are important data governance components in data mesh architectures?	Systematic literature review	3.5
3. How are data governance structures designed?		
a. What are the key processes for designing the structure of organizational data governance?	literature research	3.6 & 4.4
b. What problems occur during the data governance design process?	Expert interviews	4
c. Which tools are currently used as guidance during the data governance design process?	Expert interviews	4 & 5.2
4. How to develop a design guidance tool for data governance structure design in a data mesh?		
a. What are the goals and requirements of the guidance tool?	Expert interviews	4 & 5
b. Which method can be used to design and develop the guidance tool?	Literature search	2.4.3
c. Which method can be used to validate the guidance tool?	Literature search	2.4.4
5. What constitutes a guidance tool for data governance structure design?		
a. What high-level components does the guidance tool need to possess?	Expert interviews	4
b. Which design decisions for data governance structures in data mesh architectures can be distinguished?	Expert interviews	5
c. Which organizational configurations for data governance structures in data mesh architectures can be distinguished?	Expert interviews	5
d. How can the tool be used during the data governance structure design process?	Expert interviews	5

Chapter 3

Theoretical background

This chapter presents the results of the literature review methods presented in the previous chapter. This chapter provides a theoretical background to construct answers to RQ1, RQ2 and RQ3a. **Section 3.1** presents an overview of the traditional data platform architectures. **Section 3.2** discusses the data mesh paradigm by introducing its principles, advantages, disadvantages and a comparison with the other data platform architectures. **Section 3.3** briefly explains data mesh migrations and presents the results of the scoping study on data mesh migration challenges. **Section 3.4** introduces the concept of data governance and the framework of Abraham (2019) that is used to define the components of data governance. **Section 3.5** presents the results of a systematic literature study on the data governance components in a data mesh. Finally, **Section 3.6** explains how data governance and DGSs are designed.

3.1 Traditional data platform Architectures

This section presents the data warehouse architecture, the data lake architecture, and the data lakehouse architecture and discusses the limitations of these architectures. This is important background information to construct an answer to RQ1a.

The recent developments in data usage in organizations create the need for data architectures that support these new usages for data. This section describes these architectures and visualizes these architectures in **Section 3.1.3**

A data platform is an integrated set of technologies that collectively meet an organization's end-to-end data needs [74]. It enables the acquisition, storage, preparation, delivery, security and governance of data [74]. An important part of a data platform is the data warehousing architecture . This describes the technology that ingests operational data, aggregates and processes these data and holds as output the analytical data that can be used to extract analytical value for decision-making [68].

3.1.1 1st generation DPA: Data warehouse architecture

The data warehousing architectures of today are influenced by concepts like facts and dimensions formulated in the 1960s [23]. These architectures are responsible for moving the data from the operational systems to the applications that support data analysis. This is done through the extraction of data from operational databases and sources, transforming this data into a universal schema and loading this data in warehouse tables which can be

accessed through SQL-like queries [23]. This data warehouse is built for processing refined structured data. A simple visualization of a data warehouse architecture is presented in **Section 3.1.3**).

These data warehouses were fulfilling business needs for a long time. Organisations that implemented a data warehouse architecture experienced for example improvements in data quality and consistency, faster data retrieval, the possibility of analysing historical data and enhanced business intelligence [52]. However, the rise of Big data has put pressure on these architectures. Big Data is the information asset characterised by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value [22]. Data warehouses are not built for data with some of these characteristics. For example, the data format of big data is not always known before it is collected, which means that the data pipelines for cleaning and processing the data, could be unsuitable for that specific format. This limits the process abilities of the data by the techniques used within a data warehouse and obstructs the analysis of this data.

3.1.2 2nd generation DPA: Data lake architecture

The concept of a data lake merges to overcome the difficulties faced when processing big data in a data warehouse [46]. Just like the data warehouse architecture, the data in a data lake architecture gets extracted from the operational systems and loaded into a central repository often in the format of an object store. It is possible to store data in its original form until processing is needed, so new data does not interfere with the data structure, which allows continuous injection of data into the data lake [46, 42]. This flexible storage preserves scalability and performance [59] and aims to improve the ineffectiveness and friction of extensive up-front modelling that a data warehouse architecture demands [23]. Because of these benefits, it is possible to train models against real production queries.

In these architectures, data is extracted from many operational database sources and represents as much as possible of the original content and structure. Data is loaded to scalable object storage and accessed through the object storage interfaces by data scientists that use the data for analytical and machine learning model training. More to the analytical side, lakeshore marts are created as fit-for-purpose data marts which could be used as data warehouses [23]. Even though a data lake architecture can have a lot of benefits, it also creates complex and unwieldy pipelines of batch and streaming jobs that should be operated by a specialized data team. Additionally, the realized value and performance of a data lake are highly dependent on effective metadata management, data quality management and data reliability which can be challenging to achieve [59].

3.1.3 3rd generation DPA: Data lakehouse

Data lakes and Data warehouses have their benefits and limitations as earlier described. An effort to combine the benefits of data lakes and data warehouses resulted in a new data platform architecture that is built around a new system called a Data lakehouse. This system is defined by Armbrust as "a data management system based on low-cost and directly-accessible storage that also provides traditional analytical database management system management and performance features. Lakehouses thus combine the key benefits of data lakes and data warehouses: low-cost storage in an open format accessible by a variety of systems from the former, and powerful management and optimization features from the latter" [7].

In a data lakehouse, a transactional metadata layer is implemented on top of the object store that defines which objects are part of a table version and which are not [7]. This layer can enforce specific metadata schemes to ensure interoperability with the stored data and it supports users to define specific constraints for the ingested data through the data lake pipelines to improve quality [48]. A data lakehouse provides direct access to BI tools to enable advanced analytics and it allows the implementation of different kinds of APIs and machine learning libraries to meet the expectations of data engineers and data scientists [48]. Data lakehouses can support streaming data from for example Kafka and Spark [42]. The metadata layer, which also contains a data governance layer, improves the architecture and can for some organizations be sufficient to solve problems like data quality, interoperability and findability. Even though the implementation of this system in the data platform architecture seems to solve some of the problems from earlier data platform architectures, a data lakehouse is not yet the answer to all current problems.
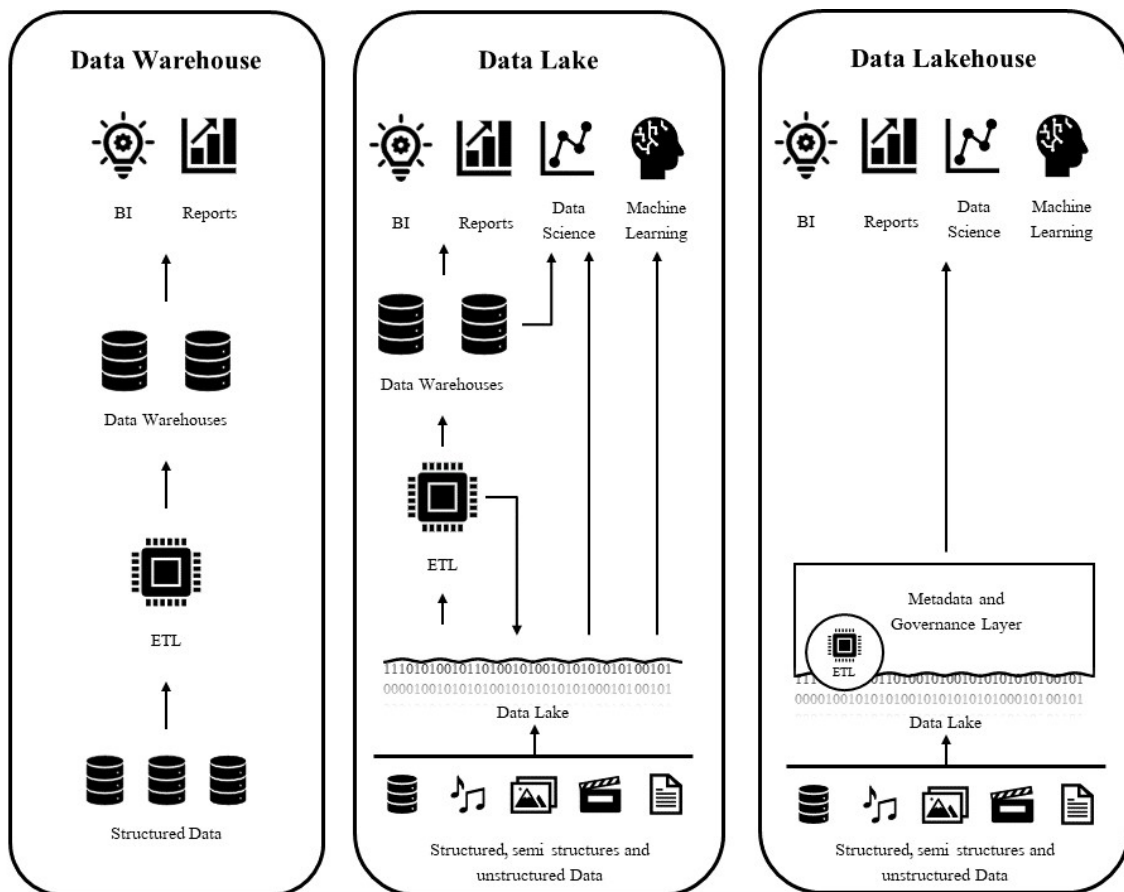
FIGURE 3.1: Simple visualization of different data platform architectures (adapted from Databricks [56])

Limitations of current DPAs

The technical characteristics of data warehouses, data lakes and data lakehouses are very different, but the organizational characteristics of these architectural patterns are rather similar. There are three different kinds of teams with specific expertise and responsibilities: source teams that generate the data, one central data team that is responsible for the flow

and storage of data within the organization, and multiple analytical teams that extract the data to perform analyse [23]. Ensuring the usability of these data takes time and resources to do. The source teams have limited incentive to ensure their information is useful because they do not have direct benefits of its usability and they do not have the responsibility to ensure the data quality [10]. This means that there is one single hyper-specialized data team with data and machine learning engineers that is responsible for all the data within the data platform. These data engineers end up spending a lot of time cleaning up data from the source teams and making it useful for the analytical teams. This makes the roles of the central data team more difficult as they need to provide useful data for a diverse set of operational and analytical needs, without a clear understanding of the data and the source domains that generate the data and with a lack of domain expertise in their team [23]. Furthermore, the lack of communication between the teams that create, transform and analyze the data, results in a lack of consensus regarding the goals and standards of the data [38]. These limitations are practically presented in a case study of one of the biggest online clothing retailers (Zalando). This study describes a data platform architecture that is based on the data lake architecture. This case study showed that their data platform architecture had a clear ownership issue: data producers are unaware of the intended purpose of the data and do not feel responsible for the quality, while the data analysts, who benefit from high-quality data, lack knowledge about the source of data and can not improve the quality of the data [2, 21]. This means that all the responsibilities end up at the central data team.

Dehghani (2022) states that there is an unsatisfactory alignment between the organizational needs and the architectures instituted [24]. Due to the growing needs of the organization in terms of volume and variety of data, the centralized data teams are overloaded with responsibilities [24]. These responsibilities range from ad-hoc exploration for analytical purposes to central ETL pipeline management, and the exact responsibilities within these operations are often not fully respected or unclear [24]. In addition, there is a lack of domain knowledge in this central data team. The analytical teams are using data for different use cases and the core data team is not able to satisfy the demand of switching business contexts for which they have no expertise [43]. This leads to data latency and increased lead time for data analytics.

Most of these limitations can be redirected to the fact that the consumption, storage, transformation, and output of data are handled in a central warehousing system (Data warehouse, data lake, data lakehouse) which is managed by one central team (the data team). In software development, an architecture with such a single deployment unit of code is called a monolithic software architecture [66]. In the architectural design pattern of the current generation's data platform architectures, this monolithic pattern can be recognized. Monolithic architectures are generally acknowledged in the literature to be tightly coupled and challenging to maintain [62]. Continuous deployment in these large monolithic architectures is very challenging because a change to a small portion of the application necessitates rebuilding the entire monolithic system [66]. Change implementations in the massive, complex, and rapidly expanding monolithic systems of today will be too slow and inefficient in the near future due to the aforementioned limitations [47]. These limitations are similar to the limitations recognized in the centralized data teams of traditional data platform architectures [23].

3.2 Data mesh architecture

This section introduces the data mesh architecture by explaining the approach based on the different kinds of data used within this approach, by explaining the principles that form a data mesh architecture and by presenting a comparison of the traditional data platform architectures with a data mesh architecture. This and the previous section collaboratively are used to construct an answer to RQ1a.

Dehghani (2019) proposes a new approach that could evolve the monolithic data architectures into a more distributed architecture and possibly function as the paradigm shift that is needed in data platform architectures [24]. This data mesh is a decentralized socio-technical approach to sharing, accessing and managing analytical data in complex and large-scale environments [23].

3.2.1 Data in a data mesh

Before diving deeper into the principles and characterisation of a data mesh, it is important to define the context of the data. Dehghani distinguishes two kinds of data that are used in data platforms: Operational data and Analytical data.

Operational data

Operational data supports running the business and keeps the current state of the business with transactional integrity [24]. This data is captured, stored and processed by transactions in real-time, by online transaction processing systems. This data sits in databases of microservices, applications or other systems of records that support the business capabilities [23]. Operational data is referred to as "data on the inside" and it is the private data of an application or a microservice that performs create, update and delete operations on this data. In an organization, the operational data alone does not allow to support decision-making processes. For this to happen, there must be mechanisms that extract perceptible analytical value from this operational data, so that it can reach the respective stakeholders [68].

Analytical data

Analytical data is the historical integrated and aggregated view of data created as the byproduct of running the business [23]. This data is maintained by online analytical processing systems. This data is a temporal, historic and often aggregated view of the facts of the business over time [23]. This data is used by analytical consumers to provide retro- or future perspective insights, train machine learning models, and create visualizations. It is used to optimize the business and the user experience. This is normally the data that is stored in a data warehouse or data lake [23].

3.2.2 Data mesh principles

The Data mesh arises as a paradigm shift that occurs both at the technological and organizational levels. This change has the purpose of solving current problems in data management such as increased pressure on central data teams, unclear defined roles and responsibilities, low data quality, long turnaround times and high data management costs [59]. Data mesh tries to solve these problems by accepting and managing the diverse origin of data. The main purpose of the data mesh architecture is to create a

decentralized data architecture that enables the extraction of large-scale analytical data [59]. This is according to Zhamak Dehghani achieved through the implementation of four principles: 1) defining domain-driven ownership of data, 2) threatening data as a product, 3) development of a self-serve data platform and 4) the implementation of federated computational governance [23].

Domain-driven ownership of data

The first principle is built upon the decentralization of the data platform into several domains. The decentralized components of the former large whole are selected on the domain they carry within the organization [21]. The domain-driven design principle encourages domains to work independently and autonomously on their own data. This way, the domains are less dependent in the large data team and they could process, manage and analyse the data by themselves. Dividing the enterprise into domains and clearly scoping the roles and responsibilities of such a domain improves the interaction between the domain experts and data engineers [14]. This way, the responsibilities of the data are closer to the source and distributed over the areas of operation (domains), instead of centred with the centralized data team. The team that is in control of the data can better understand the data because they are closer to the domain knowledge experts on that data [58].

These domains could have a lot of flexibility on how to create, process and analyse their data, as long as it adheres to global standards. The existence of these global standards still creates the need for one centralized IT/Data team that performs organization-wide implementation and supports standardization practices [18]. The domains can provide their own APIs as well as their analytical data endpoints.

Data as a product

The second principle proposed by Dehghani is based on the concept of product thinking which is used as a perspective to look at data [23]. The focus of the organization should shift from creating data pipelines and storing the data to the data itself [59]. The data should be ready to use for analytical purposes. With this way of thinking, current roles within the organization change and new ones arise. Examples of new roles are the domain data product owner and the data product developer. Compared to the current architectures, the responsibilities of data quality are shifted. In data lakes and data warehouses, creating data with quality and integrity resides downstream from the source and remains with the centralized data team, while a data mesh shifts this responsibility close to the source of the data [23].

The people using the data are considered customers with their own customers' needs. These needs could differ per "customer", per product and per context. However according to Dehghani, in any case, the data product must possess the "baseline data product usability attributes" [23]. These attributes are Discoverable, Addressable, Understandable, Trustworthy, Natively accessible, Interoperable, Valuable and Secure. These principles extend the FAIR Principles (findable, accessible, interoperable and reusable) [41]. Dehghani states that the newly introduced characteristics are necessary to make distributed data ownership work [23].

Self-serve data platform

Because of the domain-oriented design principle, the data management capability and data literacy are federated across the organization. People with relatively little data experience are designated to build maintain and manage their data products. Therefore, they need a platform that could encapsulate the complexity of the offered data products. The self-service infrastructure and platform functions could function as the needed platform by providing tools and interfaces that can be used for the creation and maintenance of data products [24]. It reduced the required level of specialization for the product developers, decreases the duplication of efforts by avoiding that each domain works on its own method of creating data products, and it reduces discrepancies between products and therefore improves interoperability [34].

For this data platform to adhere to the requirements, it should be developed by the central platform team that consists of highly specialized architecture developers. There should be a balance between implementing functions according to the domain needs and thereby possibly making the platform resource-heavy to maintain, and offering too few capabilities and thereby leaving the domains to do their tooling and possibly reducing the compatibility of the general platform and the possibility to implement computational governance [34].

The development of this platform is an iterative process that needs an alignment between business and IT. The central platform team (IT) has to communicate with all the domain teams (business) to collect and understand the business needs and plan the implementation accordingly.

Federated computational governance

Finally, the interoperability between these domains should be enabled by the implementation of policies and standards [31]. Because the data, the data model or the understanding of data is flexible, domain-specific and flexible, the governance of these policies and standards should be dynamic. The model used for this governance should allow decentralization, and independence of each domain in the data mesh and it should be able to automatically execute decisions. This needs global normalization (federated computational governance according to Zhamak Dehghani) which aims to apply a set of rules to all interfaces of the various data products, however, the domains and data teams need autonomy to model and push the data and therefore, the governance model should also respect flexibility within the domains [24].

3.2.3 Advantages of data mesh

Firstly, the division into domains brings the operational, data and analytical team of a single domain closer and therefore the use of the data closer to the source and the domain knowledge of that data. Also, this division shortens the lead time between the operational data teams and the analytical data teams by bringing them together in one domain [21, 34]. This results in shorter ETL pipelines within the domain. Additionally, the responsibility of the data and its quality is shifted closer to the source of the data. These responsibilities are easier to carry out with domain knowledge of that data. These clear responsibilities also improve the usage of data due to the way data products can be published and discovered on the self-serve infrastructure [21]. Other people within the organization can access these products through this infrastructure. Finally, the

computational governance within the data mesh reduces the workload on the data specialists by automating decision-making.

3.2.4 Disadvantages of data mesh

Even though the minimal amount of case studies have shown that the implementation of this architecture in an enterprise has a lot of potential and could solve problems like lack of data quality, lack of data findability, data latency and lack of interoperability [59], the implementation of this socio-technical system and its principles, influences the enterprise architecture as a whole and is very complex [59].

First of all, there is a greater need for data literacy within a data mesh architecture because the responsibilities of the data are no longer centralized in one single team [21, 34]. Because of the division of ownership and responsibilities among the domains, data management capabilities are needed in all these domains. This requires more data experts or the training of current employees. In addition, to ensure global interoperability, there will always be a need for a central body within the data mesh. The boundaries between the dependence of the central data team and the independence of the domains themselves can sometimes be vague. Furthermore, the domains use the data product of other domains for new data products. This leads to data from source domains being duplicated in different data products. This increases the storage costs of data and as data is stored in different locations in an organization, it will become more difficult to govern [76].

3.2.5 Comparison of different data platform architectures

The structure, implementation and use of data warehouses, data lakes, data lakehouses and data meshes are different [59]. A data lake deals with raw data that can be structured, semi-structured and unstructured, while a data warehouse only deals with structured data. The type of repository in a data warehouse is less agile due to the stricter data requirements and the storage costs are much higher compared to the data lake [44]. The general users of data lakes and data warehouses also differ, the data warehouse mainly targets professional business users while the data lake mainly targets data experts with relatively more data expertise and experience due to the complexity of unstructured data [44]. The Data lake can be considered more effective and efficient to deal with heavy workloads, compared to the already established data warehouse [59]. A data mesh architecture, on the other hand, does not limit the data platform to a single data warehouse. It is more an organizational approach towards data platform architectures than the technical implementation of a data platform. The full comparison is shown in the **Table 3.1**

3.3 Data mesh migration

This section presents the results of a review of different roadmaps that are available on data mesh migrations and a scoping study on the challenges of these data mesh migrations. The findings of these reviews contribute to constructing an answer to RQ1b and RQ1c.

The migration towards a data mesh architecture impacts both the organizational and technical structure of the organization. This makes it a large organizational transition which can be very complex. Strategic roadmaps are commonly used in large organizational transformations like these to make the migration simpler, faster and less costly [15].

Because of the novelty of the concept, there is a limited amount of roadmaps available

TABLE 3.1: Comparison of the different generations data platform architectures and the data mesh

Concepts	Data Warehouse	Data Lake	Data Lakehouse	Data Mesh
Data Format	Relational data store, physical storage proprietary	File/object storage based on open file formats	File/object storage based on open file formats	Different data formats with Polyglot Persistence
Data type	Structured data, limited support for semi-structured data	Structured data, semi-structured data, textual data, unstructured (raw) data	Structured data, semi-structured data, textual data, unstructured (raw) data	Structured data, semi-structured data, textual data, unstructured (raw) data
Data actuality	Usually daily, limited by batch/ETL cycles	Almost real-time possible through streaming	Combination of batch/ETL cycles and streaming possible	Up to real time possible through data virtualization (depending on the data product)
Data access	SQL	Open APIs for file/object access, limited SQL access	Open APIs for file/object access, SQL	Open APIs for operational data products, SQL (with data virtualization) for analytical data products
Reliabiity	High quality and reliability, ACID transactions	Low quality and reliability	Low quality and reliability for raw data, high quality, ACID transactions for prepared data	Depending on the data product
Data Governance and security	Fine-grained security and governance (row/column level)	Weak security and governance due to file level	Fine-grained for SQL access, weak security and governance for file/object access	Depending on the data product
Performance	High because it can be specifically optimized	Rather low as file/object-based, dependent on data usage (MapReduce, Spark)	Medium, as up to now limited possibilities for optimization	Depending on the data product
Scalability	Scaling becomes exponentially more expensive	Highly scalable for large amounts of data at low cost	Highly scalable for large amounts of data at low cost	Depending on the data product
Supported use cases	Classic BI, reports, dashboards, SQL	Data Science, specifically machine learning	Diverse, from classic BI to self-service BI to machine learning	Depending on the data product, classic BI, self-service BI, machine learning, also operational applications
Advantages	Reliable, high performance	Support semi-structured data and can be used for Machine learning	Low-cost storage, support semi-structured data and can be used for Machine learning, interoperability through metadata layer	Data usage closer to source, shorter lead times, clear responsibilities, the distributed workload of the central data team and reduces workload due to computational governance.
Disadvantages	limited support for semi-structured data, domain knowledge and data responsibilities are separated, lack of ownership, centralization of responsibilities	Complex pipelines of batch and streaming jobs, domain knowledge and data responsibilities are separated, lack of ownership, centralization of responsibilities	limited support for semi-structured data, domain knowledge and data responsibilities are separated, lack of ownership, centralization of responsibilities	Complex to implement in an organization, need for data literacy, need for well balanced local vs global data governance, increase storage costs due to data duplication

for data mesh migrations. The available roadmaps are generic which is understandable because every company's migration will be unique. However, two of the three available roadmaps present several common elements and considerations that could be useful in data mesh migrations.

3.3.1 Data mesh migration roadmaps

One of these two roadmaps is the "Enterprise Data Mesh Journey" of towardsdatascience.com [15]. This roadmap divides the migration process into 3 different phases: The first phase is the strategy stream phase in which the strategy is defined for the data mesh journey, the second phase is the parallel working phase where multiple streams of work are executed in parallel and last phase is the rollout phase in which the Enterprise Data mesh is rolled out. This roadmap extends a previously designed roadmap of www.thoughtworks.com [32] which does not contain the socialization stream within the parallel stream of work segment. The roadmap with the socialization stream is presented in **Figure 3.2**.
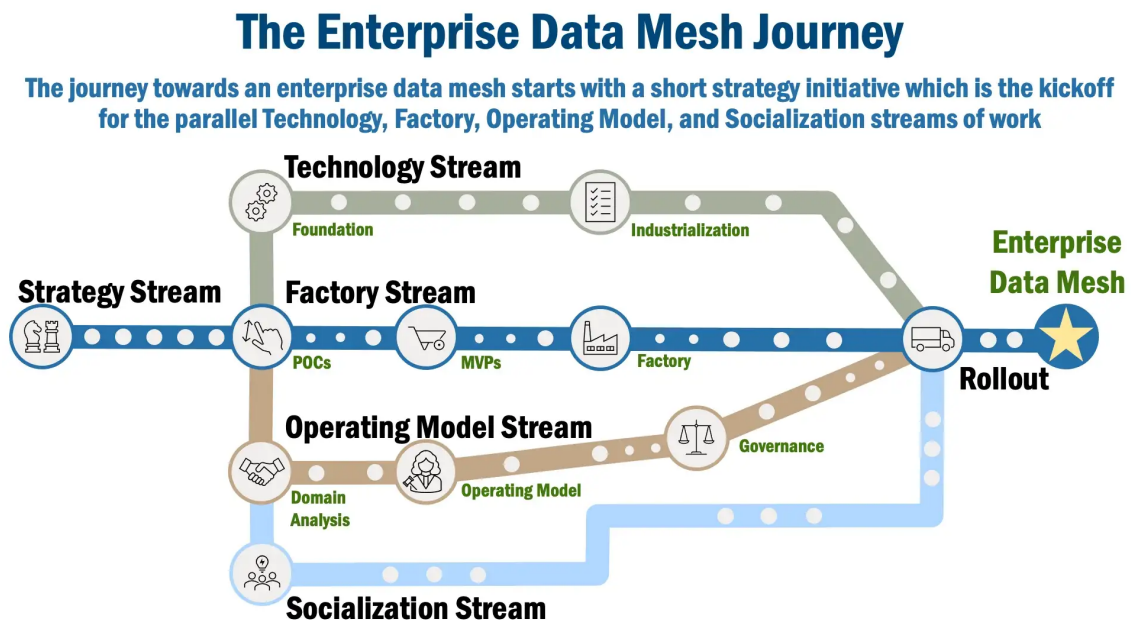
Figure 3.2: The Enterprise data mesh Journey adapted from [15]

This roadmap consists of the following streams (adapted from [15]):

- **Strategy Stream**, which provides the broader business and technical vision and expected outcomes for the data mesh.

- **Technology Stream**, which defines the technology foundation and industrialization activities required for the data mesh.

- **Factory Stream**, that introduces repeatable processes and templates to permit rapid scaling of the data mesh.

- **Operating Model Stream**, which defines the team structure, interactions, and governance techniques to build and operate the data mesh.

- **Socialization Stream**, that is used to not only communicate successes but also to continuously build the momentum required to build the data mesh.

- **Rollout Stream**, that accelerates the adoption of data products with the data mesh.

The problem described in the problem statement mostly addresses the Operating Model Stream. This Operating Model Stream is presented in **Figure 3.3**. Within this operating model stream, the problem occurs in the operating model definition. During this process, the DGS is designed and the data governance program is defined.
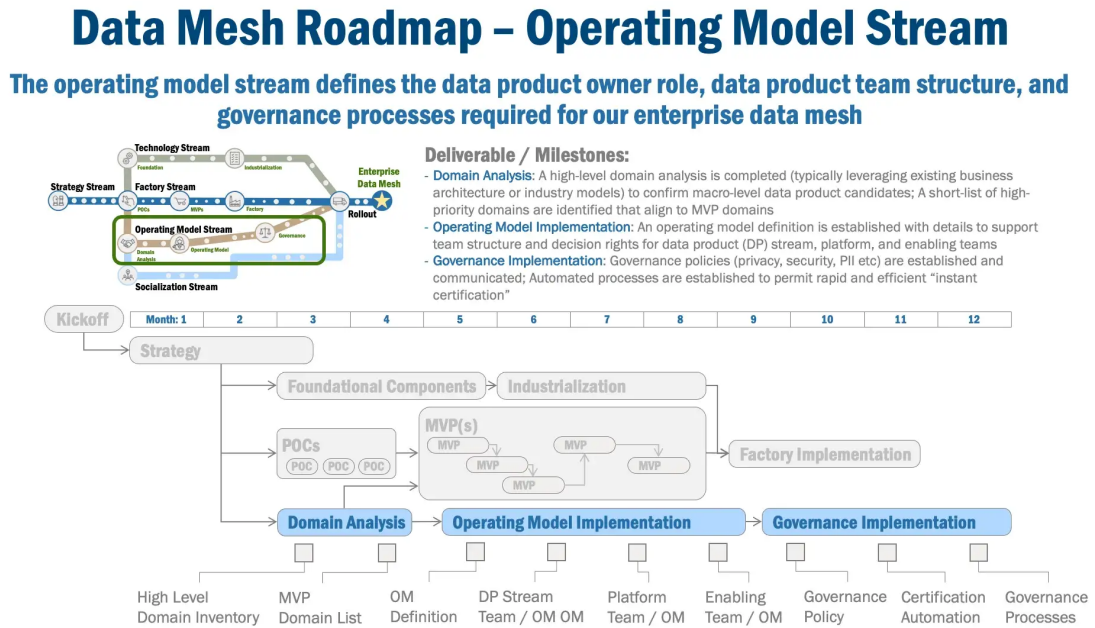
FIGURE 3.3: The Operating model stream of the data mesh journey adapted from [15]

3.3.2 Data mesh migration challenges

Because of the earlier described differences between the traditional platform architecture and the data mesh architecture, there are a lot of components of the traditional data platform architecture that need to change in order to migrate to a data mesh architecture. Changing from one data platform architecture to another can be difficult. The scoping study found that multiple challenges hamper the implementation of and migration to a data mesh. These challenges occur both in the organizational and technological layers of the organizations. This result is expected due to the fact that data mesh is a socio-technical approach and that it significantly influences both layers. However, it is interesting to see that creating changes in both layers also actually results in challenges in both layers.

When looking at the specific challenges, the results of the scoping study presented in **Figure 2.1** show that most challenges are classified as "process challenges". However, when diving deeper into the meaning of the challenges, these process challenges often find their origin in other perspectives. For example: "defining and describing the governance structure" is a process, but to be able to define and describe the governance structure, the people currently in this structure, their capabilities, and their responsibilities should be well communicated in order for the structure to fit the current organization and the

people doing this process should have the understanding of how to define and describe the governance structure. Multiple challenges interface with more than one perspective. Therefore, a note of caution is due here since the classifications of tier 1 challenges are not mutually exclusive.

The results of tier 2 and tier 3 challenges, on the other hand, are interesting and useful to discuss. One expected finding is that achieving the required level of data literacy is a challenge. This is a widely known challenge in the transformation to a data-driven organization. This in combination with the dispersion of data responsibilities to people that normally do not work with data and therefore are expected to lack data literacy for their new role, could be the reason why data literacy is a challenge mentioned in literature.

The most interesting finding of the scoping study is that multiple challenges can be redirected to the data governance within a data mesh [43, 59, 26, 65, 37, 21, 16, 20, 57]. These challenges are setting up the DGS and realizing data capabilities, roles and responsibilities in the business domains. In addition, the data quality and technical integration & standardization challenges are indirectly related to data governance as these challenges could be partly addressed by the technical implementation of a well-defined automated DGS [59]. Apparently, designing the DGS and defining its components is complicated according to the literature. The literature underexposes why this is challenging during data mesh migrations, however, an explanation could be that the DGS in a data mesh differs significantly from the traditional DGSs in data platforms. The main difference is the federated data governance model. In this model, there should be a balance between global standardization and domain sovereignty. The balance is difficult to achieve and can be different for each organization. In addition, the migration to a federated governance model can be complex due to the redistribution of roles and responsibilities. This, in combination with the novelty of the data mesh approach, the differences in governance between organizations and the lack of guidelines and best practices supporting the data mesh migration could cause problems while defining and describing the governance structure within an organization.

In short, there are multiple challenges hampering the implementation of a data mesh in organizations, these challenges are organizational challenges, technical challenges that result from organizational challenges and purely technical challenges. A lot of these challenges are data governance related and addressing these challenges requires a governance structure. **Section 3.4** introduces data governance and a framework to structure the components of data governance. **Section 3.5** discusses the findings of the systematic literature review on components of DGSs in data meshes according to this framework.

3.4 Data governance

This section presents the results of a literature review on data governance and the findings of a systematic literature review on the important components of data governance in a data mesh based on the framework of Abraham (2019). The findings of these reviews contribute to constructing an answer to RQ2a, RQ2b and RQ2c.

Data governance is more related to corporate governance than to data management. Governance is defined as the exercise of authority and control over a process, organization or geopolitical area. The process of setting, controlling, administering and

monitoring conformance with policy [51]. Data governance is in essence similar but the authority combined with policy to ensure the proper management of data.

Good data governance can help organizations to create a clear mission, achieve clarity, increase confidence in using the organizational data, establish accountabilities, maintain scope and focus, and define measurable successes [4]. In addition, other benefits include training management and staff to adopt common approaches to data issues, build standard, repeatable processes, reduce costs and increase effectiveness through coordination of efforts, and ensure transparency of processes [4]. The standards, processes, roles and responsibilities resulting from data governance activities can address all knowledge areas of the management of data: Data Architecture, Data Modeling & Design, Data Storage & Operations, Data Security, Data Integration & Interoperability, Documents & Content Management, Reference & Master Data, Data Warehousing & Business Intelligence, Metadata and Data Quality [39]. It makes sure that the enterprise principles are translated into the way data is managed.

3.4.1 Data governance frameworks

Making decisions on data management can be complex, so to support organizations, there are multiple frameworks developed that try to capture the different components of data governance and align these components to the objectives of the organization. Some of these frameworks are: the DAMA-DMBOK Functional Framework [13], the Path to Modern Data Governance [80], the DGI Data Governance Framework [73], the PwC Enterprise Data Governance Framework [67] and Abraham's Conceptual data governance framework [3].

The following section uses the framework of Abraham to describe common components of data governance programs in data mesh architectures. The framework of Abraham (2019) is chosen due to its explicit focus on data governance, wide adoption in literature, extensive coverage of data governance concepts, and perceived usability in the context of this research.

3.4.2 Antecedents for data governance

The antecedents of data governance describe the external and internal factors that precede or predict the adoption of data governance practices [3]. These antecedents can be internal or external. Internal antecedents comprise organization strategy, IT strategy, diversification breadth, Corporate location of decisions making authority, degree of business process harmonization, IT architecture, Organization culture, and Senior Management support. External antecedents comprise legal and regulatory requirements that are context-dependent, highly regulated markets that require a more centralized organizational structure than markets with fewer or no regulations and the specific industry or location-dependent requirements that enforce a (new) DGS.

3.4.3 Data governance mechanisms

Mechanisms help to plan and control data management activities. Governance mechanisms comprise a formal structure connecting business, IT and data management functions, and formal processes supporting the active participation of and collaboration among stakeholders. [3] distinguishes three different kinds of mechanisms: structural
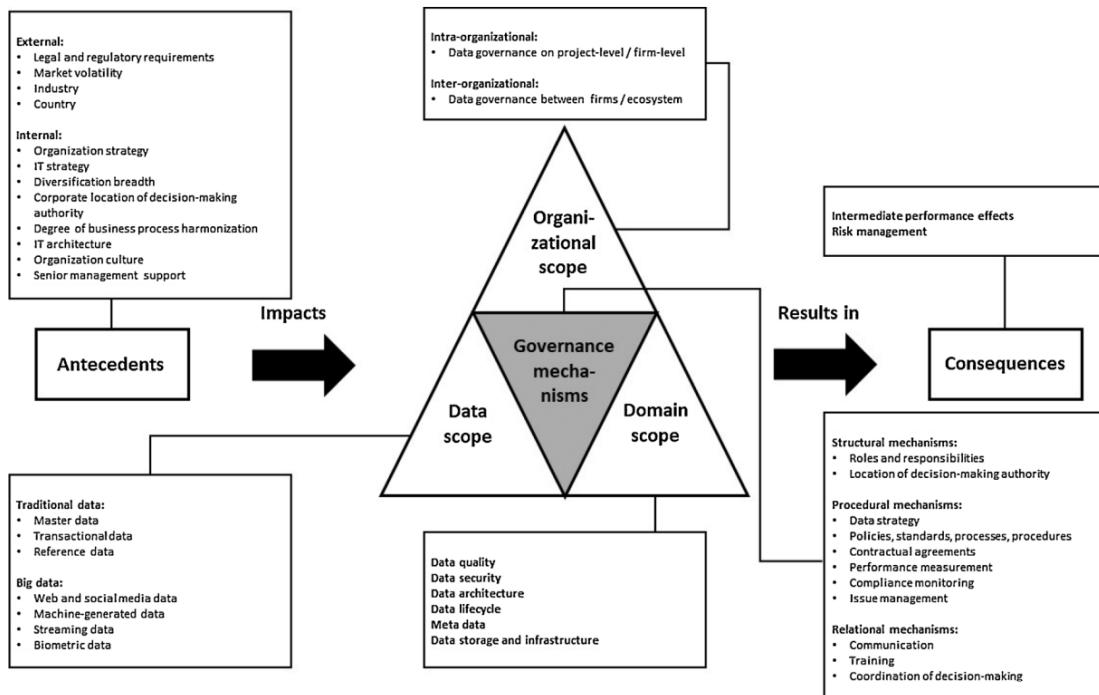
FIGURE 3.4: Concepts within the conceptual framework for data governance from Abraham et al. [3]

governance mechanisms, procedural governance mechanisms, and relational governance mechanisms.

Structural mechanisms

Structural mechanisms describe the roles, responsibilities and relational decision-making authority (central vs decentralized). Roles and responsibilities can be different for every DGS. Well-defined roles and responsibilities awareness are important for the governance structure to get the people acting in the system aligned. Examples of data governance roles are executive sponsor, data governance leader, data owner, data steward data governance council, data governance office, data producer, and data consumer. The second concept in structural governance mechanisms is the allocation of decision-making authority, this determines which organizational unit has the mandate for action related to data governance [3]. This concept distinguishes hierarchical positions and functional positions, and centralized and decentralized decision-making.

Procedural mechanisms

Procedural mechanisms aim to ensure that data is recorded accurately, held securely, used effectively and shared appropriately [3]. They comprise the data strategy, policies, standards, processes, procedures, contractual agreements, performance measurement, compliance monitoring and issue management. These concepts are described by Abraham (2019) as follows:

- The data strategy defines the high-level course of action based on strategic business objectives and it consists of a vision statement, a business case, guiding principles,

long-term and short-term objectives and an implementation roadmap.

- Data policies are the high-level guidelines and rules regarding the creation, acquisition, storage, security, quality and permissible use of data.

- Data standards ensure that the data representation and the execution of data-related activities are consistent and normalized throughout the organization. They facilitate interoperability and fit for purpose. The data standards are internally defined by data stewards and data architects or externally by standardization organizations like ISO).

- Data processes are standardized documented repeatable methods used to govern data. Examples are development and maintenance rules for data handling as well as for data modelling, but also processes for the current state, processes for the alignment and validation of policies, processes for decision-making and performance measurement

- Data procedures are the documented methods techniques and steps followed to accomplish a specific activity or task.

- Data provisioning and data sharing settings are the required contractual agreements between participating internal departments or external organizations.

- Performance measurement aims at assessing the effectiveness of data governance by measuring the level of goal attainment. Performance measures on the firm level are based on strategic business goals such as revenue growth, increased profitability, and cost savings.

- Compliance monitoring aims at tracking and enforcing conformance with regulatory requirements and organizational policies standards and SLAs.

- Issue management is the identification, management and resolution of data-related issues.

Relational mechanisms

Relational governance takes the relations between stakeholders in the governance structure into account. It encompasses communication, training and coordination of decision-making. The communication concept describes the way data governance awareness is created and how it is presented to other stakeholders in the system. Effective communication in this scope depends on the goal of the data governance program and the people that are targeted to contribute. When the stakeholders are already in the program but need to have additional knowledge or qualification to support the implementation of data governance, the organization should focus on training. This can help operationalize data policies, processes and procedures. This in combination with communication can facilitate the creation of a data culture that values their data [3]. Lastly, the coordination of decision-making is a mechanism that can be important in a DGS. This explains a vertical or horizontal approach to the coordination of decision-making. A vertical approach is an approach with decision-making authority located at the top level which uses steering and control to spread the data governance program within the organization. The horizontal approach makes use of collaborative behaviour and utilizes formal coordination mechanisms such as working group committees, task forces and integrator roles [3].

3.4.4 Organizational scope

The organizational scope represents the expansiveness of data governance and roughly corresponds to the unit of analysis, which is subdivided into the intra-organizational scope and the inter-organizational scope. Implementing a DGS in an ecosystem of organizations with different stakeholders prioritizes the concepts differently and brings other challenges than implementing a DGS within a single project. Inter-organizational data governance requires better-defined policies and standards which can be challenging to achieve due to organizational differences like data culture.

3.4.5 Data scope

According to Abraham (2019), every data governance program must specify which type of data is in focus. The scope of the data can be classified into two different categories: traditional data and big data. Traditional data comprises master data, transactional data and reference data. Addressing big data in the DGS can require concepts like architecture, storage or Metadata requirements to be better defined than when an organization only processes traditional data. The data scope could also determine what kind of roles and responsibilities are present in the governance structure because of the expertise that is required to work with the different kinds of data.

3.4.6 Domain scope

Data governance programs can address different goals or different data design decisions. It is important to determine the goal of the data governance program and according to this goal determine the focus domains which should explicitly benefit from the governance program. [3] identified the following domains: Data quality, data security, data architecture, data lifecycle, metadata and storage and infrastructure.

3.4.7 Consequences

Consequences of data governance can be categorized as intermediate performance effects, which describe the positive effect of data governance on KPIs for example data utilization, operating capabilities or data quality. Another consequence described by [3] is the management of risks. It is possible that data governance does not have a direct impact on KPIs but has a more preventive role by creating, for example, risk-mitigating policies.

3.5 Data governance in a data mesh

As described in **Section 3.3.2**, defining and describing the DGS in a data mesh is perceived by multiple researchers as a challenge in the migration and implementation of a data mesh. This subsection discusses the results of the systematic literature study presented in **Figure 2.2** and compares these results to other findings. The most important findings during this study are related to the antecedents and the governance mechanisms concepts in the framework of Abraham (2019)[3].

3.5.1 Data mesh antecedents

Antecedents are mostly the implementation of a new organizational approach (data mesh). But in the articles, the motivation for the implementation of a data mesh (and therefore

the antecedents for the new governance structure) can be very important and interesting to take into account. The literature showed that there is not a single Antecedent significantly more represented in the DGSs. This is totally case-dependent. All the advantages of a data mesh could be the motivation for the implementation of a data mesh and therefore for the implementation of the new DGS. Even other antecedents like the perceived future value or the regulatory obligations could be antecedents.

An interesting finding is that the internal concepts that lead to the implementation of data governance are interrelated. Limitations of the traditional data architecture could result in unaligned business and IT, this could require a change in the IT and organizational strategy, which subsequently recommends a change in IT and organizational architecture [59, 37, 58, 21, 34, 43].

The exact motivations for and goals of data mesh migration differ per organization. According to Abraham (2019), these differences can impact the governance structure, however, these differences only influence the specific implementation of the DGS and not the conceptual structure itself. For example: if an organization is adopting a new DGS to improve data privacy within the company, the conceptual governance structure is the same, but the access management roles and responsibilities can be relatively more extensive. This is supported by Joshi (2021), which explains that the implementation incentives mainly influence the choice of tools and the choice of standards while the conceptual structure remains the same. Therefore, to support organizations in defining and describing their DGS, it is important to define the structure on a conceptual level while upholding its usability.

3.5.2 Data mesh organizational scope

One obvious finding that emerges from the analysis is that the organizational scope of most DGSs in data meshes is intra-organizational. This is expected as the data mesh is an architecture that covers a single company or a part of a company. Within the intra-organizational scope, there are differences between DGSs for a single project or for the entire organization. The instances which are describing the structure in only one project, mention that this is a test project and that they strive towards a broader implementation within the organization in parts they think a data mesh could be beneficial. Hence, it could conceivably be hypothesised that for defining and describing the DGS within a data mesh, it is important to assess whether the implementation needs a test project, which one this would be, and which parts of the organizations should be included and excluded in the mesh to reach the full potential.

Donvcevic (2022) mentions that organizations should already be aware of the possible fast adoption of data mesh and suggests that an inter-organisational perspective could be used to look at interoperability between different data meshes. Donvcecic (2022) does not further discuss the inter-organizational scope of data governance in a data mesh architecture. Increasing the inter-organizational scope of the data mesh requires scalable data governance. Managing the scalability of data governance within a data mesh to realise inter-organizational growth is a rarely discussed topic in the literature therefore, it is out of the scope of this research. Additionally, Donvecic's paper is excluded from this SLR due to the lack of discussion on data governance and the lack of relevance to the current developments in data mesh migrations.

3.5.3   Data mesh governance mechanisms

Governance mechanisms are really important and it looks like there is the most deviation between organizations and how they defined and described these mechanisms.

An interesting observation regarding the **structural mechanisms** is that there are four different kinds of entities that collaboratively are responsible for data governance in a data mesh. The global governance team, the global infrastructure team, the domain team(s) and the data product team(s).

All of the articles agree that the global governance team is responsible for issuing and defining the organization-wide standards and policies. In addition, the presence of domain representatives in this global team to create the balance between global interoperability and domain autonomy is in line with the responsibilities described by Dehghani [23]. However, the other roles proposed by Dehghani, Data platform representatives, SMEs and facilitators, are not mentioned. It could be possible that the SMEs are included in the roles present in the global governance entity and labelled as "domain representatives", but this is not clear from the context. Only a few articles discuss the global governance team in depth. The articles that did, included performance measuring as the responsibility of the global data team. This is likely to be related to the overarching responsibilities of the global team. How KPIs for the data mesh are defined and who is responsible for this is not discussed, but it is expected to fall within the responsibilities of the global governance team.

The global infrastructure team is mentioned by multiple sources, but these sources do not go into detail on the exact roles and responsibilities within this team. According to Dehghani, this team is responsible for the implementation of data governance in the infrastructure. Most of the current implementations are not yet mature enough to implement computational governance and therefore this is not represented by many articles in the current responsibilities, however, it is expected that in the desired governance structure the infrastructure team is responsible for the implementation of the computational governance in the infrastructure. De Boer (2022) included the responsibility to enable interoperability of the various components within the Data mesh. Even though this responsibility is rather generic, the global, technical and infrastructure-focused nature of this team suggests that this responsibility is included in the global infrastructure team. A possible explanation for the lack of specific roles within this team could be that the team is made up of many teams with fractal topology [23]. The exact division depends on the complexity, scale and revisions of the platform in an organization and when starting small, this team could consist only of general members. These members should be able to align the infrastructure with the enterprise-wide technology and data strategy and the abstractions that create the self-serve element of the infrastructure should be aligned with the policies and frameworks present in the enterprise. Finally, this team should be in narrow contact with the other teams to set a strategy for and the road map of the data platform.

Several articles agree that the domain team is responsible for fulfilling their domain practices and that their domain decisions are adhering to global standards. The other responsibilities are less agreed upon, but when looking at the context, the responsibility to develop the ubiquitous language of the domain data model and the responsibility to implement and maintain data pipelines within the domain could be best fitting for the domain teams. An interesting finding is that multiple articles do not distinguish between the data domain team and the data product team. This might be because the domain

team consists of multiple data product teams and because Dehghani does not go into much detail about this team. However, The responsibilities of the domain team are expected to be significantly different from the data product teams and therefore should be distinguished. This domain team could consist of a domain owner, domain expert(s), domain data engineers, application developers and data product representatives.

The data product team is one of the most discussed teams and the results of the SLR show that multiple articles agree that tracking the performance of, building and maintaining data products is the main responsibility of the data product team. This comprises access management, compliance management with local and global standards and with service level objectives, data quality management, metadata management, and customer management. During this study, there are only 2 specific roles within this team identified, the data product owner and the data product developer. However, for a data mesh to be successful, this team is expected to consist of more roles. Grey literature articles like the article of Google Cloud architecture Center about data mesh [1], distinguish the data product producer and data product, consumer teams. Surprisingly, this distinction is not found in the literature examined for this study. However, this could be beneficial while defining and describing the DGS because the roles and responsibilities within a producer and a consumer team can be different [1]. A data product team could be a producing team, a consuming team or both. It could be interesting to determine whether this is the case for the different data product teams in an organization to know what the composition of the data product team should look like.

Regarding the **procedural mechanisms**, the types of standards and policies used in a data mesh are not significantly different from standards and policies used in other data architectures. However, their implementation of them is different. For example, due to the federated governance structure and the independence of domains regarding their data models, global modelling language and semantic modelling language standards are in particular very important [34]. There is a difference between the domain models and without these modelling language standards, these differences cannot be (automatically) recognized. Also, because of the need for data products to be findable, accessible and usable, the metadata standards regarding data products in a data mesh can be more strict and inclusive [37]. It is case dependent on how strict or flexible these standards should be and whether living up to these standards is automated and therefore the responsibility of the architecture team or not automated and the responsibility of the data product owner. These decisions should be made within the global governance team.

The **relational mechanisms** are mostly more discussed on a technical level by defining the ways stakeholders communicate via the self-serve infrastructure. This infrastructure facilitates collaboration between stakeholders and guides teams towards their goals. Interestingly, training is rarely discussed. This wasn't expected because this implementation requires people to gain capabilities that they currently do not have. Even though intuitively said, this is probably achieved by training, this is not clearly mentioned in the literature.

3.5.4 Data mesh data scope

One unanticipated finding was that the articles do not discuss the influence of data governance on the data itself. A possible explanation for this might be that the data barely changes in the migration towards a data mesh. The governance structure defined

by the different teams should be fitted to the type of data present and desired in the organization. The articles do not discuss in detail the influence that governance has on the types of data besides mentioning them. The data is mainly important during the process of data domain specification [21]. Despite the absence of clarity on which data is governed in the cases, it is expected to be important while defining and describing the structure of the data governance because this can change accordingly. For example, the need for capabilities of the team and the type of standards can be different when using only traditional data instead of both.

3.5.5 Data mesh domain scope

Most of the implementations of a data mesh and therefore the implementation of a new DGS is motivated by the need to improve data quality.

Data quality inside of a data mesh is one of the domains which is highly influenced by the new DGS. Due to the organizational architecture of a data mesh and the new division of ownership, the roles and responsibilities regarding data quality change. New roles have been introduced that divide the responsibility and ownership of the data and its quality. These responsibilities are mostly found in the role of a data product owner (DPO) and the global governance team. The DPO has more knowledge of the meaning of the data, so this person can better assess the required data quality and align this with the needs of the data consumer.

Metadata is really important in a data mesh. Especially with the self-serve data platform and the computational governance that are highly reliant on correct metadata management. It is likely that the global governance entity defines the required minimal metadata for data products, and for some data products, this metadata should be extended according to global or local standards.

The **data architecture** is vastly affected by the implementation of a data mesh. Domain models are flexible in relation to different domains, so the global governance that should ensure consistency and interoperability between domains should be able to support this flexibility. How this is implemented and how exactly global governance can deal with local inconsistencies is not yet discussed in detail in the included literature articles. One of the promising aspects that could help with this is by improving the data object definitions. This is done through standardizing modelling languages, data formats and syntax schemes, and by creating semantically precise domain models which are interoperable and detect the inconsistencies [37]. Defining the semantics of data models can be a time-consuming task and requires both business knowledge and software development skills. Even though this could be a solution, the resources it takes to semantically define the data and the domain models until the point where it can have a flexible domain model challenges the relative value of this solution.

3.6 Data governance design

In this section, the results of a literature review of the data governance design process are presented. The design of data governance operating frameworks (DGOFs) and the design of data governance operating models (DGOMs) is discussed. This contributed to answering RQ3a1

For data governance design, it is important to distinguish two distinct components of the

design of a data governance program. The design of the data governance scope and the design of the data governance mechanisms that enable the design of the data governance initiatives regarding scope [3]. This study focuses on the latter. Data governance mechanisms can be complex and should be designed, implemented and maintained with specific aspects of the organization in mind. Ladley describes the design, implementation and maintenance processes of these data governance programs extensively in his book "Data Governance: How to Design, Deploy and Sustain an effective data governance program". Here is explained that this consists of an iterative process of 5 areas: engagement, strategy, architecture & design, implementation, operation and change [51].

The data mesh migration challenges found in the scoping study described in **Section 3.3.2**, i.e. defining and describing the DGS, mostly address the architecture and design area of this process. During this process, the operating framework that describes the various arrangements and interactions of the organizational elements that will operate data governance is designed [51]. This means describing roles, responsibilities, interactions and core processes [51]. This corresponds to the design of the data governance mechanisms described by Abraham (2019) [3].

3.6.1 Data governance operating framework design

The development of the operating framework(s) is executed after the high-level data governance strategy is created and after the capabilities needed to execute data governance are identified. Ladley divides the process of developing the operating framework into four activities which collaboratively help with detailing how the data governance will be executed [51].

In the first activity, data governance processes are identified or redefined by creating a generic list of processes, rationalizing these processes to make sure that the policies in place do not conflict with the new data governance processes, addressing regulatory items and considering process design for key functions. The second activity is identifying accountability, responsibility and ownership. Here, the generic role and responsibility division is created together with the various layers of authority within the operational data governance. These layers present the accountabilities and the responsibilities, and the roles within these layers can be used to create a RACI table [40]. With the RACI matrix, the various layers of authority, and the processes in place, the functional framework can be refined. The layers of the operating framework come from the levels of authority, the interchange between the layers comes from the workflows required to support processes of the list of processes, and the federation of the governance results from the RACI matrix [51].

The data governance operating framework is designed by identifying the data governance processes, identifying roles, responsibilities and accountabilities based on these processes and based on this, creating an organizational structure that describes the interaction between these components [51]. The goal of a data governance operating framework is to design the high-level organization structure in which the data governance operations are executed and specific operating models can occur [51]. Making decisions on the design of a new operating framework can be difficult, especially when a strategic decision is made to change the organizational structure and the processes on which the traditional data governance operating framework is built [51]. The redesign of the operating frameworks requires a good understanding of the current organization's operating framework, the

organizational strategy, the principles of Data governance operating frameworks and the interaction between these components.

3.6.2 Data governance operating models

An operating model (OM) is a representation of how a specific organisation's components are configured, and function together to execute on strategy [50]. An OM is a model for an organisation to align itself with the overall business strategy and goals. When executed well, an enterprise operating model can improve resource utilization and efficiency, reinforce an outcome-based culture and mindset, and aid communication, collaboration, and knowledge-sharing [49]. Even though the name might suggest that an OM is a single all-encompassing and integrating model, in reality, an operating model is likely to be a collection of different models, visualizations, explanations and process models that are interlined to describe the operating of a corporate function [33].

In the case of this research, this corporate function is the execution of data governance-related practices. The operating model for data governance describes in concrete terms how data governance delivers the added value promised to the company and which processes are necessary to achieve this. The DGOM does this through the visualization and description of the procedures of the data governance enterprise functions [12].

The design and development of DGOMs for an organization is a process which is also explained by [51]. In this publication, DGOM design is part of the operating framework section in the Architecture and Design work area of data governance program design how this process relates to the rest of data governance program design is visualized in **Figure 3.5**.
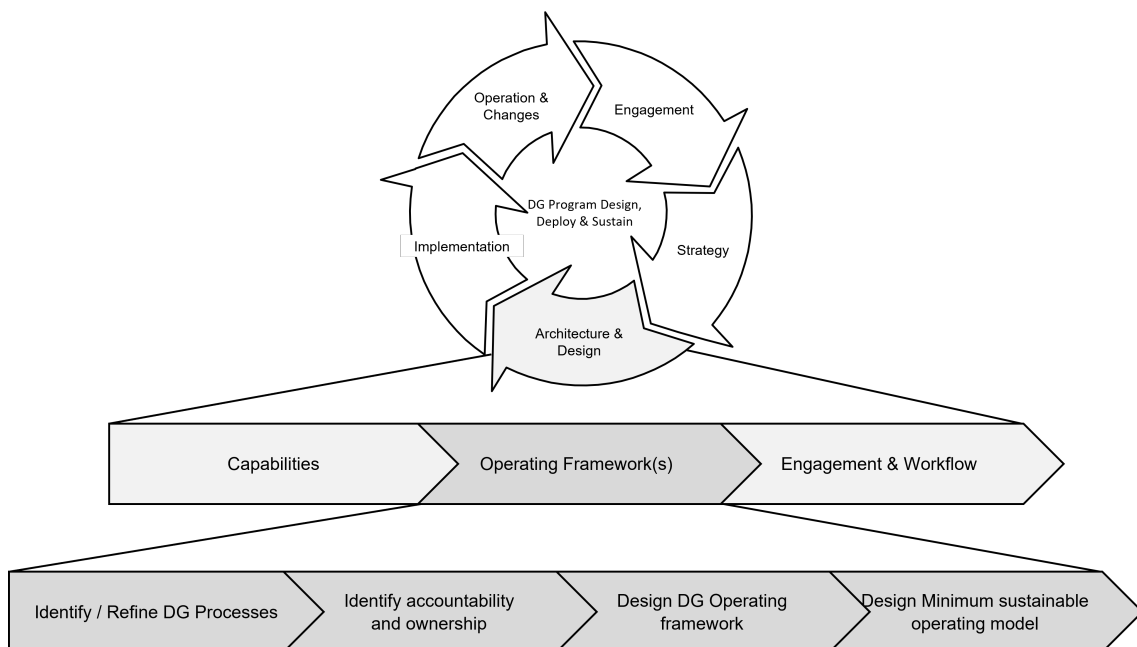
FIGURE 3.5: A visualization of the placement of DGOM design process according to [51]

3.6.3 Available guidance for data governance design

Data mesh evolution model

Dehghani (2019) describes some examples of different configurations of the DGSs in a data mesh. It presents phases of evolution that an organization has to go through to fully migrate to a data mesh architecture. These phases are the explore, expand and extract phases. For the data governance, these phases discuss the evolution of governance characteristics such as the number of domains joining the federated governance operation, maturity of the federated operating model, the focus of governance development, and coverage of computational policies [23]. This explains what characteristics should evolve, but not how these characteristics could evolve. "The members of the existing governance group are either taking a subject matter expert role, joining the platform team to help with product management of cross-functional policy automation, or joining the domains as data product owners" [23].

Data Management Body of Knowledge

Another resource that is frequently used for the design of DGSs is "DAMA DMBoK - Data Management Body of Knowledge" [39]. This book presents the following consideration areas for the construction of operating models for data governance:

- Value of data to the organization: For organizations where data is an operational lubricant, the form of data governance will be less intense than for an organization that uses data to make a profit.

- Business model: Decentralized business vs centralized, local vs international, etc. are factors that influence how business occurs, and therefore, how the data governance operating model is defined.

- IT strategy, Data Architecture, and application integration functions should be reflected in the target operating framework design.

- Cultural factors: Such as acceptance of discipline and adaptability to change. Some organizations will resist the imposition of governance by policy and principle. Governance strategy will need to advocate for an operating model that fits with organizational culture, while still progressing with change.

- Impact of regulation: Highly regulated organizations will have a different mindset and operating model of data governance than those less regulated. There may be links to the Risk Management group or Legal as well.

- The operating model also defines the interaction between the governance organization and the people responsible for data management projects or initiatives, the engagement of change management activities to introduce this new program, and the model for issue management resolution pathways through governance.

The DAMA-DMBOK also provides descriptions of the data governance roles and responsibilities. However these roles are based on a single data governance team and do not take the domain-driven design- and the product thinking principles into account.

These resources lack usability to improve guidance in the organization design for the DGS as practitioners use these resources and experience a lack of guidance during the design process. According to the problem investigation interviews, the organizational structure

of a data mesh lacks practical implication and DAMA-DMBOK lacks applicability in a data mesh context. Therefore, there is a need to develop a model or tool that provides guidance during the design process.

3.6.4 Decision trees for organization design

As described in the sections **Section 3.6.1** & **Section 3.6.2**, in the design of data governance operating models and operating framework, the design of the organization is important. The key question for enterprise architects concerned with organization design is what are the different structures available to choose from [27]. Contingency theories of organization design and data governance design have shown that there is not one best DGS for all organizations and that the best structure is really organization-specific [78]. Duncan (1979) showed that decision trees can help with designing the organizational structures for specific organizations [27]. To create a decision tree, the influencing factors of the environment of the organization, the existing organization structure and the different kinds of organization structures should be known.

A correct organization design decision tree provides a broad framework for identifying the key factors an enterprise architect should think about in considering an organization design, it forces the architect to diagnose the decision environment, and it causes the architect to think about how much interdependence there is among the segments of the organization and points out what can be done to meet the increased need for information [27]. The standard decision tree helps with choosing an alternative. However, due to the contingency approach taken during this study towards data governance and organizational design and because of the unlimited amount of influential factors on the organization structure, creating a decision tree that chooses the best alternative for the organization is not feasible for this research. In addition, the most challenging during the design decisions regarding the DGS is related to the gathering of information, generating alternatives and evaluating alternatives. Therefore a different model that improves guidance is needed.

A type of decision tree that can be used in organization design is a Design Option Tree [8]. This is a representation of the sequence of engineering decisions required for the resolution of a design problem and described the design options available at each decision point [8]. This is expected to improve the guidance during the design decision-making process and with that the quality of the decisions.

3.7 Theoretical background summary

The findings presented in this chapter are expected to contribute to the growing area of data mesh research by conducting a literature study on data mesh, a scoping study on data mesh challenges and systematic literature on data governance components in a data mesh. Additionally, the introduced theoretical concepts, relations between theoretical concepts, and results from literature studies provide a theoretical foundation for the rest of this research to build upon. Furthermore, the investigation of the literature contributes to the understanding of the problem context which is further elaborated in the next chapter.

Chapter 4

Problem investigation

This chapter played a crucial role in the research by exploring the current data governance design process, which served as a foundation for addressing RQ3 and RQ3c. This contribution is achieved through several key elements. Firstly, it restates the problems identified in the literature review, emphasizing their significance (**Section 4.1**). Secondly, it defines the social context surrounding the problem addressed in the study, providing valuable insights (**Section 4.2**). Additionally, it introduces two organizations utilized as practical case studies, offering relevant real-world perspectives (**Section 4.4**). Moreover, it introduces the Enterprise Data Management (EDM) team of Deloitte and how this team is affected by the problems (**Section 4.3**). Lastly, in the chapter the conceptual problem framework developed specifically for this study is presented, solidifying the research's direction.

4.1 Problem according to literature

The results of the preliminary scoping study presented in **Section 3.3.2** pointed out that designing the DGS is one of the main challenges of migrating towards a data mesh architecture. The findings of Bode (2023) state that current guidelines do not assist the design of data governance for data mesh architectures [11]. In the remainder of this chapter, the influences on the DGS design process are investigated by an investigation of the available guidance and an analysis of the key stakeholders in the DGS design process.

4.2 Stakeholders and goals

There are multiple actors involved in or affected by the process of designing the DGSs for data mesh architectures. These stakeholders are internal enterprise architects, internal actors with data governance responsibilities, internal actors that are financially responsible for the data mesh migration and external EA consultants that support organizations in designing the DGSs.

The internal enterprise architects are responsible for the design of the DGS in the organization. They communicate with actors from the organization to gather information about topics that influence the DGS like its strategy, culture and current DGS [51]. The goal of internal enterprise architects is to efficiently design effective DGSs for the organization (SG1). The external EA consultants support the internal EA

consultants with the design of the DGS or, if there are no internal enterprise architects, take over the responsibilities of internal enterprise architects. These roles are different as the external EA consultants act from outside the organization and support multiple data mesh migrations in different organizations. Because of this, for external consultants, it is important that the provided guidance is applicable in different organizations (SG2).

The management of the organization and other actors participating in the data governance design process contributes to the process by providing resources in the form of time and knowledge. Who these actors are depends on the organization, but examples of these roles are the Chief Information Officier (CIO), Chief Financial Officer (CFO), Subject matter experts (SMEs) data governance lead, human resource lead, members of the data governance teams, team leads from business units and sponsors [51]. Their goal is to provide the resources efficiently and to design a DGS that fits the organization (SG3). Additionally, the CFO is responsible for the financial resource delegation of the data governance design process. Therefore, this role aims to reduce the resource costs of this process (SG4).

As described, there are different kinds of stakeholders affected by the design process and therefore affected by the guidance during this process. Alexander (2015) proposes a taxonomy for categorizing these stakeholders. This taxonomy distinguishes stakeholders that are interacting with the artifact, in the immediate environment of the artifact, in the wider environment of the artifact and involved in the development of the artifact [82].

The stakeholders interacting with the artifact are the internal enterprise architects and the external EA consultants. According to the taxonomy of Alexander (2005), these stakeholders are the normal operators. The stakeholders in the immediate environment of the artifact are the actors participating in the data governance design process. According to the taxonomy of Alexander (2005), these actors are the functional beneficiaries of the artifact as these actors benefit from the output produced by the system. The stakeholders in the wider environment of the artifact are the CFO as he benefits from the system financially.

The stakeholders involved in the development of the artifact are the SMEs and case representatives supporting the development of the artifact (Consultants [5]) and the researcher that develops the model (Developer [5]). Furthermore, the Enterprise Data Management (EDM) team of Deloitte, which is introduced in the next section, and the University of Twente both provide a budget for developing the artifact and therefore are categorised as sponsors of the project according to Alexander (2005).

The stakeholders, their types according to Alexander's taxonomy and their goals are presented in **Table 4.1**. A visual representation of the social context is presented in **Figure 4.1** in an onion diagram as proposed by Alexander(2005). In the next section, the position of the EDM team of Deloitte is further elaborated.
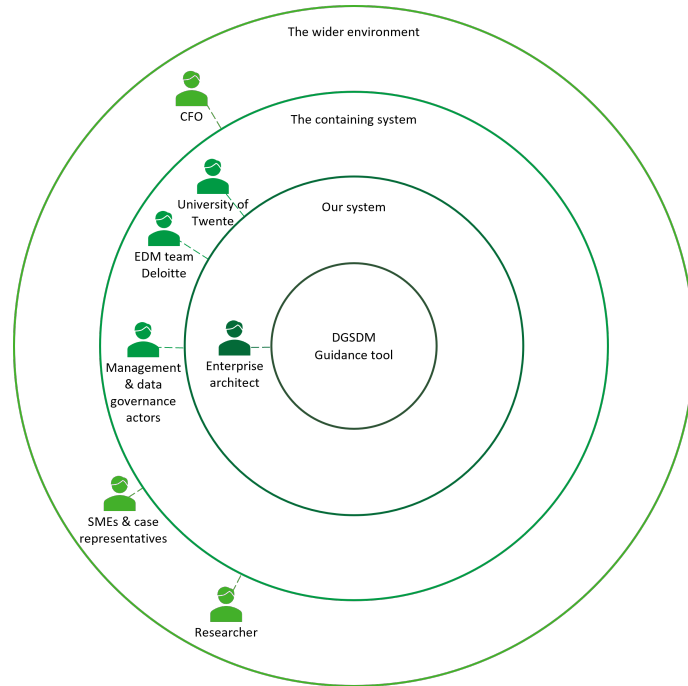
FIGURE 4.1: Onion diagram of stakeholders based on the taxonomy of Alexander (2005)

TABLE 4.1: Stakeholders of the research, their types according to the taxonomy of Alexander (2005) [5], their goals and the used id as reference for the specific stakeholders' goals

Goal ID	Stakeholder	Type	Stakeholder's Goal
SG1	Internal enterprise architects	Normal operator	Efficiently design effective data governance structures for the organization
SG2	External EA consultants	Normal operator	Efficiently design effective data governance structures for different organizations
SG3	Management and other actors participating in the data governance design process	Functional beneficiaries	Provide resources efficiently and to design a data governance structure that fits the organization
SG4	Chief Financial Officer	Financial beneficiary	Reduce costs of effective organization-specific data governance structure design process
SG5	Deloitte EDM Team	Sponsor	Develop a guidance tool that reduces costs of the data structure design process performed by EA consultants
SG6	University of Twente	Supplier of knowledge	Contribute to research and practice
SG7	Subject matter experts	Consultants	Share knowledge within the domain or organization
SG8	Case representatives	Consultants	Share knowledge within the domain or organization
SG9	Researcher	Developer	Develop an artifact that fulfills the goals and meets the requirements of stakeholders

4.3 Deloitte Enterprise Data Management team

The Deloitte enterprise data management team is a team of consultants who provide management consulting practices in enterprise architecture and data management. Consultants from this team advise their clients on how to design and manage the IT infrastructure and how to align it with the business processes, technology, data and people within the organization. When the strategy of the client includes the migration towards a data mesh architecture, the EDM team supports the organization with this migration. This consists of multiple phases, one of which includes making design decisions regarding the DGS. For this, the consultants from the EDM team conduct interviews with data governance actors and C-level executives to assess the current data governance structure, the goals of the organization, and its strategy and use the information from these interviews to create an as-is architecture.

This architecture and the goals of the company are used to create a target architecture that fits the organization and contributes to reaching its goals. This target architecture is discussed with the stakeholders at the organization for refinement and to define a roadmap to fill the gaps between the as-is architecture and the to-be architecture. Based on this refinement process, the target architecture and roadmap can be adjusted. This is an iterative process with continuous from the client. A simple visualization of the process is presented in **Figure 4.2**. The activities with a red outline are the activities where the challenges identified in the problem investigation interviews are most prominent. The subsequent section delves into an empirical analysis of this data governance design process and the problems that occurred during this process in two cases.
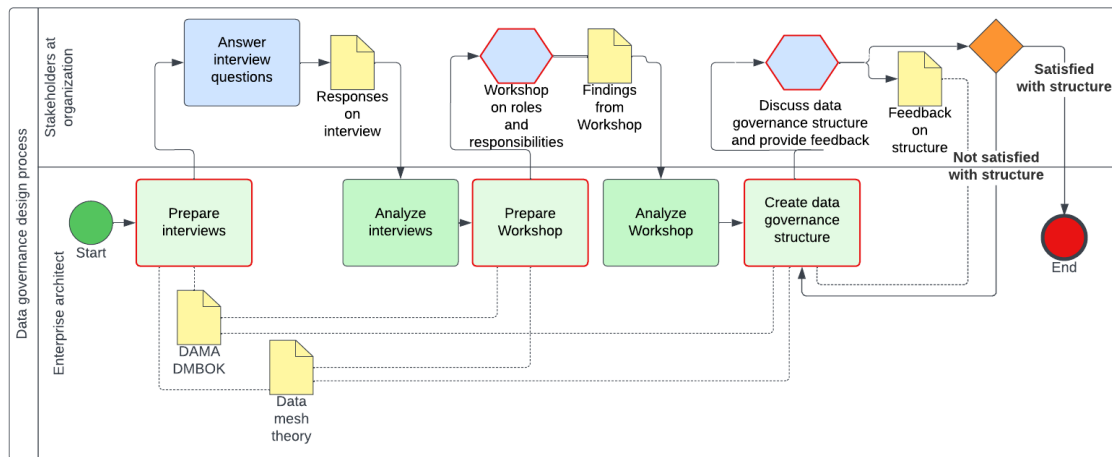
FIGURE 4.2: Visualization of the DGS design process

4.4 Case studies

To further examine the problem, two cases are investigated. This investigation explored how potential issues influenced the design process of the DGS and follows the methodology presented in **Section 2.4.2**. The results are presented on a case-by-case basis, beginning with an introduction to each organization and outlining their motivations for migrating to a data mesh architecture. Subsequently, the DGS design process is discussed including an analysis of the problems that arose during this process. For the Energy 1 case, initial problem investigation interviews are conducted but most of the findings originate from the treatment design interviews presented in **Appendix F**.

4.4.1 Energy 1

Energy 1 is a global company that operates in more than 100 countries with a vision to implement the principles of a data mesh architecture in their operating model and as their way of working with data (TDIE2). According to the problem investigation interviews, before the implementation of the data mesh principles, the organization was dependent on a very centralized DGS as Energy Consultant 2 *"Energy 1 has a long history with established centralized organization structures and processes"* (TDIE2). Energy consultant 2 explains the culture as follows: *"It's a very established company which is not willing to change much. So changes really takes time"* (TDIE2). The organization was decently mature regarding data governance as it had a centralized data team with people responsible

for data governance activities (TDIE1).

Motivation for data mesh

The goal of the migration is to improve data quality and increase the value realized through data sharing and data usage. The organization already put effort into the creation of data products and had an infrastructure to share these products on, so the gap between their current architecture and a data mesh architecture on a technical level seemed relatively small (TDIE1). However, the migration required a significant change in the organization structure in order to further support the implementation of the data mesh principles (TDIE1).

Data governance design process

The process is described as follows: *"It was a top-down approach to introduce a new central data office (CDO) organization and to push from the top down the principles and paradigms of data mesh in all matters, covering new roles that get introduced, new processes, different data mindset in terms of data sharing, and on the other hand, as well, in the technology side, to introduce a new platform"* (TDIE2)

For this transformation, multiple working streams are created which were executed in parallel. One of these working streams was the data excellence work stream in which the data governance roles and organizational structure were defined (TDIE1). This working stream was responsible for the design of the DGS and for aligning this with the current organizational structure. Energy consultant 1 states the following: *"In order to come up with certain roles definitions [...] you need to define certain processes because then you can identify what kind of responsibilities do you actually need in order to complete a process"* (TDIE1).

The identification and definition of these processes were accomplished through focus groups and workshops involving enterprise architects and data governance actors from the organization.

The second process was the data product creation process (TDIE1). *"This was a key process we were working on as part of the data governance team coming up with a data product creation process, meaning how do we get order requirements and who should be involved in requirements gathering"* (TDIE1)

The DAMA-DMBOK reference model was used for this phase and was considered the most valuable resource (TDIE1). The defined roles and responsibilities are compared to the theoretical data mesh to design a DGS that supports the data mesh principles and fits the structure of the organization.

Design process challenges

During the design of the DGS for Energy 1, the EA consultants ran into multiple problems. One of which was that the available resources were not giving enough guidance during the design process (TDIE1 & TDIE2). The practical resources did not incorporate the data mesh principles (TDIE1). Therefore, the roles, responsibilities and team structures presented in these resources were not directly applicable to the design of a DGS for a data mesh architecture (TDIE1). Moreover, the theoretical DGS suggested by Dehghani (2019) was not directly applicable to the existing organization due to significant disparities

in roles, responsibilities, and team structures between the current centralized architecture and the proposed federated architecture (TDIE1). Implementing the theoretical structure directly would entail substantial risks and was perceived as unfeasible during the migration process (TDIE1 & TDIE2).

Because of the lack of direct applicability of both the theoretical and the practical reference structures, the reference models and structures required significant adaptations to the roles, responsibilities and team structures presented. *"With regards to the data governance structure, there hasn't really been a fitting reference model. So what we've used is a collection of best practices for example, so the DAMA DMBOK and tailored these to the data mesh principles"* (TDIE1).

Furthermore, discussions around the design of the governance structure were often lengthy and unstructured (TDIE1). According to Energy Consultant 1, it was difficult to structure the meetings because they did not have identified important design decisions to discuss with the stakeholders (TEIE1). Because of this, the meetings consisted of a lot of brainstorming about how the organization could be configured and formulating and discussing the potential advantages and disadvantages of these different configurations (TEIE1). Furthermore, the absence of a structured decision-making process meant that important design choices were often delayed, causing delays in the overall implementation timeline (TDIE1).

4.4.2 Telecom 1

Telecom 1 is a global leader in the telecommunications industry, specializing in the development and manufacturing of advanced technologies for wireless communication, security, and digital networking. The company is headquartered in Europe and has operations in Asia, the Americas, and other regions with a workforce of over 50,000 employees worldwide.

Motivation for data mesh

The organization is motivated to adopt a data mesh approach because it is a distributed data governance approach that provides a solution to the siloed data structures currently within the organization (PIT1). *"They were looking for a solution to enable this data sharing and organization and quite early on, the choice was made to adopt a data mesh approach, also in designing the operating model for the future organization"* (PIT1). The data mesh approach promotes collaboration among distributed teams by providing a framework for data sharing and governance. The distributed DGS also ensures that responsibilities are distributed among teams and that there is a central data office (CDO) that coordinates data domains within the organization. This approach ensures that the organization can collaborate effectively and leverage the expertise of all teams to achieve its goals.

Data governance design process

The process of designing the DGS involved several stages that required continuous discussions and workshops. The process started with an assessment of the current operating model and how the client should organize itself to have the roles and responsibilities in place to adopt this data mesh (PIT1). This assessment is used to analyze the operating model and the roles and responsibilities and prepare for the first

workshop with the client. This workshop defined the services that teams would provide for the whole organization or for other teams. The findings from the services workshop were used to define the responsibilities and accountabilities through the creation of a RACI matrix (PIT1). For this, DAMA-DMBOK is used because according to Telecom Consultant 1, *"it is quite extensive and contains a lot of puzzle pieces to create an operating model"*. The RACI matrix was analyzed, and a follow-up workshop was conducted to discuss contradicting responsibilities in the way of working. A DGS was designed and based on the findings of the second workshop. This structure is presented to stakeholders for refinement and to discuss contradicting roles and responsibilities. The proposed DGS is adjusted during a four-month period based on continuous feedback from stakeholders (PIT1).

Design process challenges

This organization also struggled with the lack of applicable reference resources. The resources used were not data mesh-specific models, and the original model assumed a traditional data governance model where all roles and responsibilities are centralized. Telecom Consultant 1 states that they *"didn't find any structured decision-making structure in the context of a data mesh"*(PIT1). This resulted in three main problems: 1) the important decisions were unknown, 2) the possible configurations were unknown and 3) the evaluation methods were unknown (PIT1 & PIT2).

Telecom consultant 1 states that they *"really struggled with are what are the questions that we need to ask and what are the real key decisions in adopting this data mesh?"* (PIT1). Telecom consultant 2 also mentioned that they searched for the real choices that they had to make (PIT2). This resulted in additional discussion preparation time (PIT1).

Furthermore, both Telecom Consultant 1 and Telecom Consultant 2 struggled with the availability of possible configurations as it was unknown what configurations can be chosen (PIT1) and which structure from theory needed to change and which ones can be directly "copied" from theory (TDIT2).

Additionally, Telecom Consultant 2 struggled with how these choices can be made and how these options can be assessed (TDIT2). She adds to this that she lacked a guide or a list of considerations that need to be considered while designing the DGS or while co-designing the structure with the client.

The need to develop these design decisions and possible configurations increased the preparation time for the discussions, resulted in unstructured discussions with the client and required the team to redo the discussion because dependencies between decisions were not accounted for (PIT1).

4.5   Conceptual problem framework

The problems experienced in the design process of Energy 1 and Telecom 1 are analyzed and presented in a problem cluster. The resulting problem cluster (**Figure 1.1**) and its explanation can be found in **Section 1.1**. For convenience, the resulting problems are also presented in **Table 4.2**.

Problems are only included if they are explicitly mentioned during the interviews and occurred during the design process described in **Section 4.3**. There is a possibility that there are more problems that have occurred than are currently presented in the problem

statement. However, according to the interviewees, the most important challenges are discussed during the interviews.

Solving the problems that occurred during the data governance design process would contribute to the goals of the stakeholders in the following ways:

Reducing the preparation time for discussions and the design time of the DGSs would increase the efficiency of the DGS design process and therefore contribute to the goal of enterprise architects to efficiently design effective DGSs for their organizations.

Providing structure in the discussions with the management and other actors participating in the data governance design process could streamline the decision-making process. Furthermore, the presentation of design decisions and configurations decreases the time these stakeholders have to spend formulating design decisions, as well as creating and evaluating configurations. This would contribute to their goal of efficiently providing their resources in the form of knowledge and time. Consequently, the CFO can achieve cost savings by optimizing the data governance design process while still ensuring the effectiveness of the organization-specific structure.

Overall, addressing the problems in the data governance design process contributes to the stakeholders' goals by enabling them to design efficient, effective, and organization-specific DGSs while saving time, effort, and costs in the process.

TABLE 4.2: The identified problems and their occurrence in the investigated cases

Reference	Problem	Energy 1	Telecom 1
P1	Theoretical resources on data mesh are regularly not directly applicable in practice	x	x
P2	Practical resources do not discuss the new proposed data mesh specific principles	x	x
P3	Enterprise architects lack resources on design decisions	x	x
P4	Enterprise architects lack resources on possible configurations	x	x
P5	Time is spent identifying and formulating new design decisions	x	x
P6	Discussions with stakeholders are inefficient	x	x
P7	Time is spent identifying and formulating new configurations	x	x
P8	There is a lack of insight into dependencies between configurations		x
P9	Time is spent formulating evaluation opportunities for configurations	x	x

4.6 Problem investigation summary

The treatment created in this study should address the problems presented in this chapter to contribute to the goals of the stakeholders. This chapter contributed to the formulation of the answers to RQ3a, RQ3b, and RQ4a by exposing the social context of the researched problem, describing the DGS design process, investigating the problems that occurred in two cases and creating a conceptual problem framework. This is used in the next chapter to define the requirements of the treatment and to further develop the treatment.

Chapter 5

Treatment design & development

In this section, the design and development process of the guidance tool is presented. This process contributed to formulating answers to RQ4a, RQ5a, RQ5b RQ5c and RQ5d in multiple sections. First, the requirements of the treatment are specified based on interviews with enterprise architects. Then, a theoretical DGS is constructed and explained in **Section 5.3**. After that, the results of the observational case studies (**Section 5.4.1** & **Section 5.4.2**) are discussed in detail. Then, the additions of the subject matter experts are presented in **Section 5.4.3**. Lastly, the first version of the guidance tool is presented in **Section 5.6**.

5.1 Requirements

The guidance tool must possess specific properties to be able to contribute to the goals previously introduced. These properties are the requirements of the tool. Consequently, the tool is evaluated with respect to its fulfilment of these requirements. Wieringa (2014) distinguishes two types of requirements: functional requirements (the desired functions of an artifact) and non-functional requirements (the requirement that an artifact has a specified property that is not a function) [82]. The functional and non-functional requirements for the treatment developed during this research are derived from stakeholder interviews and presented in **Table 5.1**.

Table 5.1: Requirements with their requirement ID's and their mentions in the interviews

Requirement ID	Requirements	Source				
	Functional Requirements	TDIT1	TDIT2	TDIE1	TDIDG1	TDIDM1
FR1	The tool must present important design decisions regarding data governance structures for data mesh architectures	x	x	x	x	x
FR2	The tool must present resulting configurations for design decisions regarding data governance structures for data mesh architectures	x	x	x	x	x
FR3	The tool must present advantages and disadvantages of the configurations resulting from design decisions			x	x	
FR4	The tool must present a graphical representation of the configurations					x
FR5	The tool must be presentable during discussions with stakeholders	x	x	x	x	x
	Non functional requirements	TDIT1	TDIT2	TDIE1	TDIDG1	TDIDM1
	Correctness					
NRC1	The design decisions and the configurations presented in the tool must be correct in the context of data mesh migrations	x	x	x	x	x
NRC2	The configurations must be correctly represent possible configurations from practice	x	x	x	x	x
NRC3	The advantages and disadvantages must be correctly represent advantages and disadvantages from practice			x	x	
NRC4	The influential factors must be correctly representing influential factors from practice	x		x	x	
	Understandability					
NRUN1	The presented design decisions must be understandable for people without data mesh experience		x			x
NRUN2	The influence of the design decisions on the data governance structure must be understandable	x	x			
NRUN3	The explanation of the configurations must be understandable	x	x	x	x	
NRUN4	The difference between the configurations must be understandable			x	x	
NRUN5	The (dis-)advanatages must be understandable			x	x	
NRUN6	The influential factors must be understandable	x		x	x	
NRUN7	The graphical representations must be understandable					
	Utility					
NRUT1	The tool must be easy to use during the discussions with stakeholders		x			x
NRUT2	The tool must be easy to use during the the design of the data governance structure		x			x
NRUT3	The tool must be usable for the design of the data governance structures for multiple organizations	x	x	x		
NRUT4	The tool must help with distinguishing important design decisions for data governance in a data mesh architecture			x		x
NRUT5	The dis(advantages) must be useful to evaluate configurations			x	x	
NRUT6	The influential factors must be useful to discuss design decisions	x		x	x	
NRUT7	The tool must present configurations that can be used to design data governance structures in practice	x	x			
	Efficiency					
NREI1	The use of the tool during discussions with stakeholders must reduce discussion time	x	x	x	x	
NREI2	The use of the tool during the design process must reduce the time it takes to formulate new design decisions	x	x			
NREI3	The use of the tool during the design process must reduce the time it takes to formulate new configurations	x	x	x		
NREI4	The use of the tool during the design process must reduce the time it takes to formulate evaluation opportunity of configurations	x	x			
	Effectiveness					
NREE1	The use of the tool must not reduce the effectiveness of the designed data governance structure	x	x	x	x	x

These requirements address specific problems in order to contribute to stakeholder goals. The relation between the requirements, problems and stakeholders' goals is represented in **Table 5.2**. As described in **Section 4.2**, there are stakeholders that are only involved with the development. This research chose to focus on the requirements that the resulting treatment should have and not go into detail about the requirements of the development of the treatment. However, the following additional requirements are taken into account: The artifact must be created within the given time for a master thesis research of the University of Twente (AR1), the development of the artifact must contribute to research and practice (AR2), and the artifact must be developed based on knowledge from SMEs and case representatives (AR3).

TABLE 5.2: Requirements mapped to the research problems and stakeholders'goals

Requirement ID	Research problems									Stakeholders' goals				
Functional Requirements	P1	P2	P3	P4	P5	P6	P7	P8	P9	SG1	SG2	SG3	SG4	SG5
FR1			x		x	x				x	x	x		
FR2				x		x	x	x	x	x	x	x		
FR3									x	x	x	x		
FR4						x				x	x	x		
FR5						x				x	x	x		
Non functional requirements														
Correctness														
NRC1	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Understandability														
NRUN1						x				x	x	x		
NRUN2			x		x	x				x	x	x		
NRUN3				x		x	x	x	x	x	x	x		
NRUN4				x	x	x	x	x	x	x	x	x		
NRUN5									x	x	x	x		
NRUN6						x				x	x	x		
NRUN7						x				x	x	x		
Utility														
NRUT1						x				x	x	x		
NRUT2					x		x		x	x	x			
NRUT3					x		x		x		x			
NRUT4				x	x	x				x	x	x		
NRUT5									x	x	x	x		
NRUT6						x				x	x	x		
NRUT7								x		x	x			
Efficiency														
NREI1						x				x	x	x	x	x
NREI2					x					x	x	x	x	x
NREI3								x		x	x	x	x	x
NREI4									x	x	x	x	x	x
Effectiveness														
NREE1		x	x	x	x	x	x	x	x	x	x	x	x	x

5.2 Available treatments

It is important to assess how the available treatments are fulfilling the requirements and how these treatments contribute to the goals of the stakeholders [82]. As described in **Chapter 4**, DAMA-DMBOK [39] and Dehghani (2019)[23] are the resources mostly used in the DGS design processes in the investigated cases and by the data governance and data mesh experts interviewed during this study. An introduction to these resources is presented in **Section 3.6.3**.

The findings from the problem investigation point out that there are problems that occur when using these treatments. The problem investigation interviews (**Appendix D**) and the treatment design interviews (**Appendix F**) point out that it is needed to develop a new artifact that fulfils the requirements that are not fulfilled by the currently available treatments. The remainder of this chapter is dedicated to the development of this artifact.

5.3   Theoretical DGS development

The methodology introduced in **Section 2.4.3** consists of the creation of a theoretical data governance structure, practical DGSs, additions from SMEs and the creation of the first version of the guidance tool.

The choice is made not to include the findings from the literature study on data governance components because there is not enough information available in these articles to create an in-depth analysis of the DGS and because these findings include practical implementations which should not be included in the theoretical organization structure.

The DGS relates to the structural mechanisms of the data governance framework of Abraham (2019) therefore, the analysis of the theoretical structure consists of the overview of roles and responsibilities and a description of the allocation of decision-making authority. The resulting theoretical structural mechanisms are described in the following section.

5.3.1   Teams, roles and responsibilities

This section described the findings from analysing the data mesh concept proposed by Dehghani. An observation made during the analysis of the book is that a clear theoretical definition of the roles, teams or responsibilities per role is absent. Therefore, the statements implying responsibility for a specific role or team are used to create team-specific overviews.

Dehghani divides the roles and responsibilities within a data mesh into three types of teams: the federated governance team, the data platform teams and multiple data product teams. "Data mesh's global governance team is accountable for defining and designing the policies and standards, the execution of these policies is left to the platform and the responsibility of applying these policies and embedding them in the data products is given to the domains" [24]. This section explains each team in detail with their team responsibilities, specific roles and their additional characteristics.

Federated governance team

Dehghani uses the team typologies [70] to explain the federated governance team as follows: The federated governance team is a team that "facilitates the decision-making around global policies"[23] and enables data product teams by establishing practices, providing informed guidelines, develop their capabilities and close the knowledge gap.

To execute these responsibilities, Dehghani proposed a team that is composed of multiple roles: data product owners, data platform representatives (data platform owner and data platform architect), subject matter experts (legal, security and compliance) and a manager or facilitator. The roles with their responsibilities are presented in **Figure 5.1**.
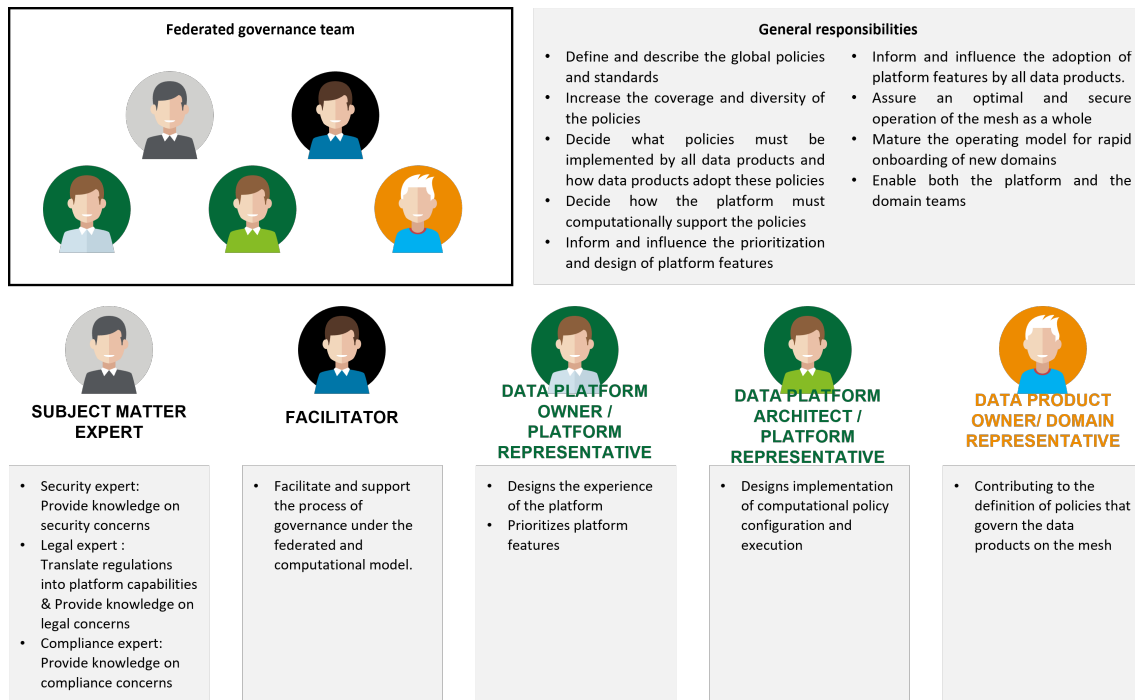
FIGURE 5.1: The roles and responsibilities of the federated governance team according to Dehghani (2019)

Besides the responsibilities of the roles and team, Dehghani mentions two other characteristics of the federated governance team. Dehghani mentions "data governance teams" (plural), however, further explanation of an organization structure with multiple federated governance teams is not given. This is interesting as later in this research (**Section 6.1.2**), the amount of governance councils are mentioned during the case interviews and further explained. The second additional characteristic explicitly mentioned by Dehghani is that "at times, the federated governance team works in strong collaboration with the platform team" to automate governance decisions in the platform" [23]. Exactly what this collaboration looks like and who is involved in this collaboration is not further described.

Data platform team

Dehghani classifies this team as a platform team according to the team typologies [70]. This team enables the data product teams to deliver their work with substantial autonomy [23].

"The data platform team is a logical team made up of many teams with a fractal topology that can have multiple stream-aligned teams, each focusing on an end-to-end delivery of a particular self-serve platform capability" [23]. How exactly a platform team divides itself into one or multiple teams depends on the scale, complexity, and revisions of the platform in an organization [23]. The explanation of the responsibilities of the teams and possible other configurations of the team by Dehghani (2019) is limited. The overview of the extracted roles and responsibilities is presented in **Figure 5.2**
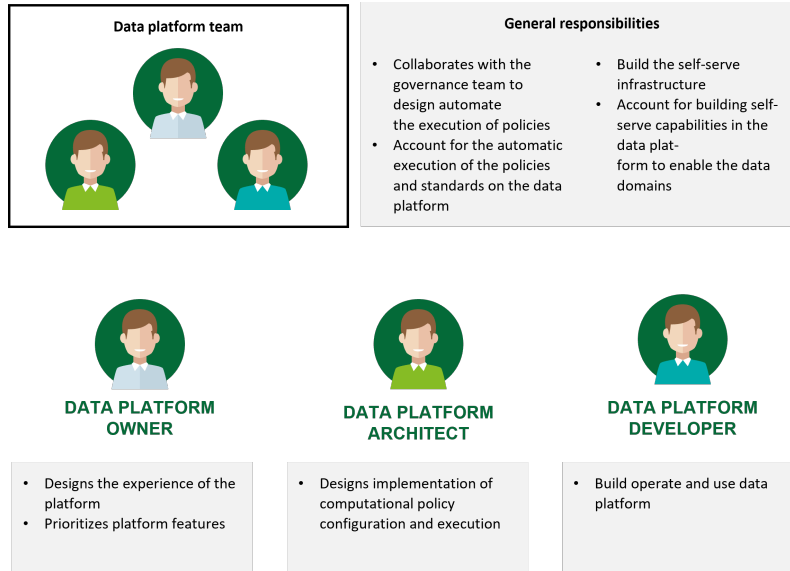
FIGURE 5.2: The roles and responsibilities of the data platform team according to Dehghani (2019)

Data product teams

The data product teams are teams that "focus on an end-to-end continuous flow of designing, modelling, cleansing, building, testing, serving, monitoring, and evolving domain-oriented data"[23]. The structure of the data product teams varies depending on the complexity of the domain and the number of data products. Each product team has their own delivery cycle for their data products. A data product team is likely to be paired with an application development team. Dehghani (2019) distinguishes two different data product teams: data product-producing teams (which are responsible for the creation, development and publishing of the data products) and data product-consuming teams (which use data products to gain insights and support decision-making in the organization). Conceptually, these are two different functions. However, these functions often come together in a single team as a data product team consumes data products for business insight and produces derived data products for the use of other domains [23]. The specific roles and responsibilities of a data product team are presented in **Figure 5.3**.

These teams cannot function on their own and are part of the DGS. In this structure, the interaction between the teams and the allocation of the roles are visualized. The complete theoretical team structure with its roles and the interactions between the teams and the roles are visualized in an organizational viewpoint presented in **Figure 5.4**.

5.3.2 Allocation of decision-making authority

The location of decision-making authority determines which organizational unit has the mandate for action related to data governance (Khatri & Brown, 2010; Otto, 2011b). Abraham (2019) distinguishes between hierarchical positioning, functional positioning, and positioning of decision-making authority on a continuum ranging from centralized to decentralized.

According to Dehghani, most data mesh decision-making processes are horizontal and the
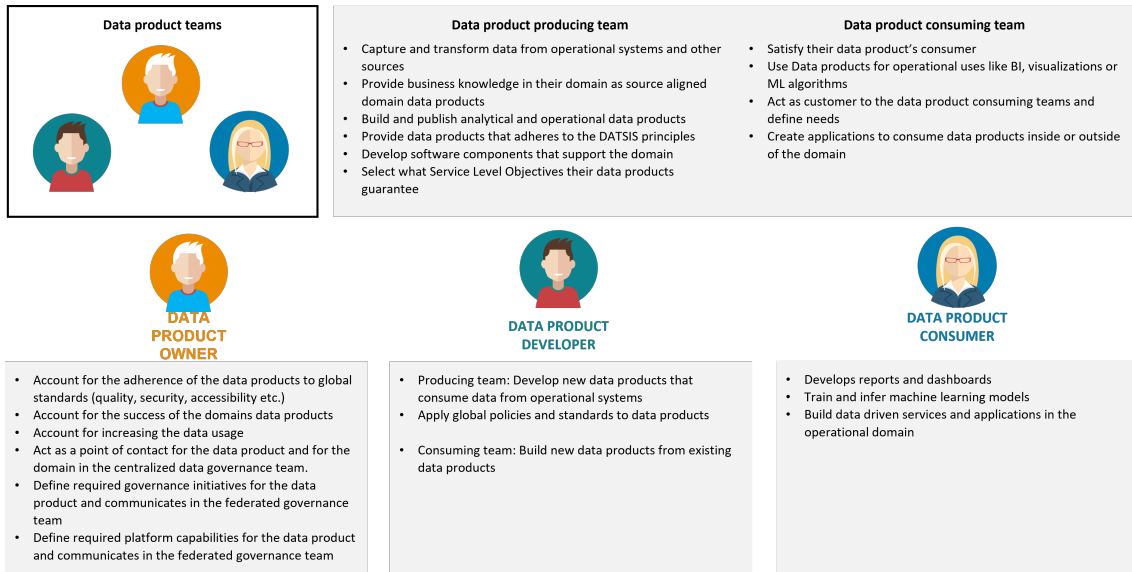
FIGURE 5.3: The roles and responsibilities of the data product team according to Dehghani (2019)

teams are "flat or hierarchical"[23]. Further elaboration on the hierarchical positioning of decision-making is limited. Functional decision-making is about which department holds the decision-making authority. In the theoretical structure, the decision-making of the data governance initiatives does not lie within a single functional department. Each functional domain can make decisions on local governance initiatives and the cross-functional decision-making is done in the federated governance team. In the federated governance team, all functional departments affected by the decision can influence the decision assuming that a representative of each functional department is located in the federated governance team.

The second type of positioning according to Abraham (2019) is the positioning of the decision-making authority which represents the level of centralization and decentralization of decision-making and is one of the focus points of a data mesh [23]. The aim of this positioning is that only decisions on governance initiatives that influence or can influence multiple domains are made in a centralized way. The members of the federated governance team collaboratively make these decisions. As Dehghani explains, these decisions should be limited to prevent the federated governance team from acting as a bottleneck but enough to ensure data quality and interoperability [23]. Decision-making on local data governance initiatives is decentralized to the domains. The way decisions are made on which and how data governance initiatives are implemented by the platform team is centralized in the federated governance team. The resulting accountabilities of the effect of these decisions are located at the positions where these decisions are made. Dehghani states that the level of the federation (centralization vs decentralization) depends on the organization and that it takes continuous feedback to find the right level for the organization [23].

The findings from this section regarding the roles responsibilities and allocation of decision-making present a conceptual model for the theoretical DGS. In the next section, this model is used as a framework to extract practical DGSs and tested on its applicability.
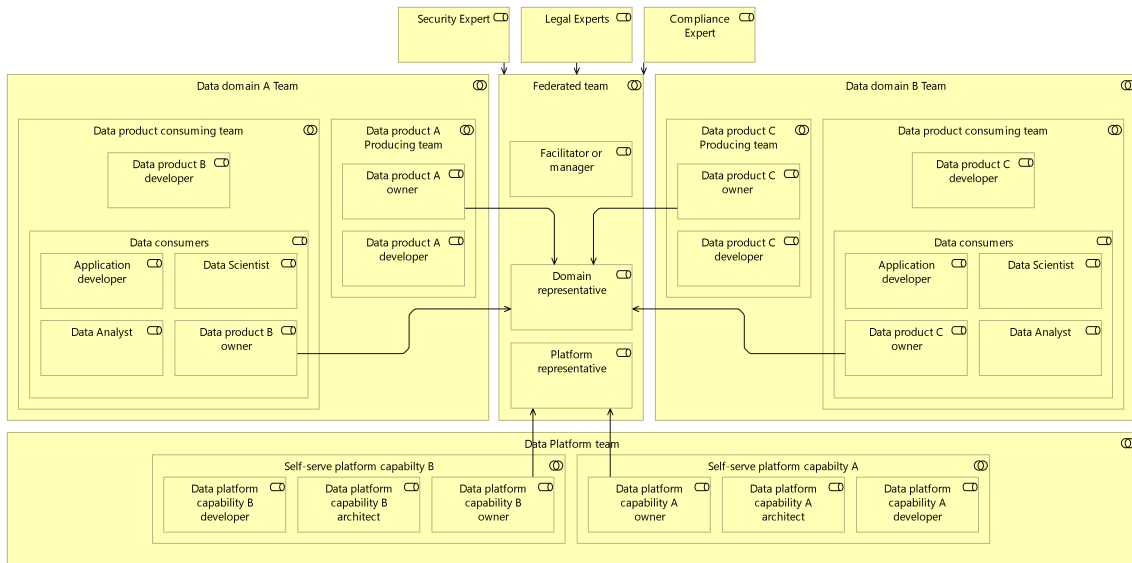
FIGURE 5.4: Organizational viewpoint of the team structure and roles in the theoretical organization structure

5.4 Practical DGSs development

For the creation of the practical structure, the methodology presented in **Section 2.4.3** is followed. For each of the cases, the organizational structure is designed according to the framework presented in the previous section and the team structures with their roles and responsibilities are defined. The resources used for this are interviews with case representatives and additional resources provided by the case representatives about the operating models and data governance roles. The interviews are referenced through the following abbreviations.

- TDIE1 = Treatment Design interview Energy 1 with representative 1

- TDIE2 = Treatment Design interview Energy 1 with representative 2

- TDIT1 = Treatment Design interview Telecom 1 with representative 1

- TDIT2 = Treatment Design interview Telecom 1 with representative 2

5.4.1 Energy 1 case

A detailed introduction to Energy 1 and the data governance design process is given in **Section 4.4.1**. This section presents the results from two interviews with case representatives (TDIE1 & TDIE2) which can be found in the **Appendix F**.

Energy 1 organizational structure

The theoretical structure presented in **Figure 5.4** is discussed during the interviews and this structure is used as a reference to create the organizational structure of the energy company presented in **Figure 5.5**. This structure consists of the central data office, the domain teams, the platform team and multiple councils in which people from the teams come together.

61
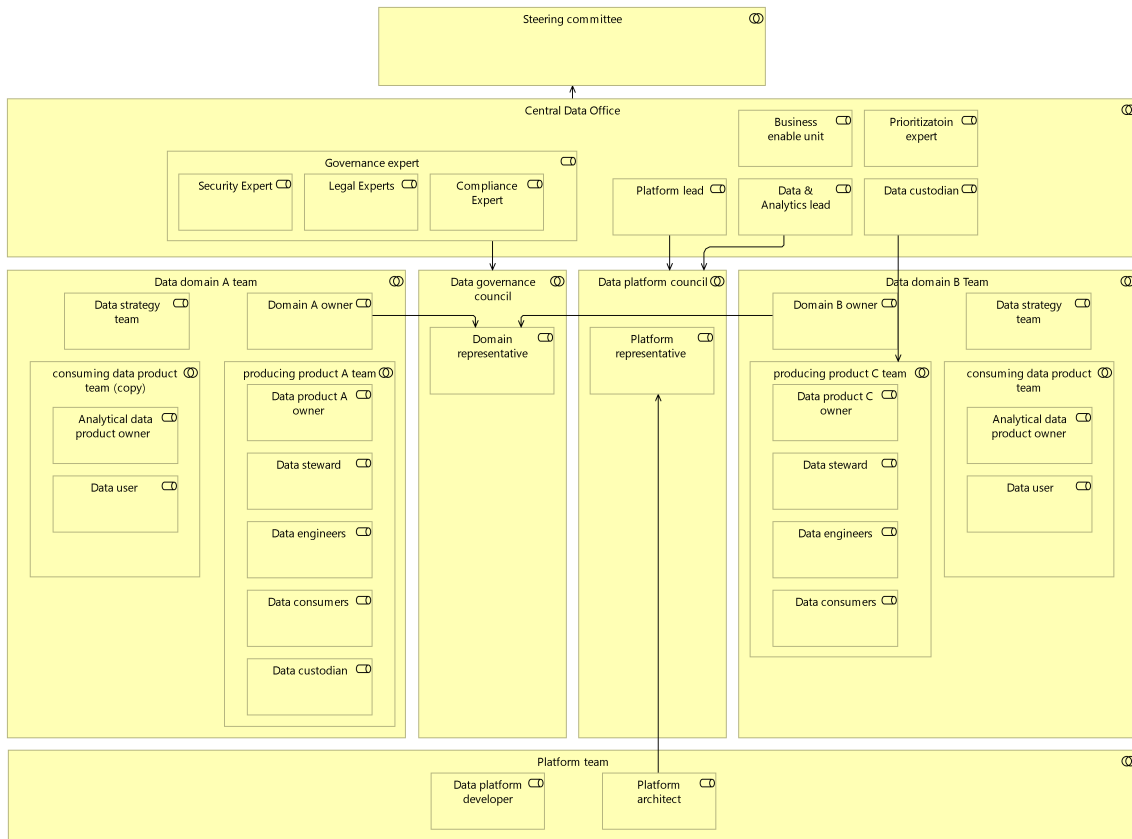
FIGURE 5.5: Simplified DGS of Energy 1 in the Archimate modelling language [53]

Energy 1 federated governance team

The federated governance team structure of Energy 1 as presented seems to be significantly different from the theoretical structure. There are multiple topic-specific councils in which roles from different teams come together to discuss global topics (TDIE1). These councils are the data governance council, data platform council and business value council. The business value council is not presented in the structure due to its lack of interaction with the rest of the teams and data governance roles. The business value council discusses the prioritization of data products and use cases. The data platform council discusses which data platform capabilities should be implemented in the platform and how the data platform automates decision-making. The data governance council discusses data governance-related topics and defines and describes standards and policies. But also, this council is a place to ask whether there already is a data product for your specific use case or whether you could make it yourself (TDIE1). The reason for the design decision to introduce multiple councils is that this is expected to increase the depth of discussions on topics and to improve communication by grouping people with the same background together(TDIE1). This choice is influenced by the size, culture and complexity of the organization. These different councils are aligned by the Central Data Office. This is a multi-disciplinary team of experts that are experienced in cross-functional decision-making. This team consists among others of the platform owner, the data governance expert and the data analytics expert (TDIE2). This team discusses the most important decisions with the steering committee of the

organization which consists of C-level executives.

Energy 1 domain teams

The second important team to discuss is the domain team. The domain teams in Energy 1 have similarities and differences in comparison to the domain teams presented in the theoretical structure. As proposed by theory, there is a clear split between data product-producing teams and data product-consuming teams (TDIE1). The producing team is responsible for the data products up until they are published in the data marketplace (TDIE2). The data product consuming team is responsible for using the data products from the marketplace to gain insights (TDIE2).

Energy 1 also defines a new team in the domain called the data domain strategy team which creates the domain specific strategy (TDIE1). *"Who were then reaching out to the various data domain owners, in order to make them understand the whole concept and convinced them from the overall idea of data mash, even though it was pushed on top down, however, you always need to make some some kind of homework in regard to convincing them and making them having them all as a joint and common understanding of the different roles"*(TDIE2).

At the start of the transformation, this team consists of external consultants from Deloitte *"So therefore, either we from Deloitte stepped in and took over this role, or someone from the central organization from the Center of Excellence or whoever could also step in order to take over this role"*(TDIE2). The data strategy team reaches out to the various data product owners to make them understand the whole concept of the data mesh, to increase the understanding of their specific roles and to assign the roles to people of the organization (TDIE1). This was needed to balance the top-down approach and to increase the understanding and willingness to change. Even though this is an important team, this is not considered a specific design decision regarding the target DGS because it is a temporary role. The responsibilities of this team are in a later stage of this research merged with the responsibilities of the enabling team.

Next to the data strategy team, the structure shows domain owners in the domain teams. A data domain owner is the overarching accountable person for any data product that gets created in his or her domain (TDIE1 TDIE2). This is different from a data product owner as that role is responsible for the specification of data products on an operational level. There is a hierarchical difference between the data domain owners and the data product owner. The data domain owners have a saying in the data governance councils while the data product owners are not present in the council (TDIE1). Cross-domain issues related to a specific data product can be escalated via the data domain owner (TDIE1). The data product owner assigns a data steward and a data custodian for his data product. These roles are similar to the roles defined by Abraham (2019).

Another important design decision regarding the data domain team is the decision on how to distribute the data product development developing capabilities. In the theoretical structure, each domain has its own data product developers. However, not all business domains do currently have the technical skills required (TDIE1). Therefore, someone from the data platform team took over this role for the domains or the domain uses resources from an external party (TDIE1). This results in a change in the responsibilities of the data platform team as they have an additional responsibility to offer these development capabilities. These capabilities are data engineering, data

architecture and software development skills. Before the data mesh migration, most of these capabilities were centralized in the central data team.

Another important finding related to the domain is the definition of domain boundaries. There are different ways to define the boundaries of the domains according to the interviews, the structure can be based on processes, corporate functions, systems or data context (TDIE1). This case uses domain boundaries based on data-object context, this goes across functions, processes and systems in order to break silos and to bring different people together that haven't been communicating earlier. Whether this is the right choice depends on the organizational culture, politics within the organization and senior management support (TDIE1). The domain boundaries do not have a significant impact on the specific roles and responsibilities but are important when operationalizing the data mesh.

Energy 1 allocation of decision-making authority

The changes in the federated team and the structure of the councils result in a change in decision-making allocation. The hierarchical decision-making changes from a flat hierarchy to a more pyramid-structured hierarchy. This is because of the introduction of the domain owner creates a hierarchy within the domain and the introduction of the central data team creates a hierarchical structure between the councils and the central data team.

The location of global decision-making authority is mostly centralized within the central data office. This is the place where all the experts are present to make the decisions on cross-functional and cross-domain data governance initiatives. This structure is expected to not affect which decisions are made centralized and decentralized. The global standards affecting multiple domains and enabling data interoperability still need to be made centralized and decisions only affecting one domain can be made locally in the domain. According to the interviews, finding which decisions belong to which category is difficult. A best practice here is to start by categorizing a lot of decisions as if they are affecting multiple domains and start with strong centralized decision-making. Slowly after reaching a higher maturity, the domains can gain more flexibility. Based on these findings, the following decisions are identified in the Energy 1 case:

- The introduction of a steering committee that oversees the decisions made by the federated governance team

- Introduction of multiple councils with their topic-focused discussions

- Introduction of a data strategy team that creates the domain-specific strategy and improves the culture towards the data mesh paradigm

- The introduction of a data domain owner that is accountable for the data product in its domain and for the overall success of the domain and solely represents the domain in the data governance council

- The use of the data developer capabilities from outside the domain or outside the organization because the capabilities are not yet present in the domains themselves

5.4.2 Telecom 1 case

Telecom 1 is an organization that was relatively decentralized without strong data governance policies or standards. The migration towards a data mesh architecture was expected to break the business domain silos and improve data sharing and data quality

within the business domains. *"The current situation could be seen as a decentralized governance model where everyone could make their own choices regarding data governance standards, rules and policies"* (TDIT1). The architects that were responsible for the creation of this structure used focus groups, workshops, literature [23, 39] and their experience to design these structures. A detailed introduction to this case is presented in **Section 4.4.2**

Telecom 1 organizational structure

To be able to compare the practical organization structure with the theoretical structure, the theoretical structure presented in **Figure 5.4** is used as guidance during the interviews. The resulting organization structure of the data governance within Telecom 1 is presented in **Figure 5.6**. After the creation of this structure, the case representatives are interviewed according to the same semi-structured interview protocol as the Energy 1 case (**Appendix F**).
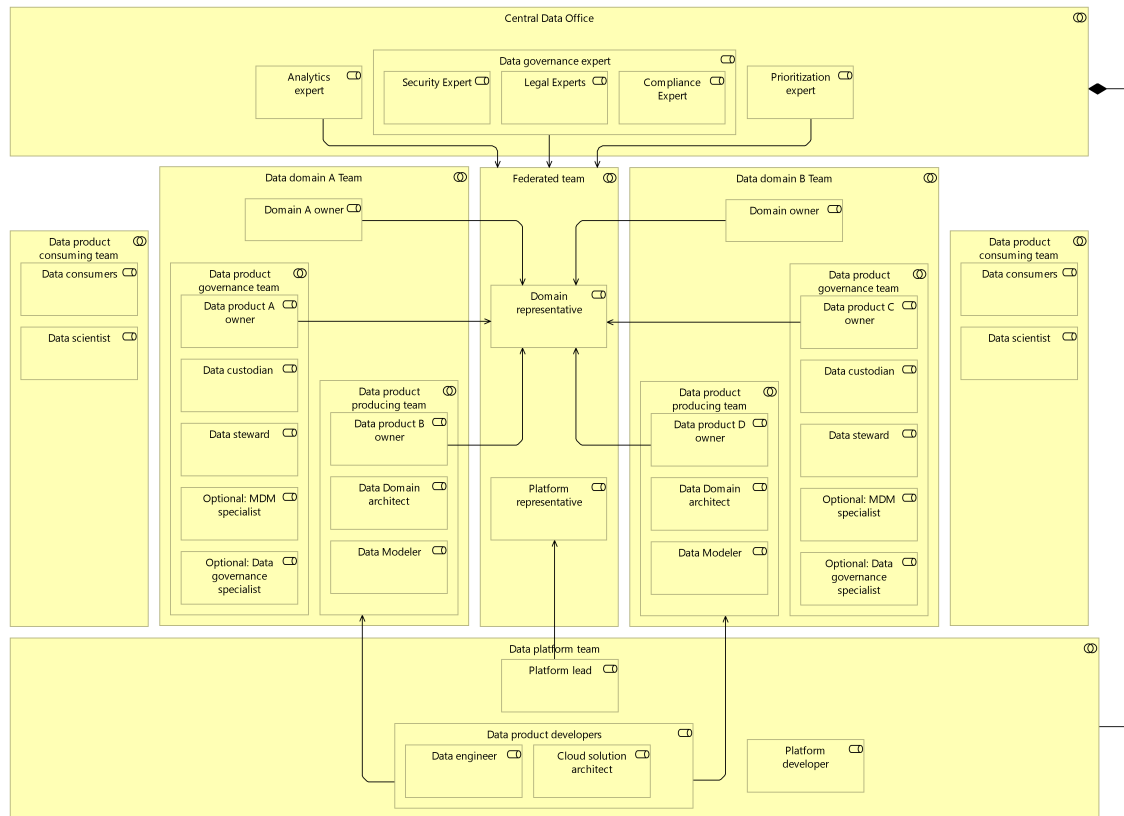
FIGURE 5.6: Simplified DGS of Telecom 1 in the Archimate modelling language [53]

Telecom 1 federated governance team

The first design decision mentioned during the interview that differentiates from the theoretical data mesh structure is the use of a Central Data Office (TDIT1 & TDIT2). This is a team that is comparable with the SMEs presented by Dehghani because they provide knowledge in the federated governance team. However, in the organization design of Telecom 1, these experts do have additional responsibilities because, besides their advisory role for the federated team, they do also have the decision-making right on

the global decisions made in the federated team (TDIT1 & TDIT2). This is different from the theoretical structure of Dehghani because there, only the domain representatives and platform representatives in the federated governance team make the decisions regarding global data governance initiatives. The reason for this design decision is that there is a strong need for central alignment regarding the data governance initiatives. To be able to make these decisions, also a prioritization expert, data & analytics expert and data governance expert are present in this central data Office. This is different from the theoretical structure proposed by Dehghani (2019) because she states that "since data governance is founded on a federated structure, there is not one specific data governance role".

Additionally, during this data governance design process, there is a deliberate choice made to design one single data governance council instead of multiple as stated in (PIT2): *"So I think that at this point, a single board should be fine to touch upon all important issues. But as more and more domains get on board, [...] that council is going to be too big to be effective. At that moment you should think about splitting up this council"* (PIT2). This supports the presence of the design decision made by Energy 1 regarding the amount of data governance councils.

Telecom 1 domain teams

A second design decision discussed during the interview is about the distribution of data product development capabilities. For this, Telecom 1 has used a "consulting model" in which the data development capabilities are centralized in the data platform team instead of having the skills and roles for data product development in the domains (TDIT1). *"Purely theoretical, there should be only infrastructure in the platform team, but they were also creating data products"* (TDIT1). The people with data product development skills help data domain teams without data product development capabilities with creating data products. The data product developers do not have a specific domain to assist. This is used as a temporary solution as there are currently not enough people with data development capabilities available to divide them in their domain and because the people with these capabilities are currently also in a centralized team and *"rearranging people in different teams is a tough discussion because people want to keep their [current] teams"* (TDIT2).

The third design decision made is the decision to assign data domain owners. Even though Dehghani explicitly mentions the role of "data domain product owner" and mentions that specific business domains can have multiple data products, the role of the data domain owner is not mentioned. At Telecom 1, they chose to define a data domain owner role which is responsible for overseeing the success of the whole domain and the prioritization of data product development within the domain (TDIT2,PIT2). These domain owners are people from the business and together with the data product owners represent their domains in the federated data governance team (PIT2).

The final decision mentioned during the interview was to create a clear split of responsibilities within the data product-producing team. Within the data product-producing team, Telecom 1 has assigned a producing team and a governance team. *"I think that there should be data product working groups in the domain and working groups that focus on the data governance"*PIT2. This producing team is responsible for making sure that the data that they use is correctly put together according to the data architecture and data models. The governance team is responsible

and accountable for the compliance of the data products according to the global governance initiatives (TDIT2). *"they need to monitor the data quality of their domain. Look at any issues that have been raised, and all the issues that have been raised, is it showing that there needs to be some sort of data governance or data management improvement initiative that needs to be introduced"* (PIT2).

Telecom 1 allocation of decision-making authority

The change from a siloed decentralized data governance framework to the more federated governance framework proposed by Dehghani resulted in a change in decision-making allocation. At the start, the hierarchical decision-making changes from a flat hierarchy to a more pyramid-structured hierarchy. This is because of the introduction of the domain owner that creates hierarchical decision-making within the domain and due to the introduction of the central data team that has a higher decision-making authority than the domain teams.

Regarding the allocation of decision-making authority, Telecom 1 focused a lot on centralization versus decentralization (TDIT1 & TDIT2). Because of the organisational structure of Energy 1 and the need for standardisation, they started with a centralization of decision-making. The aim for the longer term is to decentralize again but with global standards. TDIT1 states the following: *"I don't think that realistically it's feasible to do it as like a one bang kind of approach. It's also not sustainable because what you want is you want people to get into the culture. So what we've done is we've introduced like intermediate targeting target operating model as they start to set up their domains and get their data management maturity up"*. During the interview is addressed multiple times that it is expected to be a process with continuous feedback on the level of federation. The process of centralization and decentralization of decision-making is similar to the Energy 1 case. This is interesting because they both come from a very different organizational structure.

Based on these findings, the following decisions are identified in the Telecom 1 case:

- The introduction of a Central data office that has a higher decision-making power than the data governance council in which the domain and platform representatives come together

- The distribution of the data developer capabilities outside of the domain because the capabilities are not yet present in the domains

- introduction of a data strategy team that creates the domain-specific strategy and improves the culture towards the data mesh paradigm

- The introduction of a data domain owner that is accountable for the data product in its domain and for the overall success of the domain and represents the domain together with the data product owner in the data governance council

5.4.3 Additions of subject matter experts

The organization structures and the design decisions are discussed per case with the SMEs and interviews are conducted according to the interview protocol presented in **Appendix F**. Validating with subject matter experts is crucial to ensure the accuracy and reliability of research findings. The interviews with the subject matter experts discussed possible design decisions that they encountered to enrich the design decisions

resulting from the case studies. The important findings during the interviews are presented in this section. The subject matter expert interviews are referenced as follows:

- TDIDM1 = Treatment Design interview with Data mesh Expert 1

- TDIDG1 = Treatment Design interview with Data Governance Expert 1

- TDIOM1 = Treatment Design interview with Operating Model Expert 1

Domain team

During the discussion on the domain team, the data product developer capability distribution is further elaborated by stating that: *"One option is to have that every data product developer is assigned to one domain at the time and stays there. But this is really hard to achieve. [...] The other option is to have a pool of data product developers that can contribute to several domains at the same time. Or you can actually bring a developer and assign this developer for a certain domain for like three months or four months until the project is over. So one domain at a time, but they are not tied to one [single] domain"* (TDIDM1). This resonated with the findings from Energy 1 and Telecom 1 where there is a mix of the data product developer pool and the data product developers specifically assigned to multiple domains.

Besides the data product developers, the role of the domain owners is really important as the data mesh expert states that *"the domain owner, depending on the complexity of the domain, or the data products might shift more towards the strategy and the business and application of these data products for a business and let the data specific knowledge go to the data product owner"* (TDIDM1). If the maturity is low and the domain only has a few data products, the data domain role and the data product role can be executed by the same person (TDIDM1). Otherwise, these roles are expected to split and the data domain owner's responsibilities shift more towards the business and application of the data products while the role of the data product owner is responsible for the components of the data products that require data-specific knowledge.

Enabling team

The second important addition is the possibility to introduce a new team outside of the current federated governance team, data platform team or domain teams that are responsible for enabling the domains (TDIDM1 & TDIDG1). This enabling team is responsible for making sure the domain knows which capabilities they are missing and then helping with acquiring these capabilities instead of lending their own capabilities towards that team. These capabilities could be data product development capabilities, but also expert knowledge needed to realize the goals of the domains like knowledge on how to implement a new workflow like DevOps or DataOps.

5.5   Resulting design decisions

The design decisions resulting from the case studies and the interviews with the SMEs are analysed and the inferred design options from theory and practice that are perceived influential for the structural data governance mechanisms are presented per team in this section. The following design decisions are identified during this study;

- Federated governance team design decisions
    - Presence of a steering board
    - Amount of federated governance councils
- Data Domain team design decisions
    - Data domain ownership
    - Domain representatives structure
    - Data product developer capability allocation
- Federated governance team design decisions
    - Presence of an enabling team

5.5.1 Federated governance team design decisions

When designing the federated governance team, there are multiple design decisions with resulting configurations. Because the design decisions influence each other, the decisions are used as classification decisions and do not represent the order in which the decisions should be made.

The first classification of components in the federated team is the presence of a centralized steering board. The centralized steering board is a team of senior executives that is not directly part of the federated governance team but acts as a team of decision-makers that have a higher decision-making authority than the federated team(s). This resulted in the following two organizational configurations:

Configuration 1.1.1: Federated governance team(s) without a steering board
In a structure that does not have a steering board with decision-making authority, all decisions are made by the federated governance team. Subject matter experts can advise the federated governance team, but the federated governance team is responsible and accountable for the success of their decisions on data governance initiatives.

Configuration 1.1.2: Federated governance team(s) with a steering board
In this structure, there is a steering board that oversees the decisions made by the federated governance team(s). This means that the federated governance teams discuss the data governance initiatives and give advice to the steering board regarding these initiatives.

The second classification of components in the federated team is the number of federated governance councils. After analysing the results of the case study and comparing them to the theory, the following two organizational configurations are identified:

> **Configuration 1.2.1: Single federated governance council**
>
> In organization structures with a single federated governance council, this team consists of subject matter experts, domain representatives and data platform representatives. In this team, all data governance initiatives are discussed. This means that both the standards affecting the domains and the ways the data platform implements computational governance are discussed within this team. Additionally, this team discusses the prioritization of data product development and platform capability development. Due to the discussion on multiple topics with people from multiple disciplines, managers and/or facilitators that guide the discussion are needed.

> **Configuration 1.2.2: Multiple federated governance councils**
>
> In organization structures with multiple federated governance councils, the discussions within the councils and the representatives within the councils depend on the focus of that specific team. This configuration is expected to be combined with configuration 1.1.2 because it is important to align the discussion within the different teams and with that provide a holistic view of the decisions. The focus of the different teams is dependent on the organization.

Besides these design decisions, it is important to mention that also the availability of specific experts can be different from the experts mentioned in theory. The interviews showed that prioritization experts, data & analytics experts and governance experts could also be part of the federated governance team(s).

5.5.2 Data domain team design decisions

The design of the data domain team can be different per organization and also different per domain team. The following design decisions are identified with their corresponding configurations. The first design decision is the decision regarding the hierarchy within the domain and the presence of domain owners.

> **Configuration 2.1.1: Collaborative data domain ownership**
>
> In configurations with collaborative Data domain ownership, the accountabilities for specific data products are located by their corresponding data product owner in the domain. These data product owners are responsible for their data products and for how their data products deliver business value to the rest of the organization. There is not a single person accountable for the success of the domain. This is a collaborative effort and is measured by the accumulating success of the data products.

> **Configuration 2.1.2: Dedicated data domain owner**
>
> In domains with dedicated data domain owners, a single person is accountable for the created business value of the domain. This way the domain is seen as an entity with its own measurable success. This data domain owner role can be carried out by someone that also has the responsibilities of a data product owner and is likely to be someone from the business side. This creates a hierarchy in the domain and because the data domain owner is accountable, it is likely that data product owners report to the domain owner about the business value of their data product and the data domain owner can make decisions on how the data product owners should improve the success of their data products.

As described, a dedicated data domain owner creates a hierarchy in the domain. This hierarchy can influence the role of domain representatives in the federated governance

structure. There are 2 configurations identified:

Configuration 2.1.2.1: Hierarchical domain representative structure

In this configuration, only the data domain owner is present in the federated governance council. This means that data governance issues or the need for new data governance initiatives experienced in the domain by data product owners should be reported to the data domain owner in order to be discussed with the federated governance council. This increases the responsibility of the data domain owner and requires good communication between the data domain owner and the data product owner(s).

Configuration 2.1.2.2: Flat domain representative structure

In this configuration, both the data domain owner and the data product owner are present in the federated governance council(s). The data product owners are there to discuss data product-specific details and the data domain owners' responsibility in this team is to ensure that with the decisions made in the data governance team, the domain can still deliver value to the organization.

Besides the ownership within the domains, the responsibilities of creating data products and making sure they are compliant with global governance initiatives can also be designed in multiple ways. This allocation of the data product development capabilities can be configured in three different ways:

Configuration 2.2.1: Dedicated data product developer per domain

The perceived ideal configuration is to have (at least) one dedicated data product developer per domain. This data product developer can be responsible for multiple data products, but it is located in a single domain and possesses specific domain knowledge that must be used for the creation of the data products.

Configuration 2.2.2: Dedicated data product developer for multiple domains

In this configuration, the responsibility of the data product developer to implement the data governance initiatives in the data products is assigned to a person outside of the domain. This data product developer can be assigned to multiple domains at the same time, or this person can be assigned to a domain for a specific period of time or until a certain maturity level is reached. This way, the data product developer is able to develop domain-specific knowledge.

Configuration 2.2.3: Non-dedicated data product developer pool

The third configuration contains a data product developer pool. This means that it is the responsibility of the data product developer to implement the data governance initiatives in the data products are also assigned to a person outside of the domain. The capabilities of these developers can be used to help with the technical implementation of the data products without the aim to develop extensive domain-specific knowledge about the data products.

5.5.3 Enabling team design decision

Within the organizational structure, a lot of roles are dedicated to enabling people or teams. An enabling team is a specialized team within an organization that is responsible for managing and supporting the capability-building of other teams within the organization. In the context of data governance, an enabling team may provide

guidance, training, and support to different domains or departments within the organization to improve their data management practices and develop their capabilities in data governance. The team does this from a Center Of Excellence (COE) point of view in which this team is also responsible for developing and sharing best practices and supporting the organization in achieving its goals related to the specific area of expertise.

The decision to have an enabling team in an organizational structure for data governance depends on various factors. An enabling team can be beneficial in ensuring the successful implementation of data governance by providing guidance, support, and resources to the different domains within the organization (TDIDM1). They can also help promote a culture of data-driven decision-making and facilitate collaboration across different departments. This enabling team centralizes the responsibility of capability enablement which promotes a central enabling plan and central steering to improve the development of the capabilities. Ultimately, the decision to have an enabling team should be based on the organization's goals, resources, and specific needs for effective data governance (TDIDM1). This design decision has two organizational configurations.

A decision added by analysing the subject matter expert interviews is the decision to create an enabling team. This team is responsible for capability acquisition within the data domain teams. This team consists of subject matter experts on data governance, data product development or ways of working. This team is responsible for building data governance capabilities in the domain and computational governance capabilities in the data platform team. In the organization structure of a data mesh, this team could be absent or present. If the enabling team is absent, the responsibility of acquiring the needed capabilities for the team is in the team itself.

Configuration 3.1.1 Absent enabling team

An organizational configuration with an absent enabling team is one in which individual teams within the organization are responsible for their own capability development in the area of data governance. In this configuration, each team has the freedom to shape their development according to their specific needs and preferences, based on their unique organizational context and objectives.

This approach can be effective in organizations where there is a high level of expertise and knowledge within the different teams. Each team can leverage their strengths to develop their own capabilities, which can then be shared across the organization. However, this approach may also result in inconsistency in the development of capabilities across different teams, leading to potential gaps in data governance practices across the organization.

Configuration 3.1.1 Present enabling team

In this configuration, a dedicated enabling team is responsible for developing and implementing data governance practices across the organization, rather than leaving it up to individual teams. The benefits include greater standardization and coordination, but there may be less flexibility to adapt to individual team needs.

The resulting design decisions and their source are presented in **Figure 5.7**.

Team	Decision domain	Influential factors	Configuration	Reference Dehghani (2019)	Energy 1 support	Telecom 1 support	SME mentions	Advantages	Disadvantages
Federated governance team	Steering board	Organizational culture, Organizational strategy, organizational size, organizational complexity, C-level Management	Configuration 1.1.1: Federated governance council(s) without a steering board	x				High decision making speed, Low human resource costs	Difficult to locate exact accountabilities, lack of holistic view on decisions, lack of strategic alignment, Lack of leadership,
			Configuration 1.1.2: Federated governance council(s) with a steering board		x	x	1	Clear accountabilities , strong strategic allignment	Low decision-making speed, high human resource costs
	Federated governance council(s)	Data mesh maturity, Organizational culture, organizational size, organizational complexity, C-level management	Configuration 1.2.1: Single federated governance team	x		x		Central alignment within a single team	Need for facilitators, relatively large team, need for prioritization of decision making
			Configuration 1.2.2: Multiple federated governance teams		x		1	Focused decision-making, in depth discussions on specific topics	Risk of misalignment, slow decision making
Data domain teams	Data domain ownership	Data Mesh maturity, Domain complexity, organizational culture, Human resource availability	Configuration 2.1.1: Collaborative data domain ownership	x				No need for business holistic view of business-person, fast decision making	Unclear allocation of accountabilities, lack of decision maker,
			Configuration 2.1.2: Dedicated data domain owner		x	x	1,2,3	Clear allocation of accountabilities	Slow decision making, people needed that can have a holistic view,
	Domain representative structure	Organizational culture, Organizational strategy, organizational size, organizational complexity, C-level management	Configuration 2.1.2.1: Hierarchical domain representative structure			x	1,2,3	High decision-making speed	Effort required for good communication
			Configuration 2.1.2.2: Flat domain representative structure		x		1	Centralization of communication, discussion of data product specific problems	Slow decision making, unnecessary people in council,
	Data product developer capability allocation	Business Data and IT alignment maturity, Domain complexity, human resource availability, capability availability	Configuration 2.2.1: Dedicated data product developer per domain	x			1,2,3	High domain specific knowledge	High need for human resources with data development capabilities, low flexibility
			Configuration 2.2.2: Dedicated data product developer for multiple domains		x		1,3	Moderate feasibility, moderate domain knowledge	Need for prioritization of data product developers
			Configuration 2.2.3: Non-dedicated data product developer pool			x	1,2,3	Low need for human resources with data development capabilities, high flexibility	Low scalability, low domain knowledge, need for prioritization
Enabling team	Presence of Enabling team	Human resource availability, organizational culture, organizational strategy	Configuration 3.1.1: Present enabling team				1	No need for extra human resources, No need for the creation of a communication or enabling plan	Lower capability development, lack of change management plan
			Configuration 3.1.2: Absent enabling team				1	Clear responsibilities regarding capability development, extra guidance in capability development	Need for extra human resources, slow development of capabilities

FIGURE 5.7: Resulting option table

The design decisions and configurations found during this investigation are used to create the first version of the guidance tool. This tool is presented in the next section.

5.6 Guidance tool version 1

The design decisions presented in the previous section are used to create a guidance tool. This guidance tool is created in PowerPoint and consists of introduction slides and design decisions. The following section discusses the components of the guidance tool and presents the slides. The red boxes with the component IDs are added for clarity during the discussion of the slides and are not present in the actual tool. Furthermore, the influential factors in the figures are red if they are removed during the validation study, green if they are added during the validation study and black if they were not removed nor added

5.6.1 Guidance tool V1 components

The first introduction slide is a tool guide presented in **Figure 5.8**. This slide presents the title and a short explanation of the slide (1.1), the usage and the goal of the tool (1.2), the method of navigation within the tool (1.3) and the relations and entities within the tool (1.4).

FIGURE 5.8: Guidance tool v1 slide 1: Tool guide

The second introduction slide is an example of how the different configurations can form a DGS and is presented in **Figure 5.9**. This slide presents the title and a short explanation of the slide (2.1), the buttons to navigate to other introduction slides (2.2), the different configurations build up the theoretical structure and the practical structure from Energy 1 (2.3), and a simple visual representation of the theoretical and practical DGSs (2.4).
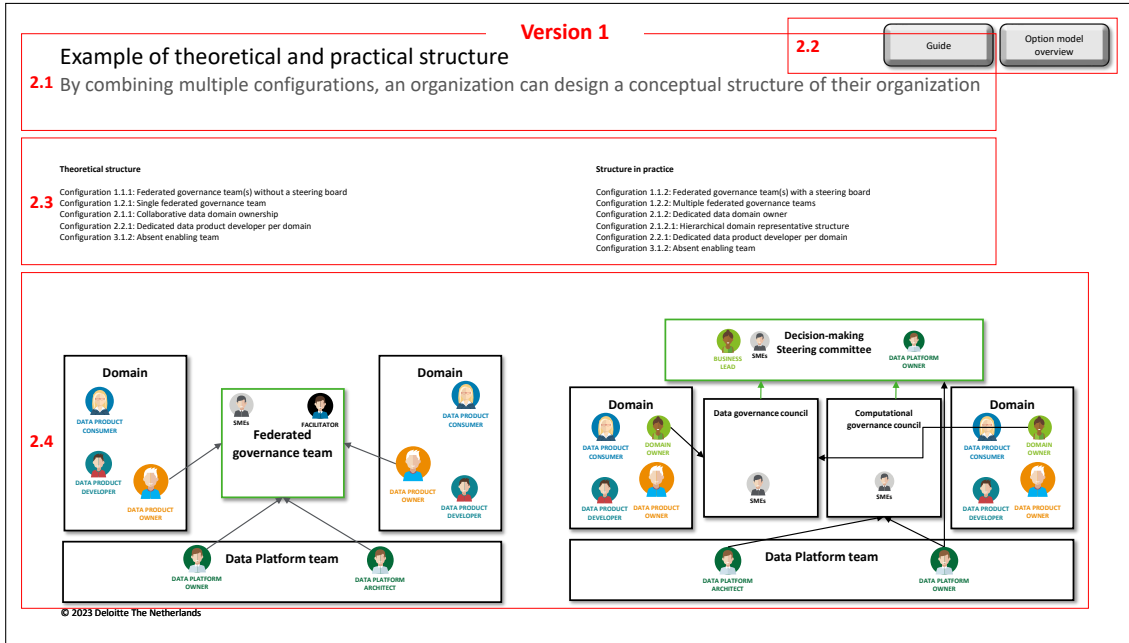
FIGURE 5.9: Guidance tool v1 slide 2: Example of theoretical and practical structure

The third and last introduction slide is a presentation of the design option model and is presented in **Figure 5.10**. This slide presents the title of the slide (3.1), the buttons to navigate to other introduction slides (3.2), an option tree with the teams, the design decisions and the configurations (3.3), and buttons to navigate to the slides of the design decisions (3.4).
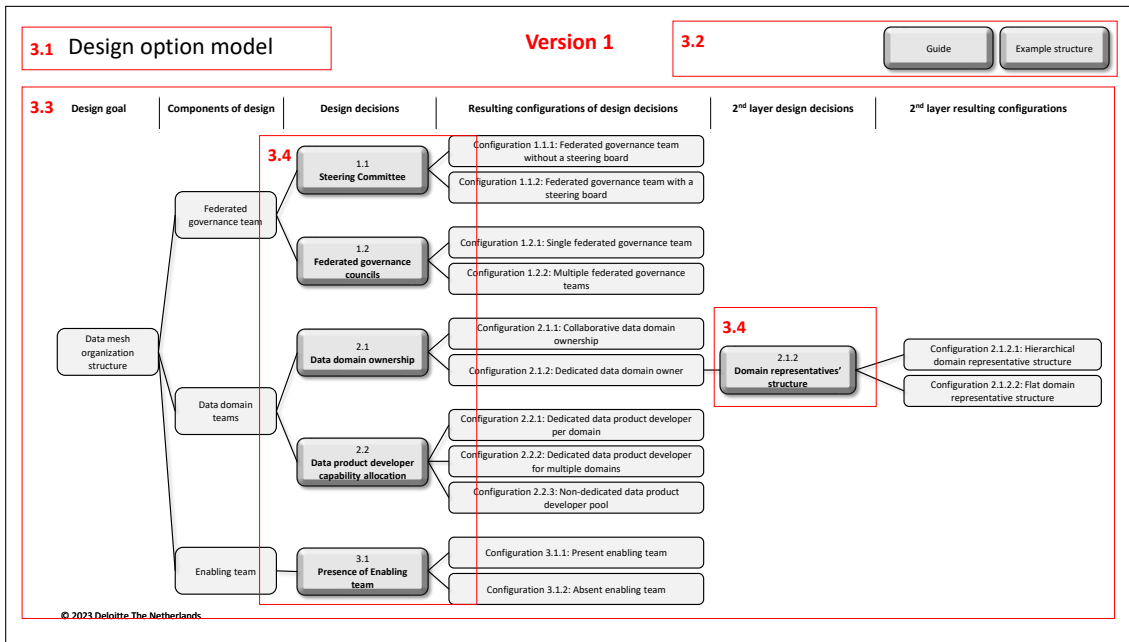
FIGURE 5.10: Guidance tool v1 slide 3: Design option model

The remaining slides are the slides explaining the design decisions and are presented in

Figure 5.11, **Figure 5.12**, **Figure 5.13**, **Figure 5.14**, **Figure 5.15** and **Figure 5.16**. All these slides have a similar structure. This structure consists of the title of the design decision with a question that can be asked during the discussion with stakeholders (X.1), buttons to navigate back to the options tree (X.2) an explanation of the organizational configurations (X.3), a simple visual representation of the configurations (X.4), the pros and cons of the configurations (X.5) and the factors influencing the design decision (X.6). Configuration 2.1.2 of design decision 2.1 has an extra button to navigate to design decision 2.1.2 because this design decision is important to discuss when configuration 2.1.2 is chosen.

FIGURE 5.11: Guidance tool v1 slide 4: Steering committee

FIGURE 5.12: Guidance tool v1 slide 5: Federated governance council(s)

FIGURE 5.13: Guidance tool v1 slide 6: Data domain ownership

FIGURE 5.14: Guidance tool v1 slide 7: Domain representative structure

FIGURE 5.15: Guidance tool v1 slide 8: Data product developer capability allocation

FIGURE 5.16: Guidance tool v1 slide 9: Enabling team

5.6.2 Guidance tool V1 requirements contribution

As earlier addressed, the tool needs to fulfil the requirements presented in **Section 5.1**. The components presented in the guidance tool contribute to the fulfilment of these requirements. The relations between the components and the requirements are presented in **Section 5.6.2**

There are additional requirements that are not directly fulfilled by the components. FR5 (The tool must be presentable during discussions with stakeholders) is fulfilled by creating the tool in PowerPoint. NREI1 (The use of the tool during discussions with stakeholders must reduce discussion time) is a requirement that is fulfilled by the use of the tool. Additionally, to fulfil NRUT3, an effort is made to present the components in non-organization-specific roles, teams, and structures and by validating with EA experts whether the design decisions and the organizational configurations are generalizable.

The contribution of the components to the requirements is validated with expert interviews, and this process is presented in the next chapter.

5.7 Treatment design and development summary

In conclusion, this chapter describes the development of a theoretical conceptual DGS according to book of Dehghani (2019). This model is applied to analyze the practical DGSs of two case studies, which enabled the retrieval of specific design decisions and organizational configurations.

The insights gained from this analysis were further enriched through engagement with subject matter experts. By incorporating their opinions, the design decisions and organizational configurations are improved to better align with practical considerations.

Table 5.3: The contribution aim of the components towards the requirements (X represents all slide numbers of the design decision slides)

Requirement IDs	Tool guide v1 components																		
	Introduction slide components												Design decision slide components						
	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	3.1	3.2	3.3	3.4	X.1	X.2	X.3	X.4	X.5	X.6	6.7
FR1											x		x						
FR2											x				x				
FR3																	x		
FR4								x								x			
FR5																			
NRC1											x		x		x	x	x	x	
NRUN1													x		x	x	x	x	
NRUN2								x					x		x	x			
NRUN3															x				
NRUN4															x	x	x		
NRUN5																	x		
NRUN6																		x	
NRUN7				x												x			
NRUT1		x	x		x					x	x	x	x	x	x	x	x	x	x
NRUT2							x									x	x	x	
NRUT3								x											
NRUT4											x		x						
NRUT5																	x		
NRUT6																		x	
NRUT7								x											
NREI1																			
NREI2											x		x						
NREI3											x			x	x	x			
NREI4																	x		
NREE1																			

Finally, the resulting design decisions and organizational configurations are presented in a guidance tool. In this chapter, RQ4a, RQ5a, RQ5b, RQ5c and RQ5d, are addressed.

In the next section, the correctness of the design decisions and configurations is validated and the quality of the tool is measured based on its utility, understandability and its influence on the efficiency and effectiveness of the design process.

Chapter 6

Treatment validation and refinement

The goal of this validation research is to justify that the guidance tool contributes to stakeholders' goals when implemented in the problem context [82]. This is done by assessing the satisfaction of its requirements.

The validation research follows the validation methodology presented in **Section 2.4.4**. The created guidance tool is validated based on the requirements presented in **Section 5.1**.

As described in the methodology section, this validation step is performed through expert interviews and based on expert opinions. These interviews are summarized and the created evaluation document with responses and important comments is presented in **Appendix G**. The findings of the evaluation research contributed to answering RQ5. A total of six enterprise architects with data mesh experience (EI1, EI2, EI3, EI4, EI5 and EI6) and two enterprise architects without data mesh experience (EI7, EI8) are interviewed. EI7 and EI8 did not include the assessment of the correctness of the model. In this section, the requirement satisfaction based on the interviews is discussed and the refinement opportunities and refined tools are presented.

For referencing of the interviews, the following abbreviations are used:

- EI1 = Evaluation interview 1 with Energy 1 representative 1

- EI2 = Evaluation interview 2 with Telecom 1 representative 1

- EI3 = Evaluation interview 3 with Telecom 1 representative 2

- EI4 = Evaluation interview 4 with Data mesh expert

- EI5 = Evaluation interview 5 with Data governance expert 1

- EI6 = Evaluation interview 6 with Data governance expert 2

- EI7 = Evaluation interview 7 EA consultant without data mesh experience 1

- EI8 = Evaluation interview 8 EA consultant without data mesh experience 1

6.1 Correctness

In this research, correctness is referred to as the state of being accurate, true, or free from error or mistake according to the knowledgeable people interviewed in this research. The interviews included questions regarding the correctness of the design decisions, configurations, influential factors, advantages, and disadvantages. The support of these components is measured, and the acceptance threshold is set at 80%. Missing components added by interviewees are included in the guidance tool and validated in the second validation session. Through this acceptance criteria and through the validation of the interviewees, the fulfilment of NCR1, NCR2, NCR3 and NCR4 are tested.

In the following sections, important statements regarding design decisions and expert opinions are discussed. Additional comments about the correctness can be found in **Appendix G**. In the following tables, the red rows represent the disadvantages and the green rows represent the advantages.

6.1.1 Design decision 1.1: Steering board

This section discusses the evaluation interviews of the steering board described in **Section 5.5**. The results of the evaluation interviews regarding this topic are presented in **Table 6.1**. This shows that the interviewees agree on most of the components of this design decision and that the first version of this design decision is reasonably correct. However, there are some disagreements regarding this design decision which are presented below.

100 % Of the interviewees recognize that this decision has to be made and that it is an important decision. One of the interviewees mentions that *"it is unlikely that C1.1.1 will occur because, in this configuration, there is a lack of power from the leadership to drive the data governance initiatives to the rest of the organization"* (EI1). Other interviewees agree that C1.1.1 is less often the best configuration for the organization, however, they could think of organizations where this would be the case (EI1, EI2, EI3, EI5 & EI6).

One interviewee mentioned that there is a configuration missing. In that configuration, the steering board would be in the domain team and would be responsible for the strategic alignment of the domain with the organizational strategy (EI3). Deciding how the steering within the domain is done is interesting, however, this is considered outside of the scope of this design decision because this decision only affects global governance and not domain governance. EI6 adds to this that there are probable configurations missing, but due to the organizational differences, it is impossible to be exhaustive.

During the interviews, additional factors are mentioned like organization agility (EI2) and data governance maturity (EI3), but these factors are mentioned as possible factors without a strong argumentation, in addition, these factors are only mentioned once and therefore not included. According to EI2, "C-Level management" is not the correct wording as this way it can be perceived that the C-level executives choose the configuration. EI5 agrees and suggests changing this to a leadership culture.

The advantage of low human resource costs of C1.1.1 has the support of 33%. EI1, EI2 and EI5 mention that the steering committee does not meet frequently, and therefore the extra human resource used by the steering committee are negligible. The disadvantages of this design decision are agreed upon with the exception of "unclear accountabilities". EI2 and EI3 mention that in both configurations, accountability should be clearly defined.

Both interviewees add to this that in this configuration, it is important to choose whether there is a collaborative decision-making structure or that a single person is accountable for the decision-making.

The advantage of clear accountabilities of C1.1.2 has a low support of (33%) because the accountabilities still need to be clearly defined for an effective DGS. The steering committee could make this difficult because this requires a clear split of accountabilities between the council and the steering committee. EI4 and EI5 add "leadership involvement" to the advantages. Because when the leadership is involved in decision-making as in C1.1.2, there is more power behind the decisions making. The disadvantage of slow decision-making speed is partly agreed upon. IE4 and IE5 extend this disadvantage by stating that there is an additional decision-making approval needed, which could decrease the decision-making speed.

TABLE 6.1: The results of the evaluation interviews regarding the correctness of the steering committee design decision (N=6), green = advantage, red = disadvantage

Validation component		
Design decision: 1.1 Steering committee	100,00%	Yes
Organizational culture	100,00%	Yes
Organizational Strategy	100,00%	Yes
Organization size	100,00%	Yes
Organization complexity	100,00%	Yes
C-level management	66,67%	No
Leadership culture	Added	Not applicable
Configuration 1.1.1: decision-making data governance council	83,33%	Yes
Fast decision-making	83,33%	Yes
Low human resource costs	33,33%	No
Unclear accountabilities	50,00%	No
Lack of leadership support	100,00%	Yes
Lack of strategic alignment	100,00%	Yes
Lack of holistic view on decisions (already covered)	100,00%	No
Configuration 1.1.2: Advisory data governance council	100,00%	Yes
Clear accountabilities	33,33%	No
Leadership support	100,00%	Yes
Strong strategic alignment	100,00%	Yes
Slow decision-making	83,33%	Yes
High human resource costs	16,67%	No

6.1.2 Design decision 1.2: Federated governance council(s)

This section discusses the evaluation interviews of the federated governance council described in **Section 5.5**. The results of the evaluation interviews regarding this topic are presented in **Table 6.2**. This shows that the interviewees agree on most of the components of this design decision and that the first version of this design decision is reasonably correct. However, there are some disagreements regarding this design decision

which are presented below.

All of the interviewees recognized this design decision and its configurations. EI1 mentioned that another configuration could be a configuration in which there are multiple layers of governance councils. Even though this is interesting, this design decision is about the separation of concerns within the councils and not about the hierarchy of the councils. This could be introduced as a new design decision. The support for data mesh maturity and C-level management is low (50%), EI5 stated that "Data mesh maturity is too broad to be used as an influential factor"

There is sufficient support for all the advantages of C1.2.1, but according to EI2 and EI5, the formulation should be more precise to be effective. Regarding the disadvantages, all configurations got sufficient support, however, according to EI5, *"The disadvantages should be more high over and changed to Difficult to oversee the council due to its size, agenda and possible competing priorities"* EI3 supports this by stating that the "need for prioritization of decision-making" should be changed to competing priorities. Findings like these point out that the interviewees agree on the fundamentals of the disadvantages, but prefer different ways of formulating them.

Both the advantages and the disadvantages of C1.2.1 are supported. EI3, EI5 and EI1 added that there is less waste of human resources if people only attend the discussions to which they contribute.

TABLE 6.2: The results of the evaluation interviews regarding the correctness of the federated governance councils design decision (N=6), green = advantage, red = disadvantage.

Validation component	Support	Approved
Design decision: 1.2 Federated governance council(s)	100,00%	Yes
Data mesh maturity	66,67%	No
Organizational culture	100,00%	Yes
Organization size	100,00%	Yes
Organization complexity	100,00%	Yes
C-level management	50,00%	No
Configuration 1.2.1: Single federated governance council	100,00%	Yes
Central alignment within a single team (alignment effort)	83,33%	Yes
Need for facilitators (lack of council manageability)	83,33%	Yes
Relatively large team (lack of council manageability)	83,33%	Yes
Need for prioritization of decision-making	66,67%	No
Low Human resource effectiveness	Added	Not applicable
Configuration 1.2.2: Multiple federated governance councils	100,00%	Yes
Focused decision-making (Human resource effectiveness)	100,00%	Yes
In-depth discussion on specific topics (High human resource effectiveness)	100,00%	Yes
Risk of misalignment (High alignment effort)	100,00%	Yes
Slow decision-making (High alignment effort)	100,00%	Yes

6.1.3 Design decision 2.1 Data domain ownership

This section discusses the evaluation interviews of the Data domain ownership described in **Section 5.5**. The results of the evaluation interviews regarding this topic are presented in **Table 6.3**. This table shows that there are some components missing, but the overall design decision and the configuration are approved.

All of the interviewees supported the design decision, the configurations, and the influential factors of this design decision. EI1, EI2 and EI4 mentioned that the communication between the domain and the council is very important and that this is advantageous in this configuration. This configuration has a person that bridges the gap between the domain and the other teams and acts as a single point of contact. This advantage is introduced as a "single point of communication".

TABLE 6.3: The results of the evaluation interviews regarding the correctness of the Data domain ownership design decision (N=6), green = advantage, red = disadvantage

Validation component	Support	Approved
Design decision: 2.1 Data domain ownership	100,00%	Yes
Domain complexity	100,00%	Yes
Organizational culture	100,00%	Yes
Human resource availability	100,00%	Yes
Data mesh maturity	100,00%	No
Configuration 2.1.1: Collaborative data domain ownership	100,00%	Yes
Fast decision-making	50,00%	No
No need for a person with a holistic view (low role fulfilment effort)	66,67%	No
Unclear accountabilities (Low responsibility clarity)	100,00%	Yes
Lack of a decision maker (Low responsibility clarity)	100,00%	Yes
Lack of clear communication point for domain	Added	Not applicable
Configuration 2.1.2: Dedicated data domain owner	100,00%	Yes
Clear accountabilities	100,00%	Yes
Clear communication point for domain	Added	Not applicable
Need for a person with domain owner capabilities (low role fulfilment effort)	83,33%	Yes
dependent on strong domain definition	50,00%	No

6.1.4 Design decision 2.1.2: Domain representatives' structure

This section discusses the evaluation interviews of the Domain representatives' structure described in **Section 5.5**. The results of the evaluation interviews regarding this topic are presented in **Table 6.4**

This design decision is supported (83%). EI1 does not agree with this design decision as "it is unlikely that organizations choose for C2.1.2.2 as the domain owner is the bridge between the data governance council and the data product developers". If the data product owners themselves are attending the data governance councils, the data domain owner is redundant. EI3 acknowledges that it is very unlikely that organizations adopt C2.1.2.2, however, "it could be the right choice if the organization and the governing council is organized and small". Even though it is unlikely that 2.1.2.2 is chosen, this configuration was preserved as other interviewees perceived this as helpful during the design of the organization structure and it would support the discussion about the organization design regarding the positioning of the data domain owner.

Five interviewees mentioned that configurations are missing. EI1 and EI4 mentioned that there should be a configuration with only data product owners for domains without data domain owners. EI2, EI3 and EI5 mention the importance of bringing specific roles from within the domain as SMEs into the councils. Like data consumers ("another configuration could be one in which data consumers are present in the council" (EI2)), data owners ("The data owner role could also be in the council" (EI3)) or data product owners ("bring a data product owner as an SME in the domain" (EI5)). This new configuration with a domain representing subject matter experts replaces the configuration with strictly the data domain owner and data product owner in the council.

TABLE 6.4: The results of the evaluation interviews regarding the correctness of the domain representatives' structure design decision (N=6), green = advantage, red = disadvantage

Validation component	Support	Approved
Design decision 2.1.2 Domain representative structure	83,33%	Yes
Organizational culture	83,33%	Yes
Organizaitonal Strategy	50,00%	No
Organization size	83,33%	Yes
Organization complexity	83,33%	Yes
C-level management	33,33%	No
Configuration 2.1.2.1: Hierarchical domain representative structure	100,00%	Yes
Fast decision-making	83,33%	Yes
Smaller federated data governance council (council manageability)	83,33%	Yes
Effort required for good communication	100,00%	Yes
Configuration 2.1.2.2: Flat domain representative structure	50,00%	No
Centralization of communication	Removed	Not applicable
Data product-specific problem discussions	Removed	Not applicable
Slow decision-making	Removed	Not applicable
Large federated data governance council (council manageability)	Removed	Not applicable

6.1.5 Design decision 2.2: Data product developer capability allocation

This section discusses the evaluation interviews of the design decision Data product developer capability allocation structure described in **Section 5.5**. The results of the evaluation interviews regarding this topic is presented in **Table 6.5**

All interviewees recognize this design decision and all three configurations. Only the advantage of the "high feasibility" of C2.2.2 is not supported because this depends on the organization. IE5 introduces a new advantage of C2.2.3 by stating that with a centralized developers pool, it is easier to standardise the way of working and leverage best practices. Low scalability is not supported and multiple interviewees added low efficiency to C2.2.3 (EI1, EI3 & EI5).

TABLE 6.5: The results of the evaluation interviews regarding the correctness of the Data product developer capability allocation design decision (N=6), green = advantage, red = disadvantage

Validation component	Support	Approved
Design decision 2.2: Data product developer allocation	100,00%	Yes
Domain complexity	66,67%	No
Human resource availability	100,00%	Yes
Capability availability	100,00%	Yes
Configuration 2.2.1: Dedicated data product developer in domain	100,00%	Yes
High domain-specific knowledge	100,00%	Yes
High efficiency	100,00%	Yes
High need for human resources (low human resource flexibility)	100,00%	Yes
Configuration 2.2.2: Dedicated data product developer for multiple domains	100,00%	Yes
Moderate domain-specific knowledge	100,00%	Yes
High feasibility	33,33%	No
Human resource flexibility	Added	Not applicable
Effort needed for prioritization	100,00%	Yes
Configuration 2.2.3: Non-dedicated data product developer pool	100,00%	Yes
Low need for human resources (Human resource flexibility)	83,33%	Yes
High flexibility (Human resource flexibility)	100,00%	Yes
High standardization	Added	Not applicable
Need for prioritization	100,00%	Yes
Low scalability	50,00%	No
Low domain knowledge	100,00%	Yes

6.1.6 Design decision 3.1: Presence of enabling team

This section discusses the evaluation interviews of the design decision Presence of enabling team described in **Section 5.5**. The results of the evaluation interviews regarding this topic is presented in **Table 6.6**

All interviewees recognize the design decision and the two configurations. During EI3, it was unclear to the interviewee whether a single person per team could be responsible for the enablement of the whole team. Therefore, the distinction between both configurations should be more clear. The distinction between both is that in C3.1, each team is responsible for its capability development and there is not a centralized effort to enable teams, while in C3.2 there is a central unit collectively responsible for enabling all teams and manages this centrally with a cross-team enablement strategy.

Both EI2 and EI5 added to the advantages of C3.1.2 that there is flexibility for the teams to manage their development and therefore. The benefit of "no need for extra human resources" is insufficiently supported because according to EI4 & EI5, the need for human resources is approximately equal, because enabling is required, however, the human resources are decentralized instead of centralized. "Lack of development capabilities" is not sufficiently supported, during the interviews, this disadvantage was replaced with lack of steering in capability development. This disadvantage could be a result of the disadvantage "lack of change management plan" however, this could also be a standalone disadvantage.

TABLE 6.6: The results of the evaluation interviews regarding the correctness of the Data product developer capability allocation design decision (N=6), green = advantage, red = disadvantage

Validation component	Support	Approved
Design decision 3.1: Enabling team	100,00%	Yes
Human resource availability	100,00%	Yes
Organizational culture	100,00%	Yes
Organizational strategy	83,33%	Yes
Team capability maturity	Added	Not applicable
Configuration 3.1.1: Absent enabling team	100,00%	Yes
No need for the creation of a communication or enabling plan	83,33%	Yes
No need for extra human resources	50,00%	No
Capability development flexibility	Added	Not applicable
Less development of capabilities	50,00%	No
low knowledge sharing	Added	Not applicable
lack of change management plan (Enablement steering)	100,00%	Yes
Configuration 3.1.2: Present enabling team	100,00%	Yes
Clear responsibilities regarding capability development (Enablement steering)	83,33%	Yes
Extra guidance in capability development (Enablement steering)	83,33%	Yes
High knowledge sharing	Added	Not applicable
Need for extra human resources	83,33%	Yes
Slow development of capabilities	66,67%	No
Low capability development flexibility	Added	Not applicable

After the validation interviews for correctness, the content of the tool is improved according to the feedback of the experts and the other requirements are tested.

6.2 Understandability

For further validation of the tool, the fulfilment to other requirements is addressed. The interviewees scored their agreement to the fulfilment of the requirements on a scale from 1 (totally disagree) to 7 (totally agree). The results with the minimum, maximum and average scores are presented in **Table 6.7**. The interviewees generally agree that the components of the tool are understandable. Especially the graphical representations seem to improve the understandability of the design decisions significantly as EI3 states the following: *"Love the graphics, this makes me immediately understand the differences between the options"*.

The two components that scored lowest on understandability are the influential factors and the explanation of the configurations. To improve the understandability of these components, experts suggested adding definitions of roles (EI1), adding definitions of influential factors (EI7 & EI8), shortening the configuration explanations, highlighting important components of configurations and making the influential factors more specific

as described in the correctness section (EI2).

Additional changes to improve the understandability are: changing the title and colours of the option tree overview (IE8), removing the steering committee in the visualization of design decision 1.2 (IE8), adding a footnote that there is likely a steering committee with multiple data governance councils, representing the (dis)advantages more comparatively and differentiate the data governance council and the federated governance team. Overall, the understandability of the tool can be improved by further explaining the influential factors and improving the explanations of the configurations.

TABLE 6.7: The results of the evaluation interviews regarding the agreement to statements on a scale from 1 (totally disagree) to 7 (totally agree) (N=8)

ReqID	Understandability requirements	Min	Max	Avg
NRUN2	The influence of the design decisions on the DGS is understandable	5	7	5,75
NRUN3	The explanations of the configurations are understandable	4	7	5,31
NRUN4	The differences between the configurations are understandable	5	7	6,00
NRUN5	The (dis)advantages of the design options are understandable	5	7	5,88
NRUN6	The influential factors are understandable	3	7	5,13
NRUN7	The graphical representations are understandable	5	7	6,38

6.3 Utility

The tool demonstrates a relatively high level of expected utility, as presented in **Table 6.8** and as expressed by the following quotes: EI1 states that "*This framework can really help with designing the structure of the teams and meetings. Without having such a tool at hand, there is a lot of brainstorming and comparing. With this tool it is easy, we can open it and discuss it which accelerates the process because many thoughts are already included in the tool*".

Furthermore, according to EI8 "*It is intuitive that you can click through from decision to decision and it is possible to directly navigate to the design decision that you want to talk about with your client [...] or with your team, to discuss the different scenarios and formulate a recommendation [..] This makes it very useful*".

The utility of the influential factors was also in depth discussed during the interviews. EI2 explicitly mentions that even though the influential factors are very organization dependent, they give a good starting point for the discussions with stakeholders. EI5 on the other hand scored the utility with a 4 and mentions that specifying the influential factors would increase their utility because they are currently very generic.

Other comments made regarding the utility are that the configurations should be shortened and sharpened to enhance utility in the design process (EI4) and that the example given in the tool and the tool guide should be more elaborate (EI8). Additionally, EI 8 mentions the following: "*The theoretical vs practical slide could be more elaborated. Add one-pager*

on how to use this tool, and extend the current page. Make the disclaimer that it is to support making the decision instead of making the decision".

The advantages and disadvantages scored to lowest on utility, especially during EI7 & EI8. These experts are less knowledgeable on the data mesh paradigm and were previously not involved in the research. They state that the advantages and disadvantages should be further elaborated and explanations for the advantages and disadvantages should be included in the tool. This explanation could help with assessing how the configuration could influence the design of the organization and how these advantages and disadvantages resonate with the current organization. Overall, the utility requirements are fulfilled and considered usable. The feedback from the experts is taken into account and the tool is refined based on their comments.

TABLE 6.8: The results of the evaluation interviews regarding the utility requirements

ReqID	Utility requirements	Min	Max	Avg
NRUT1	This guidance tool is easy to use during discussions with stakeholders	5	7	6,13
NRUT2	This guidance tool is easy to use during the design process	5	7	5,88
NRUT3	This guidance tool is useful during the design process for multiple organizations	6	7	6,50
NRUT4	This guidance tool is useful to distinguish different design decisions	5,5	7	6,13
NRUT5	The (dis) advantages are useful to support decision-making in the design process	4	7	5,63
NRUT6	The influential factors are useful to support decision-making in the design process	4	7	6,00
NRUT7	This guidance tool gives design configurations that could help during the design process	5	7	6,13

6.4 Efficiency

As described in the introduction, the tool aims to increase the efficiency of the organization design process in two ways. It tries to reduce the discussion time with clients during the design of organization structures and it tries to reduce the actual design time of the organization structures. The tool is evaluated based on its influence on efficiency and the results of this validation study are presented in **Table 6.9**

Experts agree that this tool could reduce the discussion time with clients by presenting key design decisions. According to EI3 and EI8, *"this could be a really helpful tool to help facilitate discussions during workshops" and "It would definitely contribute to the efficiency of the organization design of a data mesh".*

Experts also agree that this tool could reduce the design time of the organization structure by presenting example configurations. *"This tool sets out the key parameters and key decisions required to be taken into account" (EI3) and "This tool gives structure to the decisions and helps you with making decisions faster".* Furthermore, the case representatives that performed the organization design for their organization agree that

this tool would have helped their process and increased efficiency (EI1 & EI3).

However, this study only presents the expectation that the discussion time and design time would be reduced. A quantification of this expected reduction or empirical evidence of this reduction is missing. **Section 7.8.1** delves deeper into this limitation of the study.

Overall, according to the results from the evaluation study, there can be concluded that experts expect that the guidance tool improves the efficiency of the DGS design process, However, there is insufficient evidence to quantify this expectation or to conclude that the guidance tool improves efficiency.

TABLE 6.9: The results of the evaluation interviews regarding the efficiency requirements

ReqID	Efficiency requirements	Min	Max	Avg
NREI1	Using this tool would reduce discussion time about the organization structure	4	7	5,88
NREI2	Using this tool would reduce the design time of the organization structure	5	7	6,00

6.5 Effectiveness

Besides the increase in efficiency, it is important that the use of the guidance tool must not reduce the effectiveness of the design process. When asked how the use of this tool influences the effectiveness of the design process, none of the experts expect that this tool would decrease the effectiveness as the lowest score is 5.

An interesting statement during EI8 is the following: *"I don't feel comfortable saying that it gives you better or worse organizational structures, but it helps you with making the design process more effective and from higher quality because it presents the criteria that should be taken into account and the pros and cons of these decisions"*. Other experts also mention that it is difficult to say whether it would increase the effectiveness because the effectiveness is dependent on a lot of other aspects of the process like the taken action from the client's side (EI4) or the funding of the process and of the roles and teams (EI8). EI7 explicitly mentions that she does not foresee any problems that could occur and result in a less effective DGS design process.

However, similar to the evaluation of the influence on efficiency, this study only raises the expectation that the effectiveness of the DGS design process would not decrease. A quantification of this influence on the effectiveness or empirical evidence proving that the guidance tool does not reduce effectiveness is not given during this study. **Section 7.8.1** also delves deeper into this limitation of the study.

Overall, according to the results from the evaluation study, there can be concluded that experts expect that the use of the guidance tool does not reduce the effectiveness of the DGS design process, However, there is insufficient evidence to quantify this expectation or to conclude that the guidance tool improves efficiency.

TABLE 6.10: The results of the evaluation interviews regarding the effectiveness requirements

ReqID	Effectiveness requirements	Min	Max	Avg
NREE1	This tool affects the effectiveness of the organization structure design process. (1= decrease, 4 = unaffected, 7 = increase)	5	7	5,63

6.6   Guidance tool version 2

The tool is refined based on the findings from the evaluation interviews. This refinement process tries to improve the fulfilment of the requirements. The refinements with their influencing requirements and affected components of the tool are presented in **Appendix G.3**. Besides these refinements, alignment, spelling and grammar errors are corrected. The resulting tool is presented in **Appendix A**.

This evaluation chapter of the research document focused on the validation and refinement of the guidance model. The model consists of introduction slides with explanations of the use of the tool, the influential factors and the roles in order to improve its usability and utility.

To summarize, the treatment validation demonstrated that, according to experts, the refined guidance tool fulfils the correctness, understandability and expected utility requirements. Furthermore, enterprise architects expect that the use of the guidance tool improves the efficiency of the DGS design process without reducing its effectiveness.

Chapter 7

Conclusion

In this chapter, the research is concluded by answering the research questions formulated in the introduction of this research and by discussing the findings, the theoretical and practical contribution, its limitations and the directions for future research. The main research question answered in this research is:

MRQ: How to improve migrations to data mesh architectures by designing a guidance tool that supports deciding on data governance options by fulfilling criteria on correctness, understandability, utility, efficiency and effectiveness in order to make the design of data governance structures more efficient without reducing its effectiveness?

The sub-questions that support the main research question are covered in **Section 7.1** - **Section 7.5**. In **Section 7.6**, the answer to the main research question is formulated. In **Section 7.8**, the research findings and the conclusions are further discussed. The contributions are described in **Section 7.7** and the directions for future research are presented in **Section 7.9**.

7.1   Data mesh migrations

The first sub-question is:

RQ1: What does the migration to a data mesh architecture look like?

  (a) *What differentiates data mesh architectures from traditional data platform architectures?*

A data mesh architecture is a socio-technical approach to building decentralized data platform architectures. The data mesh approach divides the operational teams, central data teams and analytical teams into self-sovereign teams based on business domains which are supported by a self-serve infrastructure and governed through a federated governance model. The results of this study suggest that the differences between traditional data platform architectures and data mesh architecture can predominantly be found in the organizational structure of the platform. The data roles and responsibilities are diffused into the business domains to balance the workload that traditionally is accumulated in the central data team. This is expected to improve the scalability of the organization's data usage. In addition, the self-serve infrastructure provides business actors with the capability to fulfil their data responsibilities and the federated data governance allows domain teams

to have autonomy and ownership over their data, making them more accountable for its quality, security, and compliance.

(b) *What comprises a data mesh migration?*

Similar to other organizational transformations, data mesh migrations are likely to be organization specific. The available roadmaps suggest that multiple processes are running in parallel during data mesh migrations. The migration starts with a process in which the migration strategy is developed. After this, four parallel processes are carried out addressing: 1) the definition of the technological foundation, 2) the standardization of processes like the creation of data products, 3) the design of operating models for the data mesh, and 4) the communication of the migration to communicate successes and build the momentum needed to build the data mesh. The last process is the rollout process that accelerates the adoption of data products in the data mesh.

(c) *What challenges occur in data mesh migrations?*

Data mesh migrations bring predominantly organizational challenges. First and foremost, this study found that designing an effective DGS for organizations that try to migrate towards a data mesh architecture is challenging. The reason for this is the significant difference between the traditional DGSs currently implemented in organisations and the DGS proposed for a data mesh. Directly migrating to a completely different DGS disrupts current workflows and creates the risk of people not correctly performing their data governance tasks which results in inaccurate and unreliable data, data silos, compliance and legal risks, lack of accountability, and inefficient data management processes. Therefore, enterprise architects need to design effective DGSs that align with the goals of data mesh migrations while considering the characteristics of the current DGS. However, the practical resources normally used as guidance in DGS design do not address the data mesh principles and the available resources on the DGS design for data mesh are not directly applicable in practice. This lack of guidance decreases the efficiency of the DGS design process, resulting in elevated costs.

Furthermore, data mesh migrations require a cultural change, which is difficult to realize in large organizations for which these migrations could be most effective. Lastly, the data mesh architecture demands that business actors possess the necessary skills to manage data effectively, which is often lacking. Developing these capabilities requires both motivation from the actors themselves and additional resources.

7.2 Data governance in data mesh architectures

The second sub-question is:

RQ2: How is the data governance in a data mesh structured?

(a) *What is data governance?*

Data governance is the exercise of authority, control, and shared decision-making over the management of data assets through data policies, data standards and procedures. Data governance roles involve actors who are responsible for data quality, integrity, ownership, and custodianship. These roles collaborate to establish data governance frameworks, develop policies, and oversee data governance activities. According to Al-Ruithe et al. (2019), effective data governance can help organizations to create a clear mission, reduce costs, increase effectiveness, ensure transparency, increase confidence in

using the organizational data, establish accountabilities, maintain scope and focus, and define measurable successes.

(b) *What conceptual components constitute data governance?*

There are multiple ways to describe the data governance of an organization. One of these ways is to define the structure based on the important four components that function as building blocks of data governance. The first component consists of the antecedents that precede or predict the adoption of data governance practices. These antecedents can be internal or external. The second component consists of the governance mechanisms that help to plan and control data management activities. These mechanisms can be structural, relational or procedural. The third component consists of the organizational, domain and data scopes that influence the data governance initiatives. The final component consists of the consequences resulting from implementing the data governance initiatives. These consequences can be intermediate performance effects or the management of potential risks.

(c) *What are the important components of data governance in a data mesh architecture?*

The findings of this study suggest that the DGS in a data mesh is an intra-organizational structure with both centralized and decentralized teams, roles and responsibilities. The exact roles and responsibilities definitions can differ from team to team and from organization to organization depending on their internal antecedents. In general, the following three teams build up the DGS of a data mesh: 1) the global governance team, which is responsible for issuing and defining the organization-wide standards and policies, 2) the data platform team which is responsible for the implementation of the computational governance and the realization of interoperability in the infrastructure, and 3) the domain teams which are responsible for fulfilling the domain-specific practises and ensuring that their practices adhere to global standards. The domain teams consist of data product teams which are responsible for producing, consuming, maintaining and tracking the performance of data products.

7.3 Data governance structure design

The third sub-question is:

RQ3: How are data governance structures designed?

(a) *What are the key processes for designing the structure of organizational data governance?*

The design of a DGS is expected to change depending on the organization and its unique objectives and goals. However, there are some common guidelines and ideas that can be used to develop an effective framework. The design of an effective DGS requires information from the organization about the organizational goals, strategy, culture and structure. This information is acquired during interviews or workshops. This information is used collaboratively with theoretical resources, practical resources and experiences from enterprise architects to define the new roles, responsibilities and team structures that fit the organization. This is likely to be an iterative process with input from managers and data governance actors.

(b) *What problems occur during the data governance design process?*

The data governance design process faces challenges due to discrepancies between theoretical and practical implementations. Theoretical resources may not directly apply in practice, and practical resources often lack consideration for data mesh-specific roles and responsibilities. As a result, enterprise architects lack understandable resources for important design decisions, leading to inefficient discussions with stakeholders and the risk of duplication of work. EA consultants formulate client-specific design decisions and struggle with formulating configurations and evaluation opportunities, resulting in additional time spent on these tasks. Furthermore, they lack insight into the dependencies between configurations, which could result in spending additional time reevaluating earlier made decisions.

(c) *Which tools are currently used as guidance during the data governance design process?*

The tools currently used during the DGS design process are primarily the Data Management Book Of Knowledge (DAMA-DMBOK) [39] which describes the DGS design process for traditional DGSs and "Data mesh: Delivering data-driven value at scale" written by Dehghani (2019) which describes the data mesh approach.

7.4 Guidance tool development

The fourth sub-question is:

RQ4: How to develop design a guidance tool for data governance structure design in a data mesh?

(a) *What are the goals and requirements of the guidance tool?*

The goal of the tool is to improve the efficiency of the design process of the DGS of the organizations by reducing discussion time with stakeholders from the organization and by reducing the design time of the DGS. The treatment design interviews resulted in requirements based on the correctness, understandability, and utility of the tool and on the expected influence on the efficiency and effectiveness of the data governance design process. A tool that fulfils the requirements is expected to contribute to the goals of the stakeholders of having an efficient and effective data governance design process.

(b) *Which method can be used to design and develop the guidance tool?*

There are no methodologies, methods or guidelines found for the design of a tool that contributes to the earlier-mentioned goal. Because of the characteristics of the problem, design science research is chosen as the methodology to follow during the design and development of the tool. The guidance tool development methodology combines multiple methods from related studies. The book introducing the data mesh approach is extensively analysed to develop a data governance structure of a data mesh according to the theory. This structure is used as a framework in case studies to identify differences between theory and practice. The choices made to deviate from theory are defined as the design decisions made in practice. The design decisions with the resulting organizational configurations are compared between cases to generalize similar design decisions. The resulting design decisions present the theoretical configuration and one or more practical configurations.

(c) *Which method can be used to validate the guidance tool?*

The first version of the guidance tool is validated by presenting the tool to experts and discussing the correctness of its content. Thereafter, the guidance tool is presented to experts to extract their opinions about its fulfilment of the requirements regarding understandability, utility, and the expected influence on the efficiency and effectiveness of the data governance design process. The validation study leads to refinement opportunities which are used to improve the tool. The guidance tool can be empirically validated by using the tool in practice and following a technical action research methodology to quantify the tool's influence on the efficiency and effectiveness of the DGS design process.

7.5 Data governance structure design guidance tool

The fifth sub-question is:

RQ5: What constitutes a guidance tool for data governance structure design?

(a) *What high-level components does the tool need to possess?*

The intended users of the guidance tool are enterprise architects involved in the DGS design process. To support them, the guidance tool is presented as a slide deck consisting of introduction slides and six design decision slides. The introduction slides contain information about the way the tool can be used during the data governance design process, the definitions of roles and influential factors needed to use the tool and an interactive overview of the design decisions that can be used to navigate within the tool during discussions with stakeholders.

The slides presenting design decisions contain a short explanation of the design decision for which the tool provides decision support, explanations of the organizational configurations that can be chosen, visualisations of these configurations, advantages and disadvantages of these configurations and the factors influencing the design decisions.

(b) *Which design decisions for data governance structures in data mesh architectures can be distinguished?*

The design decisions that are found during this study affect the federated governance team, the domain teams and a potential enabling team. These decisions are the presence of a steering committee, the amount of federated data governance councils, the presence of a data domain owner role, the structure of domain representatives, the allocation of the data product developers and the presence of an enabling team. While there are presumably additional design decisions to consider, the research findings provided only adequate support for these design decisions.

(c) *Which organizational configurations for data governance structures in data mesh architectures can be distinguished?*

The tool offers two or three organizational configurations for each design decision. These configurations are specific arrangements of teams, roles and responsibilities that are found in theory or practice. The experts confirmed that all configurations are practically relevant and can be used to design the data governance structure. The findings provide sufficient support for these configurations, although further exploration may uncover additional options.

(d) *How can the tool be used during the data governance structure design process?*

Enterprise architects can use the cool as support during the formulation of design decisions and configurations and for the overall design of the data governance structure. Additionally, enterprise architects can use the tool to structure their discussions with stakeholders and introduce a structured decision-making process. Furthermore, EA consultants can use this tool to formulate recommendations for their clients.

7.6 Main research question

The main research question is:

How to improve migrations to data mesh architectures by designing a guidance tool that supports deciding on data governance options by fulfilling criteria on correctness, understandability, utility, efficiency and effectiveness in order to make the design of data governance structures more efficient without reducing its effectiveness?

The research findings provide conclusive evidence that the methodology followed during this research can effectively be used to design and develop a guidance tool that is expected to improve the efficiency of the data governance structure design process for data mesh architectures without reducing its effectiveness. The validation process confirmed that the guidance tool fulfils the correctness, understandability, and utility requirements, making it a valuable resource for enterprise architects in designing data governance structures. Moreover, the tool is expected to provide structure to discussions with stakeholders by presenting crucial design decisions and their influential factors, along with organizational configurations and their advantages and disadvantages. The interviewed EA consultants and experts expect that the tool will serve as a valuable starting point for data governance structure design, ultimately enhancing the efficiency and effectiveness of the design process. Given the significant role of this process in data mesh migrations, these improvements hold the potential to enhance the overall success of data mesh migration.

7.7 Contributions

7.7.1 Practical contribution

First and foremost, The guidance tool presents design decisions and configurations for practitioners to use during the organization design process. It can serve as a knowledge base for enterprise architects to design the structural mechanisms of data governance within a data mesh and support discussions with stakeholders. This tool complements existing data governance design resources like DAMA-DMBOK or Dehghani (2019) by providing applicable configurations and helps enterprise architects guide discussions, use configurations to assess impacts on DGSs, and evaluate design decisions based on its pros, cons, and influential factors.

Furthermore, the guidance tool presents general design decisions and organizational configurations which makes it usable by EA consultants for formulating recommendations for multiple clients. Increasing the efficiency of this process reduces its costs and could create a competitive advantage.

7.7.2 Scientific contribution

The present study attempts to address multiple gaps and in doing so makes important contributions. Firstly, it extends the existing research on data mesh by conducting the first

literature reviews on data mesh challenges and important data governance components specific to a data mesh context. Additionally, it demonstrates the applicability of the data governance framework of Abraham (2019) in a federated governance context, which is considered valuable given the limited research on federated data governance.

Secondly, This study provides the first thorough investigation of data governance in a theoretical data mesh by developing a theoretical DGS. This structure can be used as a framework for the assessment of practical data mesh implementations facilitating the identification of differences between DGSs and learning from these findings.

Third, the research follows a new methodology that uses techniques from different research domains, such as design science, design options, organizational structure, and data governance design. The demonstration of this method emphasises its potential relevance in other contexts and recommends its use in future research.

The scientific community is increasingly recognizing the importance of data governance, and the rise of data mesh has made it more complex and critical. This research contributed to the scientific knowledge base by researching how the fields of data governance and data mesh interact with each other.

7.8 Discussion

This research followed a design science research methodology to develop the guidance tool and in this section, this methodology and the resulting guidance tool are discussed.

7.8.1 Reflection on methodology

The main foundation for the methodology is the design science research methodology of Wieringa (2014). This methodology consists of an Implementation evaluation/ problem investigation, the design & development of the treatment and the validation of the treatment. These steps of the methodology are reflected in these sections.

Problem investigation method

Through a thorough problem investigation process, this research identified significant problems with the current DGS design process. Even though the problems with the currently available treatments are investigated, a systematic assessment of these treatments is not part of this research. This is not perceived as disruptive to the research, however, it could contribute to better determining the added value of the guidance tool.

Treatment design and development method

The second phase is the treatment design and development phase. The first process of this phase is the analysis of the data mesh book of Dehghani (2019) because this is currently the most exhaustive source on the data mesh approach. The analysis of this book is done systematically through axial and open coding and validated with a data mesh expert, however wrong interpretations and research biases should be taken into account while interpreting the results.

The second process within the treatment design and development phase is the creation of the practical DGSs. This consists of the assessment of two cases. During this assessment,

the theoretical DGS is used as a framework to create the case-specific DGS supplemented by semi-structured interviews with the case representatives and provided resources by the case representatives. This is done in collaboration with the case representatives to prevent misinterpretations and improve its validity. Due to the complexity and size of the investigated organizations, the DGSs are conceptual models. These structures are created to compare these to theory and extract design decisions and configurations and therefore should not be interpreted with caution.

The third process of the treatment design and development phase is the extraction of the design decisions and configurations. The differences between the theoretical DGS and the practical DGSs are addressed as design decisions that are made by enterprise architects to deviate from theory. The findings from both cases are complemented by findings from expert interviews. The design decisions and configurations from the cases and experts are compared and the similarity is determined based on the influenced team and the definition of roles and responsibilities. Similar design decisions from the cases are merged. Because of possible misinterpretations, the generalized design decisions and configurations are validated based on their correctness and the experts and case representatives are given the opportunity to address incorrect generalizations. According to the experts, this resulted in important and useful design decisions. However, also these results should be interpreted with caution because of the small sample size (N=2), and because the tool is presumably not comprehensive. Other design choices or configurations not included in this study could be important for organizations. Also, if both cases deliberately chose to follow the theory, this is not identified as a design decision even though this could be of great importance for other organizations. Nevertheless, during this research, the correctness of the design decisions is assessed and the participants did not have additional design decisions or configurations that should be included to improve the completeness of the guidance tool.

Treatment validation method

The final phase of the design science research methodology is the validation and evaluation phase, where the guidance tool is validated based on its fulfilment of the requirements. In this stage, the guidance tool is presented to enterprise architects and evaluated based on its expected ability to support enterprise architects in making decisions about the design of DGSs.

The DGS design process is a lengthy and resource-heavy process that requires contributions from numerous stakeholders. Because of the available resources of this research, validating the tool through implementation in this process is perceived as infeasible during this research. Therefore, the validation is only based on expert opinions on a hypothetical interaction between the artifact and the context. A quantification of the influence of the guidance tool on the efficiency and effectiveness of the DGS design process could significantly strengthen the findings. In **Section 7.9**, multiple methods are presented that could be followed for this quantification.

Even though the lack of quantification is a limitation of this study and with that, the study does not provide strong empirical support/evidence, the findings from the expert interviews can be used to validate the treatment and assess its fulfilment of the requirements.

In conclusion, the design science research methodology is demonstrated as an effective methodology for the development of the guidance tool. The three stages of the

methodology, problem investigation, design and development, and validation and evaluation provided a structured approach to problem-solving. The design and development phase resulted in the creation of a guidance tool that was informed by existing literature and feedback from stakeholders. The validation and evaluation phase enabled the assessment of the expected influence of the guidance tool on the efficiency and effectiveness of the DGS design process. The methodology followed during this research can be used as a framework for future research in this area, enabling the development of other artifacts that support decision-making in organizational design.

7.8.2 Reflection on guidance tool

The guidance tool developed in this study is validated by eight enterprise architects with specializations in data governance, data management, operating models and data mesh. The evaluation consisted of two rounds with the tool refinement process and showed that the guidance tool is expected to be useful during the initial steps of the DGS design process. However, a few debate issues emerge when the model is examined critically.

Generalizeability

The first and most important discussion point of the guidance tool is its generalizability. The tool provides design options, advantages, and disadvantages resulting from the case studies and expert interviews, but how this resonates in practice is highly dependent on the organization. In organizations with specific needs, this could mean that the tool may not be directly applicable. Enterprise architects will need to adapt the design decisions to fit their specific needs. This limitation highlights the need for organizations to conduct a thorough analysis of their goals, objectives and current DGS. The challenge of generalizability is not unique to this tool, but it is a common challenge in tools that support decision-making.

Comprehensibility

Another limitation of the tool is its comprehensibility. The tool is based on a literature study and input from experts, but there may be additional factors that should be considered. Due to the complexity of organisational structures, there are often many interrelated factors that need to be considered. It can be challenging to identify all possible design options and their configurations. Therefore, the tool's design options are expected to be not comprehensive enough and organizations still need to supplement the tool with their own research and analysis. The impact of this limitation is reduced by asking interviewees whether they are missing design decisions in the tool and by being open-minded during interviews.

Usability

The final important point of reflection is its direct usability. The tool is expected to be usable during the formulation of design decisions, configurations, and evaluation opportunities and during the preparation of sessions with stakeholders as the design decisions present important discussion points that should be addressed during these sessions.

In addition, the tool only discusses the data mesh-specific roles and that does not

address the traditional data governance roles. The users of the tool rely on other resources or expertise to determine how to integrate the traditional data governance roles of the organization into the organizational configurations provided by the tool.

In conclusion, the guidance tool contributes to the field of data governance, but it is important to acknowledge its limitations. The tool provides a structured approach to decision-making, but its generalizability is limited, there may be additional factors to consider, and it does not consider traditional data governance roles. Future research could address these limitations and further refine the decision-making process.

7.9 Future research

There are several promising directions for future research related to the extensive study on data governance in a data mesh and the guidance tool developed in this study.

First and foremost, future research could quantify the influence of the guidance tool on the efficiency and effectiveness of the DGS design process. This quantification could be realised in multiple ways.

The first method is by running through the process twice, once without the guidance tool and once with the guidance tool. The differences in efficiency between the processes should be traced back to the guidance tool. For this research, the dependencies between the processes should be taken into account.

Another way to quantify this influence is by selecting data mesh migrations where the guidance tool was not utilized and comparing their outcomes to similar migrations conducted with the support of the tool. During this comparison, both the resource spent on the design of the DGS and on the discussions with stakeholders should be tracked. These comparisons could give insights into how efficiency and effectiveness are influenced. For this research, an effort should be made to validate the validity of the findings as the design processes of different organizations are complex and dependent on multiple organizational factors which could hamper the generalizability. Therefore, it can be difficult to attribute the impact solely to the usage of the guidance tool, as other factors, such as organizational culture, leadership, and resource availability, can also contribute to a difference in the efficiency and effectiveness of the process.

Therefore, the most promising method to empirically validate the guidance tool is by conducting Technical Action Research. This is an attempt to scale up treatment to conditions of practice by actually using it in a particular problem [81]. For this, the methodology presented by Wieringa (2014) can be used. The technical action research should answer the question: How effective is the guidance tool in improving the efficiency of the DGS design process in data mesh architectures? To answer this question, a client organization should be acquired and a treatment plan should be agreed upon. Specific measurements should be designed (for example based on the metrics for efficiency and effectiveness of data governance design proposed by [51]. These measurements should be taken during the DGS design process and the collected data should be analysed in detail to further validate its contribution to stakeholders' goals. These findings should be used to draw conclusions, highlighting the strengths and limitations of the use of the guidance tool in practice.

Another potential area for future research is to address the issue of components in the case structures that are similar to theory but not considered design decisions. This could

be accomplished by extending the tool with more cases to identify additional alternative designs and validate these design decisions. Additionally, as the field of data mesh architectures continues to grow, including design decisions from the literature could help identify more potential design decisions.

Further research could also be conducted on specific design decisions to identify more influential factors and configurations. For example, a decision regarding the hierarchy of federated data governance councils could be explored in more depth. Additionally, researching the influence of design decisions on each other could provide valuable insights into how different decisions interact and impact the overall design.

Furthermore, future research could focus on developing a more comprehensive tool that includes guidance on how to position traditional data governance roles within the organizational structure. This could involve providing templates or frameworks for defining and assigning roles and responsibilities, as well as guidelines for effective communication and collaboration between different parts of the organization.

Finally, given that data mesh is still in its early stages and rapidly evolving, there is a need for future research to update this study with additional findings from both theory and practice. As more organizations begin to adopt the data mesh paradigm, there will be opportunities to gather new insights and data on the effectiveness of different design decisions and their impact on organizational outcomes. Furthermore, ongoing research by multiple researchers will continue to shed light on key issues and challenges in implementing data mesh, which can inform future updates to this study and ensure its relevance and usefulness for practitioners in the field.

Bibliography

[1] Architecture and functions in a data mesh — data mesh on google cloud.

[2] Data mesh in practice: How europe's leading online platform for fashion goes beyond the data lake, Sep 2020.

[3] Rene Abraham, Johannes Schneider, and Jan Vom Brocke. Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, 49:424–438, 2019.

[4] Majid Al-Ruithe, Elhadj Benkhelifa, and Khawar Hameed. A systematic literature review of data governance and cloud data governance. *Personal and Ubiquitous Computing*, 23:839–859, 2019.

[5] Ian F Alexander. A taxonomy of stakeholders: Human roles in system development. *International Journal of Technology and Human Interaction (IJTHI)*, 1(1):23–59, 2005.

[6] Hilary Arksey and Lisa O'Malley. Scoping studifes: towards a methodological framework. *International journal of social research methodology*, 8(1):19–32, 2005.

[7] Michael Armbrust, Ali Ghodsi, Reynold Xin, and Matei Zaharia. Lakehouse: a new generation of open platforms that unify data warehousing and advanced analytics. In *Proceedings of CIDR*, 2021.

[8] William B Askren and Kenneth D Korkan. Design option decision tree: A method for schematic analysis of a design problem and integration of human factors data. In *Proceedings of the Human Factors Society Annual Meeting*, volume 18, pages 368–375. SAGE Publications Sage CA: Los Angeles, CA, 1974.

[9] Olivia Benfeldt Nielsen. A comprehensive review of data governance literature. 2017.

[10] Ashish Bijawat. Data warehouse vs. data lake vs. data lakehouse vs. data mesh, Jun 2022.

[11] Jan Bode, Niklas Kühl, Dominik Kreuzberger, and Sebastian Hirschl. Data mesh: Motivational factors, challenges, and best practices. *arXiv preprint arXiv:2302.01713*, 2023.

[12] Lars Michael Bollweg. Development of a responsive operating model. In *Data Governance for Managers*, pages 81–100. Springer, 2022.

[13] Michael Brackett and Production Susan Earley. The dama guide to the data management body of knowledge (dama-dmbok guide). 2009.

[14] Susanne Braun, Annette Bieniusa, and Frank Elberzhager. Advanced domain-driven design for consistency in distributed data-intensive systems. In *Proceedings of the 8th Workshop on Principles and Practice of Consistency for Distributed Data*, pages 1–12, 2021.

[15] Eric Broda. Towards a practical data mesh roadmap, Nov 2022.

[16] Susanna ES Campher. Semantic metadata requirements for data warehousing from a dimensional modeling perspective. In *ICEIS (1)*, pages 129–136, 2022.

[17] Amber G Candela. Exploring the function of member checking. *The qualitative report*, 24(3):619–628, 2019.

[18] Silvia Anna Chiusano. Data mesh: The newest paradigm shift for a distributed architecture in the data world and its application. *Ingegneria Informatica*, 2021.

[19] Edward Curry. The big data value chain: definitions, concepts, and theoretical approaches. In *New horizons for a data-driven economy*, pages 29–37. Springer, Cham, 2016.

[20] EcoStruxureTM Data. Faster. 2022.

[21] Willemijn de Boer. A meshed up data architecture design. 2022.

[22] Andrea De Mauro, Marco Greco, and Michele Grimaldi. A formal definition of big data based on its essential features. *Library review*, 2016.

[23] Zhamak Dehghani. *Data Mesh: Delivering data-driven value at scale*. O'Reilly Media, 2022.

[24] Zhamak Dehghani. Data mesh principles and logical architecture, 2022.

[25] Ivan R Diamond, Robert C Grant, Brian M Feldman, Paul B Pencharz, Simon C Ling, Aideen M Moore, and Paul W Wales. Defining consensus: a systematic review recommends methodologic criteria for reporting of delphi studies. *Journal of clinical epidemiology*, 67(4):401–409, 2014.

[26] Juraj Dončević, Krešimir Fertalj, Mario Brčić, and Mihael Kovač. Mask-mediator-wrapper architecture as a data mesh driver. *arXiv preprint arXiv:2209.04661*, 2022.

[27] Robert Duncan. What is the right organization structure? decision tree analysis provides the answer. *Organizational dynamics*, 7(3):59–80, 1979.

[28] Kathleen M Eisenhardt. Building theories from case study research. *Academy of management review*, 14(4):532–550, 1989.

[29] Anton Engström. Towards agile data engineering for small scale teams. *LU-CS-EX*, 2020.

[30] Arlene Fink. *Conducting research literature reviews: From the internet to paper*. Sage publications, 2019.

[31] Chris Ford. *Data Mesh and Governance*. Thoughtworks, Aug 2021.

[32] Ammara Gafoor, Ian Murdoch, and Kiran Prakash. Data mesh in practice: Product thinking and development, 2022.

[33] Heath Gascoigne. The business transformation playbook - preview, Jan 2019.

[34] Abel Abraham Goedegebuure. *Data Mesh: Systematic Gray Literature Study, Reference Architecture, and Cloud-based Instantiation at ASML*. PhD thesis, TILBURG UNIVERSITY, 2022.

[35] Shivakumar R Goniwada. Cloud native architecture and design. 2021.

[36] Simo Hokkanen. Utilization of data mesh framework as a part of organization's data management. Master's thesis, It"a-Suomen yliopisto, 2021.

[37] Y Hooshmand, J Resch, P Wischnewski, and P Patil. From a monolithic plm landscape to a federated domain and data mesh. *Proceedings of the Design Society*, 2:713–722, 2022.

[38] Cuelogic Insights. Data automation for healthcare, Jul 2021.

[39] Dama International. *DAMA-DMBOK: data management body of knowledge*. Technics Publications, LLC, 2017.

[40] J Mike Jacka and Paulette J Keller. *Business process mapping: improving customer satisfaction*. John Wiley & Sons, 2009.

[41] Annika Jacobsen, Ricardo de Miranda Azevedo, Nick Juty, Dominique Batista, Simon Coles, Ronald Cornet, Mélanie Courtot, Mercè Crosas, Michel Dumontier, Chris T Evelo, et al. Fair principles: interpretations and implementation considerations, 2020.

[42] Nathalie E Janssen. The evolution of data storage architectures: Examining the value of the data lakehouse. Master's thesis, University of Twente, 2022.

[43] Divya Joshi, Sheetal Pratik, and Madhu Podila. Data governance in data mesh infrastructures: the saxo bank case study. Technical report, EasyChair, 2021.

[44] Pwint Phyu Khine and Zhao Shun Wang. Data lake: a new ideology in big data era. In *ITM web of conferences*, volume 17, page 03025. EDP Sciences, 2018.

[45] Barbara Kitchenham, O Pearl Brereton, David Budgen, Mark Turner, John Bailey, and Stephen Linkman. Systematic literature reviews in software engineering–a systematic literature review. *Information and software technology*, 51(1):7–15, 2009.

[46] Krish Krishnan. *Data warehousing in the age of big data*. Newnes, 2013.

[47] Petar Krivic, Pavle Skocir, Mario Kusek, and Gordan Jezic. Microservices as agents in iot systems. In *KES International Symposium on Agent and Multi-Agent Systems: Technologies and Applications*, pages 22–31. Springer, 2017.

[48] John Kutay. Data warehouse vs. data lake vs. data lakehouse: An overview of three cloud data storage patterns, May 2022.

[49] Anne Kwam and Maximilian Schroeck. Architecting an operating model - deloitte us.

[50] Anne Kwan and Maximilian Schroek. Architecting an operating model.

[51] John Ladley. *Data governance: How to design, deploy, and sustain an effective data governance program*. Academic Press, 2019.

[52] Elise Lakey. Top 5 benefits of a data warehouse for your data-driven organization, Jun 2022.

[53] Marc M Lankhorst, Henderik Alex Proper, and Henk Jonkers. The architecture of the archimate language. In *Enterprise, Business-Process and Information Systems Modeling: 10th International Workshop, BPMDS 2009, and 14th International Conference, EMMSAD 2009, held at CAiSE 2009, Amsterdam, The Netherlands, June 8-9, 2009. Proceedings*, pages 367–380. Springer, 2009.

[54] Harold J Leavitt. Applied organization change in industry: structural, technical and human approaches. 1964.

[55] Yair Levy and Timothy J Ellis. A systems approach to conduct an effective literature review in support of information systems research. *Informing Science*, 9, 2006.

[56] Ben Lorica and Ali Ghodsi. What is a lakehouse?, Jan 2020.

[57] Antti Loukiala, Juha-Pekka Joutsenlahti, Mikko Raatikainen, Tommi Mikkonen, and Timo Lehtonen. Migrating from a centralized data warehouse to a decentralized data platform architecture. In *International Conference on Product-Focused Software Process Improvement*, pages 36–48. Springer, 2021.

[58] Inês Machado, Carlos Costa, and Maribel Yasmina Santos. Data-driven information systems: the data mesh paradigm shift. 2021.

[59] Inês Araújo Machado, Carlos Costa, and Maribel Yasmina Santos. Data mesh: Concepts and principles of a paradigm shift in data architectures. *Procedia Computer Science*, 196:263–271, 2022. International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2021.

[60] James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers, et al. *Big data: The next frontier for innovation, competition, and productivity.* McKinsey Global Institute, 2011.

[61] Chitu Okoli and Kira Schabram. A guide to conducting a systematic literature review of information systems research. 2010.

[62] Guadalupe Ortiz, Juan Boubeta-Puig, Javier Criado, David Corral-Plaza, Alfonso Garcia-de Prado, Inmaculada Medina-Bulo, and Luis Iribarne. A microservice architecture for real-time iot data processing: A reusable web of things approach for smart ports. *Computer Standards & Interfaces*, 81:103604, 2022.

[63] Vinit Parida, Tobias C Larsson, Ola Isaksson, and Pejvak Oghazi. Towards open innovation practices in aerospace industry : challenges and opportunities. 2011.

[64] Mark Petticrew and Helen Roberts. *Systematic reviews in the social sciences: A practical guide.* John Wiley & Sons, 2008.

[65] Nikolai J Podlesny, Anne VDM Kayem, and Christoph Meinel. Cok: A survey of privacy challenges in relation to data meshes. In *International Conference on Database and Expert Systems Applications*, pages 85–102. Springer, 2022.

[66] Francisco Ponce, Gastón Márquez, and Hernán Astudillo. Migrating from monolithic architecture to microservices: A rapid review. In *2019 38th International Conference of the Chilean Computer Science Society (SCCC)*, pages 1–7. IEEE, 2019.

[67] PricewaterhouseCoopers. Global and industry frameworks for data governance, Dec 2019.

[68] Margy Ross and Ralph Kimball. *The data warehouse toolkit: the definitive guide to dimensional modeling.* John Wiley & Sons, 2013.

[69] Denise M Rousseau, Joshua Manning, and David Denyer. 11 evidence in management and organizational science: assembling the field's full weight of scientific knowledge through syntheses. *Academy of Management Annals*, 2(1):475–515, 2008.

[70] Matthew Skelton and Manuel Pais. *Team topologies: organizing business and technology teams for fast flow.* It Revolution, 2019.

[71] Nuno Soares, Paula Monteiro, Francisco J Duarte, and Ricardo J Machado. Reference models for intelligent cities: an aligned template. In *developing and monitoring smart environments for intelligent cities*, pages 28–60. IGI Global, 2021.

[72] Carl Spetzler, Hannah Winter, and Jennifer Meyer. *Decision quality: Value creation from better business decisions.* John Wiley & Sons, 2016.

[73] Gwen Thomas. The dgi data governance dgi data governance framework, Aug 2019.

[74] Unknown. What is a data platform?, 2020.

[75] Victoria Uren and John S Edwards. Technology readiness and the organizational journey towards ai adoption: An empirical study. *International Journal of Information Management*, 68:102588, 2023.

[76] Stefan van Duin. From data mess to a data mesh, Jul 2022.

[77] John Venable, Jan Pries-Heje, and Richard Baskerville. Feds: a framework for evaluation in design science research. *European journal of information systems*, 25:77–89, 2016.

[78] Kristin Weber, Boris Otto, and Hubert Österle. One size does not fit all—a contingency approach to data governance. *Journal of Data and Information Quality (JDIQ)*, 1(1):1–27, 2009.

[79] Jane Webster and Richard T Watson. Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly*, pages xiii–xxiii, 2002.

[80] Dave Wells. The path to modern data governance, Aug 2019.

[81] Roel Wieringa and Ayşe Moralı. Technical action research as a validation method in information systems design science. In *Design Science Research in Information Systems. Advances in Theory and Practice: 7th International Conference, DESRIST 2012, Las Vegas, NV, USA, May 14-15, 2012. Proceedings 7*, pages 220–238. Springer, 2012.

[82] Roel J Wieringa. *Design science methodology for information systems and software engineering.* Springer, 2014.

[83] Iris de Wildt. Framing the implementation of data governance platforms. Master's thesis, 2022.

[84] Nam-Hong Yim, Soung-Hie Kim, Hee-Woong Kim, and Kee-Young Kwahk. Knowledge based decision making on higher level strategic concerns: system dynamics approach. *Expert Systems with Applications*, 27(1):143–158, 2004.

[85] Michal Zasadzinski, Michael Theodoulou, Markus Thurner, and Kshitij Ranganath. The trip to the enterprise gourmet data product marketplace through a self-service data platform. *arXiv preprint arXiv:2107.13212*, 2021.

Appendix A

Guidance tool 2.0

FIGURE A.1: Guidance tool v2 slide 1: Cover

FIGURE A.2: Guidance tool v2 slide 2: Tool guide

FIGURE A.3: Guidance tool v2 slide 3: Components, legends and buttons
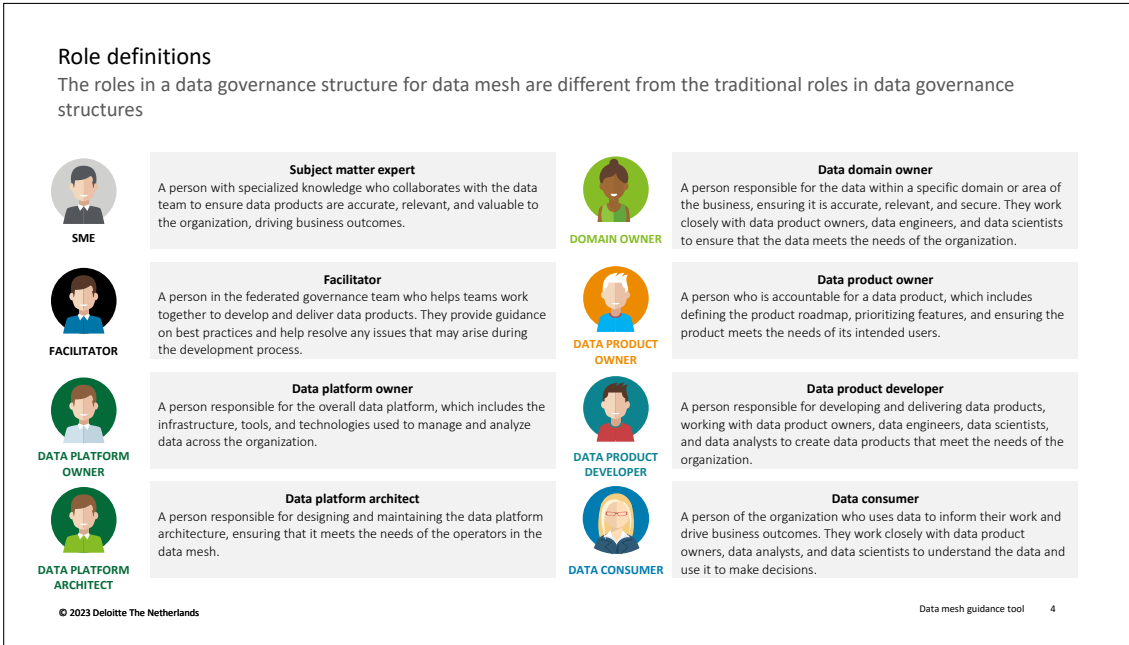
Figure A.4: Guidance tool v2 slide 4: Role definitions
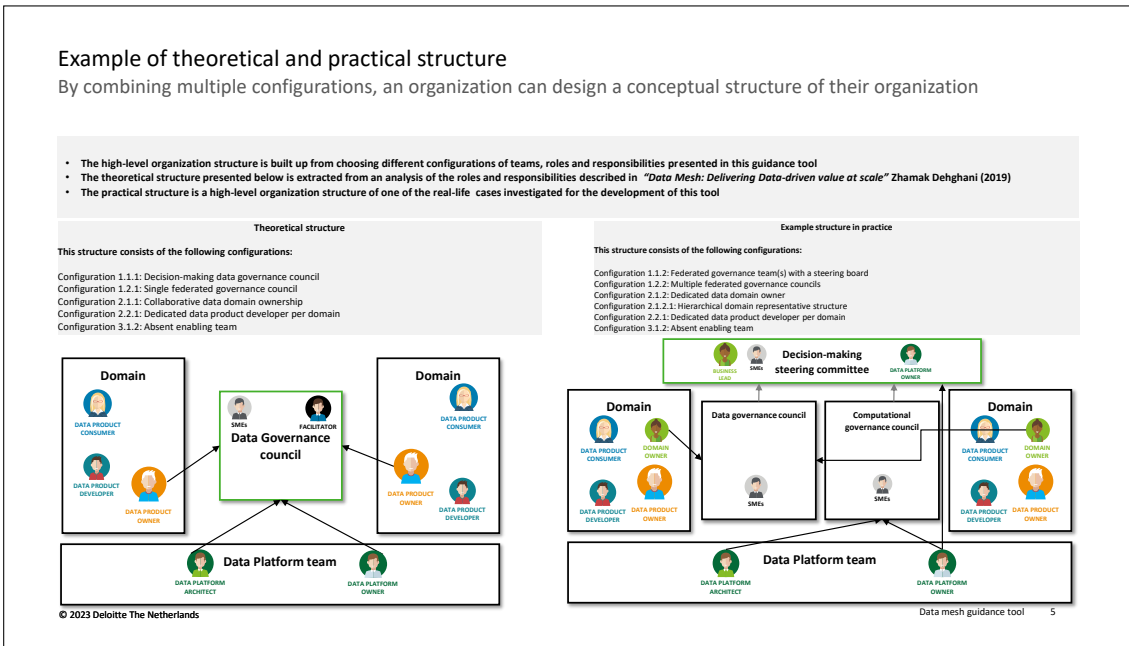
Figure A.5: Guidance tool v2 slide 5: theoretical and practical example

FIGURE A.6: Guidance tool v2 slide 6: Influential factor definitions

FIGURE A.7: Guidance tool v2 slide 7: Design option tree separator slide
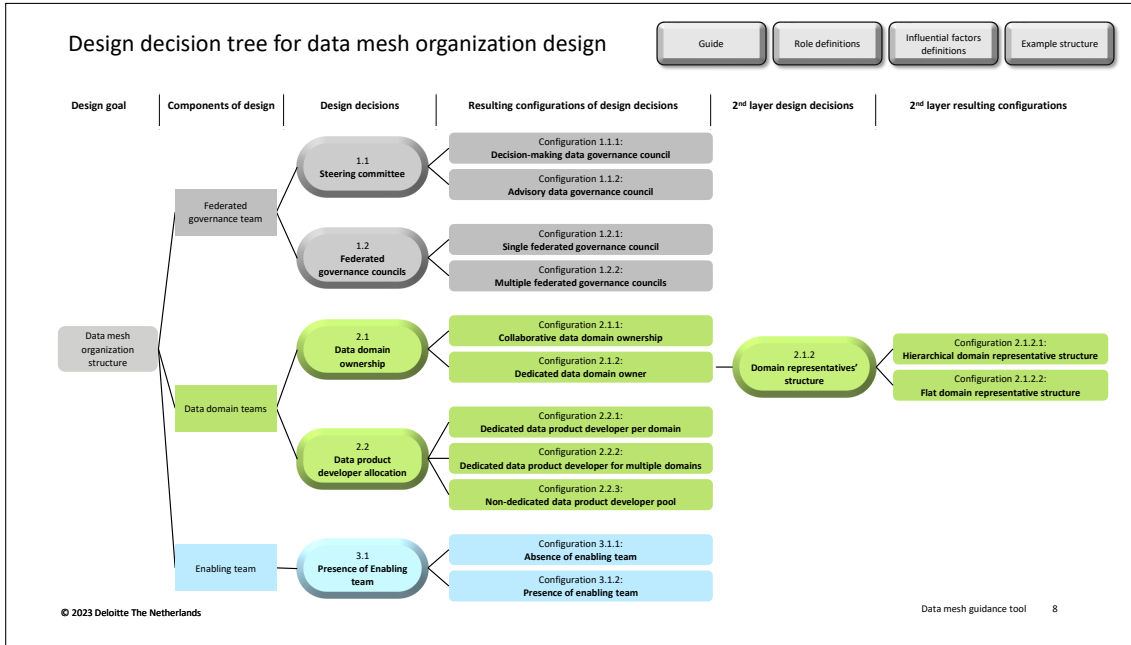
FIGURE A.8: Guidance tool v2 slide 8: Design option tree
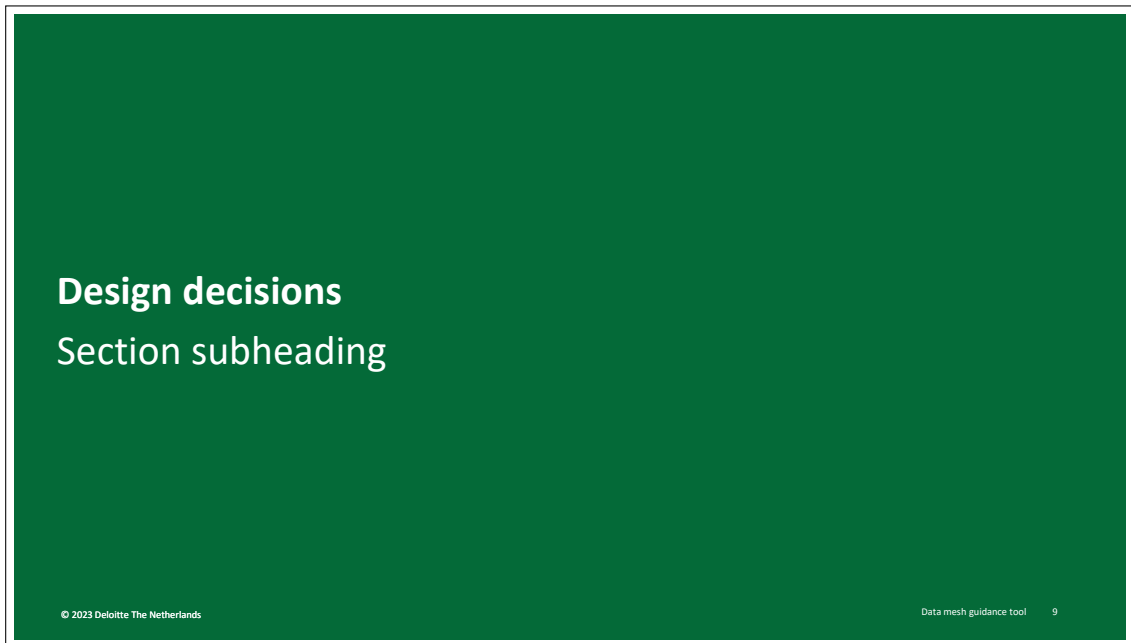
FIGURE A.9: Guidance tool v2 slide 9: Design decisions separator slide

FIGURE A.10: Guidance tool v2 slide 10: Steering committee

FIGURE A.11: Guidance tool v2 slide 11: Federated governance council

FIGURE A.12: Guidance tool v2 slide 12: Data domain ownership
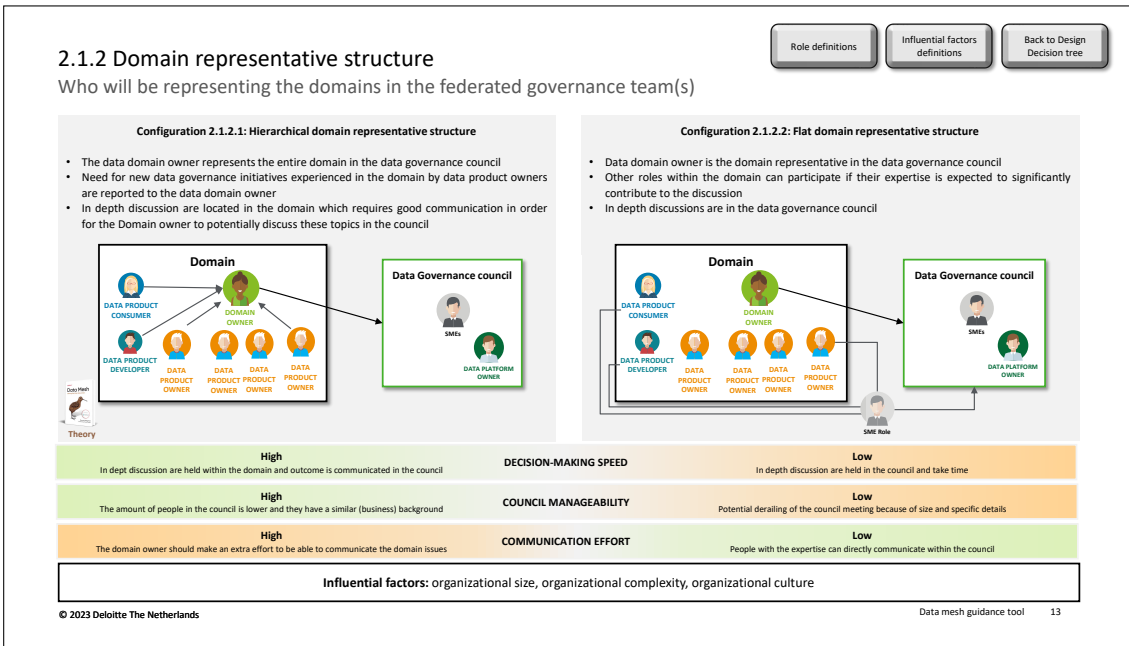
FIGURE A.13: Guidance tool v2 slide 13: Domain representative structure
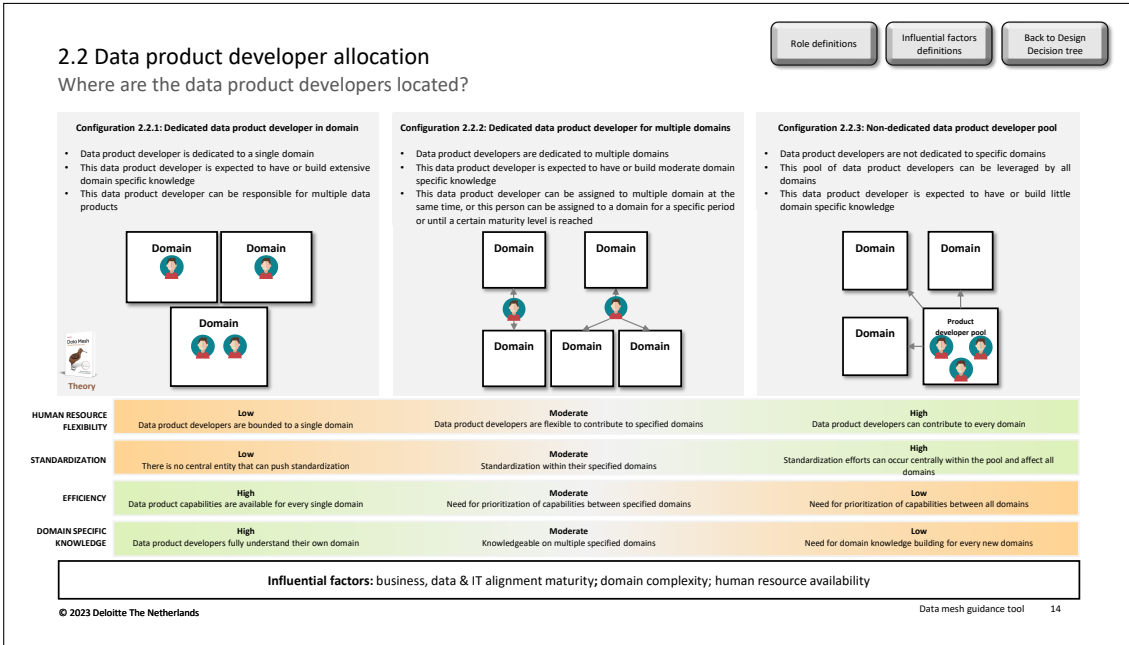
FIGURE A.14: Guidance tool v2 slide 14: Data product developer allocation
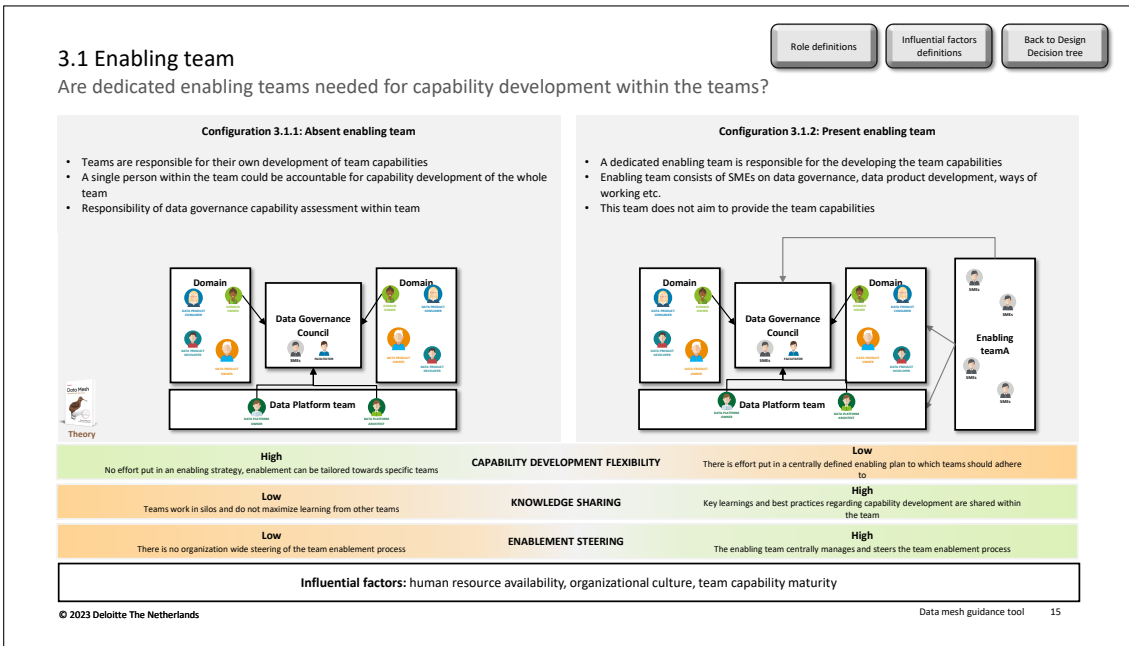
FIGURE A.15: Guidance tool v2 slide 15: Enabling team

Appendix B

Scoping study

B.1 Extensive scoping study methodology

B.1.1 Step 1: Identify the research question

The main research question of this scoping study is formulated as follows:

What challenges occur in data mesh migrations?

B.1.2 Step2: Identify relevant research studies

To be able to answer a proposed research question, Arksey advises identifying relevant research studies. This identification process is done through the creation of a search protocol. This protocol is presented in **Figure B.1**. Due to the novelty of data mesh implementation and the scarcity of literature on this topic, a wide approach is maintained to generate a breadth of coverage on this topic. Different sources are used to find these research studies. The electronic databases that are queried are Google Scholar, Science Direct and Semantic Scholar. Also, the reference lists of the selected articles from these databases are reviewed and potentially valuable articles are included in this review.

The articles found in electronic databases are queried according to the iterative search strategy described in **Figure B.1** based on search terms and filters. Synonyms for the search terms are also included in the query. The final iteration uses the following search terms:

Search term:

Data mesh AND "challenges OR bottlenecks" AND organization AND NOT "2d" AND NOT "3d" "3d mesh" AND NOT "EEG-based" AND (implementation OR migration OR transformation)

Filters:

- Publication period = 1-1-2019 - 1-10-2022
- Language = English OR Dutch.
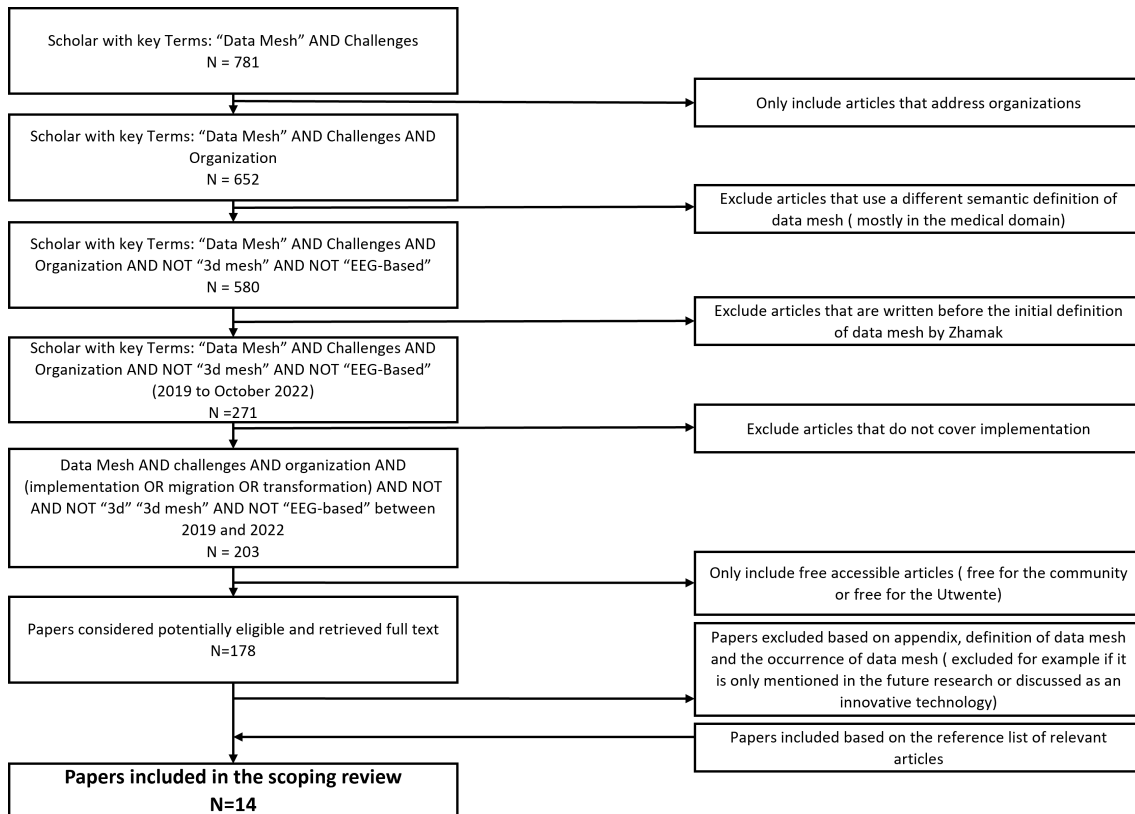
FIGURE B.1: Study selection for systematic scoping study

- Type = Article

This publication period is chosen because 2019 is the year that the book of Dehghani [23] is published in which the data mesh concept used in this research is first described. English and Dutch are chosen because these are the two languages understandable to all of the stakeholders of this research. Only free accessible articles or articles that are accessible through the license of the University of Twente, are included in this study. The combination of the search terms and the filters resulted in a total of 178 studies.

B.1.3 Step 3: Selecting studies

Of these 178 studies, the abstract is read. However, according to Badger et al. (2000), abstracts alone cannot be assumed to be representative of the full article that follows or to capture the full scope of an article. Therefore, the content of the article is also scanned for the used definition of data mesh and the coverage of the data mesh implementation challenges. Based on the abstract, the context, and the used definitions, exclusion and inclusion criteria are created to further filter articles and to endure consistency in decision-making [6]. After following this selection procedure, a total of 14 papers are selected for data extraction. The search and selection procedure is also presented in **Figure B.1**

B.1.4 Step 4: Charting the data

The selected papers are synthesized and the qualitative data is interpreted by sifting, charting and sorting material according to key issues, themes and discussed challenges. Arksey states that "a descriptive-analytical method within the narrative tradition, which

involves applying a common analytical framework to all the primary research reports and collecting standard information on each study, stands more chance of being useful" [6]. Therefore, an analytical framework is developed for this chartering process.

B.1.5   Step 5: Collating, summarizing and reporting the results

With the help of the analytical framework, the challenges mentioned in the articles are extracted. A classification scheme is developed to structure this process. This classification scheme is based on the people, process and technology (PPT) framework [54]. This framework is previously successfully used for the classification of different challenges [63]. Due to the Data-centric problem context, also the PPTD (People, Process, Technology & Data) framework [75], which extends the PPT framework, is considered. However, the PPT framework is chosen because of its broader adoption in literature and the organizational origin of the data mesh approach in which the content of the data itself is scarcely considered. Predominantly, only the data product is considered which is defined as a technological object [24].

The exact statements that describe challenges are classified as people process or technology challenges. The PPT classification is considered the Tier 1 challenge. After this classification, the statements that describe identical challenges are grouped together in Tier 3 challenges. Finally, tier 3 challenges that address the same kind of challenges are grouped together with Tier 2 challenges. The classification and the further grouping of similar challenges are done through an iterative process and evaluated during multiple consultation rounds. These consultation rounds followed the Delphi study methodology [25] with the contribution of experts in data management and enterprise architecture with knowledge of classification schemes and data mesh principles. This iterative process continued until there was consensus on both the classification scheme and the classification. The classified challenges are presented in **Section 3.3.2** of this research.

B.2   Analytical framework for scoping study

TABLE B.1: Analytical framework for scoping study

Article information	Author(s), Title, Journal BibTex	Divya Joshi, Sheetal Pratik and Madhu Podila ,Data Governance in Data Mesh Infrastructures: The Saxo Bank Case Study
REFERENCE & PURPOSE	Purpose	0.8 Providing an hollistic view of data governance that is a synthesis of academic and practitioner viewpoints and conclude by giving an example of a pilot case study where authors worked on tech and cultural interventions to address the data governance challenges.
SUBJECTS	Subject Characteristics	Data Governance, FEDERATED DATA GOVERNANCE, Financial Sector
	Sample Design	Use case SAXO BANK
	Definition of Data Mesh	Defined data mesh according to the principles of Zhamak Dehghani
DATA	Researched Enterprise	BANK SAXO BANK
AIM	AIM	provide a holistic view of data governance that is a synthesis of academic and practitioner viewpoints and conclude by giving an example of a pilot case study (Saxo Bank) where authors worked on tech and cultural interventions to address the data governance challenges.
CHALLENGES	Challenges in short	Governance of the data mesh, Technology& Business allignment, Tool Selection, creating data culture & change management.
	Challenges	1. Agreement and compliance to the governance in a dynamic policy environment is a moving target. 2. getting business and technology on the same page is very important because the need for self-serve tools and reorganisation is best driven by business, not tech. 3. seamless integration of tooling with the existing data ecosystem is tough. Available tools may not map exactly to the organisational needs. Tool selection. 4.creating data culture and change management is a humongous task. Culture and change management are usually the hardest part in any equation. and this requires partnerships and buy ins from the top management.
	Challenges solutions	Implementation of a data catologue, business glossaries. For Data Quality implementation, the DSL was defined to facilitate the data producers to create the quality rules and pushed these to the respective domain repository.Saxo Bank is committed to assimilating data governance across the organization. Milestones in this journey included formulation of DG committee, defining roles and responsibilities, and creation of domain ownership model (domain teams are responsible for owning and managing the data). s Consumer, user is able to discover the data products, view the metadata, view data lineage to understand the data flow better, view quality snapshot along with metadata and able to access to quality notifications/alerts for the datasets of interest As a data product owner, a user is able to link data products/elements to the business terms, that encourages the consumptionand avoids data inconsistencies, define information classification at data element level, and profile the data before onboarding to Kafka .Net-net, Data Workbench platform of Saxo enables the domain teams to publish transparent and trustworthy data for the consumers and will pave the way to onboarding of new partners with increased confidence. This is a big milestone in the journey of open banking.
CONCLUSION	Conclusions	"For attaining intelligent federated data governance, an organisation must solve some crucial challenges that are a mix of technical and social, making it so hard."

B.3 Classification scheme for scoping study

TABLE B.2: Classification scheme with challenges and mentions in literature

#	Tier 1	%	#	Tier 2	%	#	Tier 3	%	Joshi, 2021	Machado, 2021	Machado, 2022	Donvcevicy, 2022	Podlesny, 2022	Hooshmand, 2022	de Boer, 2022	Zasadzinski, 2021	Hokkanen, 2021	Dehghani, 2022	Vestues, 2022	Martin, 2021	Southwood, 2022	Loukiala, 2021
C 1	People	64%	C 1.1	Data literacy	50%	C 1.1.1	Increasing data experience, learning and knowledge sharing	43%						x	x		x			x	x	x
						C 1.1.2	Creating a data culture	7%	x													
			C 1.2	Incentive	50%	C 1.2.1	Specifying value of Implementation	29%		x		x					x			x		
						C 1.2.2	Getting support from top management	21%	x						x		x					
						C 1.2.3	Motivating source teams to develop data products	14%												x	x	
			C 1.3	Data teams	36%	C 1.3.1	Defining and realization of roles, responsibilities and ownership	29%							x		x			x	x	
						C 1.3.2	Reshaping data teams	14%									x					x
C 2	Process	86%	C 2.1	Defining data governance	64%	C 2.1.1	Defining and describing federated governance structure	36%	x		x	x			x		x					
						C 2.1.2	Defining granularity of global vs local data standardisation	29%								x	x	x		x		
						C 2.1.3	Defining data publication standards	14%							x							x
			C 2.2	implementation support	29%	C 2.2.1	Lack of data mesh implementation guidelines	21%				x	x		x							
						C 2.2.2	Managing change	21%	x	x					x							
			C 2.3	Organization structure	29%	C 2.3.1	Increasing agility of teams within the organization	21%							x			x	x			
						C 2.3.2	Alligning Business and IT	14%	x						x							
C 3		79%	C 3.1	Integration & Standardization		C 3.1.1	Implementing Data Interoperabilty	21%											x	x		x
						C 3.1.2	Implementing Computational data governance	29%								x	x	x			x	
						C 3.1.3	Implementing Data Standardization	21%						x	x				x			
						C 3.1.4	Implementing Common Interfaces	7%						x								
			C 3.2	Data Attributes	43%	C 3.2.1	Implementation of Data Accessability	21%							x						x	x
						C 3.2.2	Realising Data privacy	14%						x						x		
						C 3.2.3	Preventing Latency	14%						x	x							
						C 3.2.4	Ensuring Data Lineage	7%						x								
			C 3.3	Development	29%	C 3.3.1	Developing Software for a data mesh	29%	x			x						x		x		

Appendix C

Inclusion and Exclusion Systematic Literature Review

TABLE C.1: Inclusion and exclusion criteria for SLR with first 4 example classes and a sample of classified articles, Y = Yes, N = No, NA = Not Applicable, GM = DG in a DM article, M = DM focused Article, G = DG focussed article, E = Excluded, U = Unavailable

	Example: DG in a DM article	Example:DM focused Article	Example: DG focussed article	Example: DM-like architecture article	The Saxo Bank Case Study	Data Mesh: Concepts and Principles	Data driven information systems	From a Monolithic Landscape to Data Mesh	Data governance platforms	Mask-Mediator-Wrapper	Enterprise data management strategies	Data Ownership: A Survey	Data Catalogs
Is the article puplically available?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
Is the article peer reviewed?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	NA
Is the article credible?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	NA
Is the article a unique/ not an earlier included or excluded article?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	NA
Does the article discuss data governance?	Y	Y	Y	N	Y	Y	Y	Y	Y	N	N	Y	NA
Does the article discuss a principle of data mesh?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	NA
Does the article discuss data governance in depth within a single data mesh principle?	Y	N	Y	N	Y	N	N	N	Y	NA	NA	N	NA
Does the article discuss governance of architectures that implement principles of a data mesh?	Y	Y	N	N	Y	Y	Y	Y	N	NA	NA	N	NA
Does the article discuss the data governance of a data full data mesh implementation in depth?	Y	N	N	N	Y	N	N	N	N	NA	NA	N	NA
Classification	GM	M	G	NA	GM	M	M	M	G	NA	NA	E	U

Appendix D

Problem investigation interviews

D.1   Problem investigation interview protocol

Problem investigation interview protocol

Goals of the interviews

The interviews have the following goals:
1. Investigating the process of designing the target governance structure
    a. Formulating this process
    b. Explaining the challenges in this process
    c. Analyzing efficiency of process
2. Investigate usage of reference models during this design process
    a. Describing the models used in this process
    b. Formulating this interaction between reference models and the design process
    c. Define assessment criteria for the models
3. Analyse important components of reference models in this process
    a. Define important components of reference models
    b. Formulate improvement opportunities for current models

Roles of interviewees

Distinct roles are selected for this interview to cover the context of the research. These roles originate from the stakeholders of the research, and each have a different requirement specification and importance to the research. A general requirement is that the interviewees must be familiar with the concepts of data governance and data mesh. The distinct roles that are covered during the interviews are: Case representatives and Subject matter experts (SME) Each of the roles are discussed in more detail below.

The case representatives are enterprise architect consulstant that supported Energy 1 or Telecom 1 during with designing the data governacne structure. The SME's are experts on the field of Data Governance, Data mesh and Operating models. These roles are included in the research to obtain subject specific knowledge. The focus of the interviews is dependent on the role of the interviewee and is described per role

Role	Interview focus
DM/DG Consultants	Focus on the process of defining the governance structure and the models used for this. Also, the requirements of the model to support this process.
Manager	Similar to the Strategy consulting focus, but additionally focussed on the way the improvement of processes influences the efficiency the entire process
Data Mesh expert	Focus on the representation of the data mesh components and the assumptions made about data mesh. This role could also be beneficial to strengthen the assertions on data mesh migration and data mesh governance.
Reference model expert	Focus on the requirements of the models like usability and coverage and how these requirements could be measured.
DG expert	Focus on the governance structure and which important parts of the structure should be in the model and how this should be represented to increase usability and coverage.

Interviews script

1. Could you introduce yourself?
2. How would you describe your experience with data mesh?
3. In what kind of data mesh project(s) did you work on? 3.1. Can you introduce the organization? 3.2. What was the motivation to adopt a data mesh? 3.3. Can you explain about the starting point of the project? 3.4. Did this project require the design of a data governance structure?

4. Can you describe the process of designing the target governance structure at your data mesh project?
   4.1. Can you tell me more about the design process in general?
   4.2. What are the challenges that you faced during this process?
   4.3. What did you do to concur these challenges?
   4.4. What would have helped with facing these challenges?

5. Can you describe the (reference) model usage during the process of designing the target governance structure at your data mesh project?

6. Can you describe the important components of these models?
   6.1. What components of the models were important for the design of the target governance structure?
   6.2. Were there components lacking in the used model and if so, which one?
   6.3. How could these models be improved to better support this process?

7. Can you explain the resulting data governance structure that is designed during this process?

8. Finalizing the Interview:
   8.1. Do you have anything to add to this interview?
   8.2. Do you have any other questions about this research?
   8.3. Do you have any recommendations for contacting an expert on this topic?
   8.4. Are you interesting in the results of this study?
   8.5. Would you like to have a review possibility of the results of this interview by receiving the interview notes before publication?
   8.6. Can I contact you for the evaluation of the developed model?

D.2 Problem investigation interviews

PIT1 Problem investigation interview Telecom 1 Telecom consultant 1

Koen Hendriks 00:02
So, let's start Could you introduce yourself?

Telecom consultant 1 00:25
Yes, of course, I am Telecom Consultant 1 with the Deloitte for three years now and mainly worked in enterprise architecture there. So on designing a data architecture, assessing architectures creating Greenfield architectures, then later on got more of of an interest in data management so started to specialize more in that, specifically on data governance. So how to organize your data organization, and data quality. So what is good data quality? And how can you ensure or improve your data quality processes. And before that, I did also my graduation on the combination of AI and data management. So also looked into that combination. And I've been active in various industries. So in the financial services, consumer industry, but also technology and in the food industry.

Koen Hendriks 01:31
Quite an extensive portfolio. So I think you can have a lot of information for his research as well. So that's nice to hear. I also know that you participated in a data mesh project recently. Could you explain a bit what that project was and what your role was within the project?

Telecom consultant 1 01:56
So I've been on the project for eight months with for a large chip designer or manufacturer, and data has a central focal point in this project, because the problem that they had is that they were unable to share data within an organization. And they were very siloed. So they were looking for a solution to enable this data sharing and organization and quite early on, the choice was made to adopt a data mesh approach, also in designing the operating model for the future organization, but also in designing the architecture, and all the related deliverables that we made there. My role in the project in the beginning was to look at the current site architecture to do an assessment there, and to find the gaps in terms of a data mesh. So what are the applications and the solutions that they require to be able to adopt this data mesh, and in a second phase, so the last three months, i was mostly involved in the vendor selection, for the specific solutions that we identified as a gap in first part of the project. And in parallel, we also, with the rest of the project team, looked at this operating model and how the client should organize itself to not only have the application and data in place that we were focusing on, but also the roles and responsibilities to adopt this data mesh.

Koen Hendriks 03:26
That's nice to hear, also good that you both participate in the first part and in the second part. So setting up the architecture and assessing the gaps. That's also really interesting and what what the company could or another organization could move to, and also that it's kind of further into until vendor selection, that means that it's like a very broad process were you involved in so that's nice. So I hope you have a lot of knowledge also as well on the processes. Can you explain how you assessed like the current architecture did the gap analysis? Yeah, exactly. What does the future architecture consists of like what what kind of gap analysis you do there?

Telecom consultant 1 04:40
So indeed, we started off with they didn't have a current state architecture. So what we did firstly, is we use a reference architecture to define the future business capabilities. So for that, we use the Deloitte framework that we also use for previous project where we designed it Data Platform that describes all the capabilities that the organization needs, such as data storage, data exchange, data infrastructure, data consumption. So all these capabilities, to get as a starting point, and then met the current applications that they are using onto these capabilities to see what they have. And then we assess and analyze those applications based on a few criteria. So is it how is the business fit? It's future proof.

And how is the technology fit. And based on this assessment, we would make a target architecture and see at some of those applications, we could reuse for these future capabilities for the from the reference model. And also could see for which capabilities, we didn't have solution yet. And those were the gaps. So for example, a solution where we could have a data marketplace so that these data products in a data mesh are discoverable, findable, and also consumers can request access. And second most important one

was the data exchange part. So once you identify which set of product you want to use, how can you actually make sure that you exchange this between a data producer and consumer. So these were two large gaps in the architecture.

Koen Hendriks 06:26
And if you look at the organizational architecture, like the roles and responsibilities that you mentioned before that you also looked into. Were there already capabilities that fit the data mesh, organizational architecture, or was there also gap that things needed to change, and that you needed to design the target architecture of the organization with the data mesh capabilities?

Telecom consultant 1 06:54
They had some roles and responsibilites in place but, they were not very mature on data governance, they were just starting out. But they were very siloed. So each business line had their own IT departments. So r&d, had r&d, it manufacturing and manufacturing it. So it was very scattered. There was also the reason why we chose to adopt data mesh, because it is also a distributed data governance. So that was the main reason. So they do have some

distributed teams already. But what they lacked was a structure in which they could collaborate. So the distributed teams, and also there was no global data governance, no oversight. So this was something that they need to set up like a central data office or CDO, which would coordinate between those different data domains, and also the current IT organizations that they had, so that they assured that they were interoperable, and use essential standards, so that they can also exchange data when they get to this data mesh. So for the CDO, central data office, there was nothing defined yet, that's also what we did. Also, for roles and responsibilities, that would be in a data domains, they didn't define it yet. So that is definitely something that we started off with and help them with.

Koen Hendriks 08:28
If you look at the starting structure of Telecom 1, the structure is already domain driven, did you use those domains directly as definition of the data domains?

Telecom consultant 1 08:48
No with those domains required large so, then you would have maybe four or five domains. So there was a large discussion on this. So what should the data domains be also fully aligned with the data mesh theory we eventually settled on, let's not define the full enterprise upfront, but start with a domain first. So we picked the domain with new products introduction, which was part of r&d, that on itself would be a viable data domain. So we efficient with r&d In the end, maybe we'd have four or five data domains. So we would split this up even further than the domains that they had at the moment.

Koen Hendriks 09:36
The structure that you mentioned, like the domains and the CDO, so the central data office is that structure with the responsibilities decentralized and one on top that enables the interoperability How did you come up with that structure? Is that one on one of from the theory?

Telecom consultant 1 10:10
Yeah, so that that is a good point. So what we did is we used the technology operating model in a box from the Deloitte TOM on a box as a reference model to start with. And then we did a workshop with the clients on the service catalog. So this is part of the TOM on a box , and explains all the services within organizations so these are mostly internal services that one team will provide for the whole organization or for another team. And we use those services to have the discussion on what should be where so we created this RACI. Responsible, accountable, consulted the informed. For all the services. So we would say for example, and who can set data standards? As if that will be the service? Would it be the CDO? Would it be the domain? Or maybe a third one, I think of mostly the it was a split between CDO under Domain, and maybe the IT organization, I think those was this was the logical split. And to get out with the client, we had the discussion about what should be where, and that lead to interesting results. Because in that case, you can say the CDO is accountable, but the domain is responsible. Yeah, that led to another discussion, because the the CDO is not making this choice, the domain is, but if the domain makes the wrong choice, then the CDOis accountable. So this was really a challenge in this operating model and the approach that we took. So after that first workshop, there was a follow up, where we explained the contradicteries, in this way of working, right, so there needs to be some link between the responsibility and accountability to make it work,

Koen Hendriks 12:23
and how did you solve that? Is it something that the people had to agree upon? Because it was in the model? Or had you some kind of solution for that specific problem?

Telecom consultant 1 12:36
So we did make a proposal, I believe on how it should be. But there's always this fine balance because now the domains are at R&D IT full degree of freedom, and they don't want to give up that much. But you do need to have a certain global standard to be able to to make sure that you can operate with each other and can you eventually share this data? So, this is what caused the discussion. So some would say we want to keep this responsibility within the domain. But then the end we would go have a discussion and explain okay, but this would mean that the CDO might not have a say certain decisions that he is accountable for.

Koen Hendriks 13:32
Were those discussions together with the organization?

Telecom consultant 1 13:42
Yeah, you can see a sort of like a Delphi. So, we first had initial discussion, then analyzed and really calculated what would be the result of such an service model. And then you are reported back replayed back, and they could indeed see that some of those decisions need to be reversed.

Koen Hendriks 14:08
Can you further explain the different layers within the data governance structure? Do you also have a data governance layer for dats products?

Telecom consultant 1 14:32
Yes, so purely in theory, we would have some responsibility for the set of products that will be a data product owner and these will be in the domain. This most likely will be somewhere from business to own the data. And also within a domain they would have the IT capabilities to create and maintain such a data products. So that's the ultimate theory. Read

Koen Hendriks 15:02
That is in theory, and how is that in practice?

Telecom consultant 1 15:05
In practice, this was quite a big step. So now they had central IT, organizations, some related to a domain, which is r&d and one Cloud Center of Excellence (CCOE) for the organization. So eventually, we decided upon this consulting model, where we would keep the central organization. So the CCOE, within that case in the organization fall under the CDO so the central team, and then provide services to the domain to create and maintain their state of products, with the idea that over time, these domains would also hire and develop their own data capabilities, so that they can do it themselves. And then in a year or five, this decentral team with consulting model would would be would decrease, or maybe you can cease to exist, once you have fully mature data domains.

Koen Hendriks 16:06
So the CCOE, was both responsible for the creation or the development of the data products, and for the implementation of the standards for the data products.

Telecom consultant 1 16:19
So the creation of the standards would be in the CDO, and this will be more business related role, The CCOEat its main goal, to provide the infrastructure for different domains, but they also have the data engineering capabilities, which ideally would be, you know, data domain itself. But in this consulting model, they can, like lend out the resources to the domain. And in that way, you can set up these data domains, even though they're not fully mature. So they don't have their own data capabilities fully, but they can still leverage some resources that will create these data products with them until they are mature.

Koen Hendriks 17:08
is it true that the CCOE is still in the future architecture, but that it is only responsible for the infrastructure?

Telecom consultant 1

Yeah, so providing the self service infrastructure, which was also part of the data mesh principles.

Koen Hendriks 17:42
Yeah, this sounds very like the theory presented in the data mesh theory book. I can imagine that it's really difficult to map the theory to the organization. So the organizational structure and the organizational structure in theory has to come together to create like an organizational structure that fits and has the desired capabilities of a data mesh. What kind of challenges did you run into during the design of the new organizaton structure?

Telecom consultant 1 18:55
Yeah, definitely. So we had many challenges. So for once, there is an CIO in the organization. So one of the questions was how to position the CDO with the CIO. The CDO is the central data office. It's not the officer, but also the main responsibility is still honor the CIO. But this was an interesting question like where was responsibility lie?

Koen Hendriks 19:25
Is it then responsibility or accountability or are the those both in a situation?

Telecom consultant 1 19:31
Accountability, so the high level responsibility to execute this is a bit lower? So there was a question so how to position the data mesh accountabilities in the leadership?

Secondly, finding the balance between the domain autonomy and a central data office is difficult. So you need to have those standards, but also those domains, they were hesitant to give up too much of their own autonomy. So they want to be able to make their own choices, which we want to facilitate, of course, how we solve that is to say, for my central team from the CCOE, but also there was a central BI team, we will help you, but only if you follow our standards, but you are free to make your own choices. So in that way, I tried to use a carrot more than the stick, because the stick wouldn't work anyway. Because every domain now I just want to call on me, so they wouldn't listen, if we try to enforce something. So what we did, if you create these data products, and other domains will use it, and you're not adherent to certain standards, that will be a problem for the consuming domain. But what we said there is that for a data products to be able to be shared, it must comply to a sort of checklist of the standards. So we basically said, if you can comply to the standards, on your own applications, and in your own way of working, that's fine. You need to comply before we can put your data product into the marketplace. If you are able to do this fine. But for most parts of the organization was domains, we don't expect them to be that mature that they can do that. So then we say we will provide some standard infrastructure for you a data producer space on AWS, that is pre configured, and already contains some some defined building blocks that you can use. And in that way, it's easier for the domain. But it's also easier for the CCOE that would support this infrastructure, because they know exactly what is in there. And they can support the domains in this consulting model. But if a domain sais so, that I want to do very specific Analytics, which I cannot do in the tools that you provides, then they can go ahead and of course, choose your own. So we want to give that degree of flexibility with maybe the risk that those data products that results from there are not fully fully according to standards and might not be able to end up in the marketplace. But we want to simplify the central support. Because we cannot support everything from a CCOE perspective,

Telecom consultant 1 23:30
Is the check for adherence to standards automated?

Telecom consultant 1 23:45
Yeah, so it's a quasi. There should be some some manual checks still. But also some things are mandatory when indeed like publishing it to the marketplace. So in a checklist contain things like a description, the location of the data, terms and conditions for you to use the data or data contract. So these were checkmarks that that you can automate probably in the marketplace, which is not live yet. But certain checks, you can do there other checks you need to do manually. So this will be mostly I think, a coordination between the data product owner, the Central data office and the data consumer. So data consumer can also have certain requirements. So let's say you make this data products available through API, but you want to have a certain update frequency or more columns in this data product, then this will also be a discussion between the data product owner. So this is not necessarily requirements, so to say, but a data product should be fit for a data consumer. So depending on the data consumer those checks can change. So for one data consumer might be fine to have a data set is uploaded once a month, and it's only available in bulk. So the consumer might require weekly updates, and the ability to filter some rows and those kinds of crimes quirements.

Telecom consultant 1 25:38

Let's go a little bit back to the design of the whole structure and the whole idea, the whole architecture, you told that you use the reference model, the Deloitte technical operating model, also to design that. And you use like the the RACI matrix. So it's nice that you follow that. Using all these models, were there. components lacking on these models?

Telecom consultant 1 27:23
Yeah, I think so what we use were not specifically data mesh models. So if there will be some reference on what would be the advice and way of working. So this can be both on architecture level, but also on roles from responsibility level, that would really help facilitate the discussion. Like what we said, so we use the services, but in the end, we ended up with a set of services, that's not really feasible to implement. Maybe you can summarize this in four or five, maybe 10 Difference variants, combinations of these services, because some are linked and some you cannot easily distribute in a distributed model, or you should distribute them all. So those dependencies were not in the model that we use. So having a reference there, which would benefit us

Telecom consultant 1 28:26
What kind of services are that?

Telecom consultant 1 28:37
So there were things like the decision on what standards to set or technology choices that should be made. And this was really something that if there is a choice to be made by the domain, but the CDO is accountable. That will not work. So the CDO needs to be in this decision making process to a certain degree. So these combinations weren't visible in the original model, because that assumed a traditional data governance model where you would have everything in the CDO. So this distinction was really missing. These combinations weren't visible in the original model, because that assumed a traditional data governance model where you would have everything in the CDO. Yeah. So this distinction was really missing. Okay, and then the services are like the gross as he's of the operating model, right?

Telecom consultant 1
Yeah.

Koen Hendriks
Okay. Okay, that's not kind of good to hear that there is some points of improvement. If you look at this specific process of defining those services, how did you approach that now without a model that is made for data mesh?

Telecom consultant 1
Yeah, so we used we use the tom in a box model, which was a general technology operating model. And we adapted this in these two rounds. So we leverage what we had for a traditional governance model, and tried to fit it on on a data mesh governance model. But that took some time, like we said, so we had an initial discussion, we said, okay, these are a list of services, these are responsible, accountable. But then we only found out that this wouldn't fly, that sort of choices were dependent on each other. And that you cannot say service a is a domain service b is a CDO, but some of them are actually independent, and should be considered as a combination. So then we had to redo the discussion. So I think if we would have had a tom in a box variation for data mash, maybe we could have done that in the first discussion. And have those dependencies and have these considerations or making these choices in a data mesh context, already ready in a discussion.

Koen Hendriks 02:00
And if you look at the services that you eventually got, as a list or after those two sessions, got as an output, do you think that those services are generalizable for other organizations? Or is it very organization specific, for example.

Telecom consultant 1 02:24
No, so the tom in a box is a generic model. So it can be used for or adapted to any organization.

Koen Hendriks 02:33

But I mean, the If you look at other organizations, do you think that if they would use the tom in a box model, they would come to the same services so that the services were the output services weren't organization specific?

Telecom consultant 1 03:16
I would say so in general, I think goes services are generic for most organizations. There can be very specific services that an organization needs requires. So you can always add a couple, but I think the core is the same. So these were really services. So as a data consumer producer, I need to have my infrastructure, who will provide me that infrastructure. So these are really generic things that that every organization needs. I want to ask the BI tool who can provision this BI tool for me? So these are really general things. So I think this this can be applied to any organization, but it needs to be adapted a little to fit their specific needs, but I think it's a perfect starting point or it's a good starting point.

Koen Hendriks 04:10
Okay, nice. I think that we found some kind of a pain point in the in the process. If you look at the rest of the process, are there other things that could be improved with the improvement of modelsbbesides this, or was this the only thing you run into? Okay. So on the services, any other things? In terms of models? purely on the organizational model or a more contents? Purely on the process, technology people, of the data governance responsibilities. Were there some other problems with reference models or lacking of functionalities for data mesh specific context?

Telecom consultant 1 06:33
Yeah, so we just talked about entity services. I think from a architecture perspective, we also just use a generic data platform reference model. So it wasn't also very data mesh specific. So we added a few capabilities. But later down the line, we realized that we were missing some capabilities, some are maybe redundant in context of a data mesh. So also, having these business capabilities for data mesh would be a really interesting reference model to have. Because then there are fundamental changes between regular capabilities or having to split in in a decentralized and federated governance model.

Koen Hendriks 07:20
So you use the data platform reference model for the business side. So for the organizational capabilities as well.

Telecom consultant 1 07:29
Yeah, yeah. So the business capabilities, and then we use that also to map the application and data on top of these business capabilities. So the togaf architecture theory.

Koen Hendriks 07:40
And what kind of data platform reference models that you use for this?

Telecom consultant 1 07:47
That's something that we use for previous clients as well. So it's not really a defined model, I guess you can say it's the Deloitte proprietary. But it's a model that we use a couple of times for organizations that wants to introduce a data platform, but that that form would be centralized. So there will be a central data platform team, where they would say, all different domains, or different parts of the organization, would all publish their data on this specific platform. And this was really a fundamental change. Because in a data mesh, we say these domains, they can store and own their own data on their own part of the infrastructure, they don't need to use a central data platform, as long as they publish their metadata in this data catalog. And it makes sure that they used to say the right standards so that the data is interoperable and can be exchanged. So this really provided some audit capabilities that you wouldn't need in a central data platform that weren't in the in the model yet. And that we also struggled later with positioning that in the model.

Koen Hendriks 09:02
I can imagine that when mapping the data platform reference model, and then the business layer to the application layers, then or the to the date data layers with such a change in the architecture could be kind of a challenge when there's not included in the reference model.

Telecom consultant 1 09:36

And also a good one would be identity and access management. If you have a centralized platform managed by a central team, it will be relatively simple, because you would all be on the same technology stack. And the central team would do IAM but if you have this federated you have maybe one domain on AWS, one domain on Azure. And they all have their own data products, but also their own consuming systems that they have. And still, you need to have some kind of global IAM. So there needs to be some match between the domains and the global IAM, which was really a tough one, and also wasn't part of a central model. Right? It would just assume you have one Iam team. But in this case, you need to ensure IAM over multiple domains over multiple infrastructures maybe. So that's really an additional challenge that wasn't in a model before.

Koen Hendriks 10:44
Yes, that I can imagine that is a challenge. Okay, I'm looking at the questions now, to see whether there are still some questions and answered. So We answered quite a lot. So that's nice. Just for some clarifications for me to be sure of your answers. Did you use DAMA DMBOK?

Telecom consultant 1 11:20
Yes.

Koen Hendriks 11:20
Why did you use this model?

Telecom consultant 1 11:25
Because it is quite extensive. So it contains the services, it also contains from Out of the box processes, some roles. So it really contains all the puzzle pieces, so to say, to create an operating model, maybe this analogy is a good one, because it has all the puzzle pieces. And normally in a puzzle, you have also a picture of what it should look like. And what it was missing is this picture in a data mesh contexts. So you have the picture in a central governance structure, where you know, okay, this is where we're, I saw reference where we should look at the miss that that picture for data mesh.

Koen Hendriks 12:16
Okay. And I hope that I can maybe add something there. During my research, that will be will be nice. definitely If we really look at the whole process of the designing the data governance structure. What is the current state of the project and how long did this take?

Telecom consultant 1 13:40
Yeah, so a lot of it was still ongoing discussion. So to prepare and do the service catalog workshop, I think they took it took around four weeks with I guess almost three people to prepare this and to host the workshop, the workshop itself was about four hours then they had to analyze or the results which took around 2 weeks and then we had a follow up discussiona. So I think before they could set on this service catalog that will be around two months before that was decided then you would have that and in the following six months there are continuous discussions on the implications of that and their role descriptions. And so once you know for example, what is the responsibility of a Data Domain? Then you need to draw up Okay, so what roles do I need in this domain? Do I need one data governance expert? Do I need data quality experts separately? Do I need of course your data product owner so then drawing this out was something that they did in the last two months, I guess from those eight months I was on a project. So that is something that should have been delivered in those two months. And we were starting up the first domain according to these roles, and also these services that we defined. And they were still in the process of setting it up. And they still are at this point. Also, finding the right people with the data capabilities was quite hard, even in this consulting model. So, they're still as far as I know, in this process.But the goal was to have this domain really live by sometime, March this year. So then from beginning of the whole process, to the end, it will be a year almost from the first workshops to actually being alive in this way of working according to the governance model.

Koen Hendriks 16:08
Nice, we talked about the different models that you used, the whole process of designing the structure, designing the roles, responsibilities, but also a little bit further into the standardizations. And the difference of standards, the adherence to standards. And well, the See, I still have to remember, it's CCOE can you explain more about that?

Telecom consultant 1 17:36

Yeah, and it's just consulting model, I think it's key for most organizations if they want to move because most of them have now a central data team. So from moment one, you cannot say, split up, and each of you will join one domain. So that will be that wouldn't work as well. So starting with this consulting approach, I think would be a really good one for organizations that want to adopt data mesh.

Koen Hendriks 18:04
And one thing just came to my mind, the central data office. Do you also know what kind of people are inside that central data office? Are those specific people that only work at the CDO? Are those representatives from different domains? How did you structure that?

Telecom consultant 1 18:32
From the start, it will be mostly business owned. So you would have a Data Platform owner, that would own the marketplace and the exchange solution for the organization and the CDO itself, underneath will be CCOE, which has the technical capabilities. But they the other IT capabilities are mostly within the domain still. So R&D IT manufacturing IT . So that's how they would start out with so the CDO will mostly have the decision making power and setting the standards, making sure that you maintain the part of the infrastructure that is common for everyone to self service infrastructure. But, all the other it and data capabilities with this would be within the domains.

Koen Hendriks 19:29
And then the decision making on standards like setting the standards. That's something that's done in the CDO,where there are also domains involved with that process according to the target structure.

Telecom consultant 1 19:49
Yes, so in most choices that we made there, we would involve architects from all the domains all the different domains and jointly mentally brainstorm on these on these choices and standards that should apply. So, in that way, we fulfilled in this project, the role of the CDO. So to have a central overview, central guidance, and really facilitate making these choices. But of course, this should be in consult with the different domains you cannot see do make all the choices and enforce them, especially if the one responsible are in the domains. So there is this power balance also. So you need to facilitate the process help them. But in this structure, the CDO didn't have that much power. And the power was in collaboration with the different domains.

Koen Hendriks 20:44
Okay, that's something I can imagine. Just want to be sure whether that's something you also implemented there.

Telecom consultant 1 20:55
Okay.

Telecom consultant 1 20:56
I think also one thing to add from my side, so what we really lacked was a reference model that would specify the differences for data mesh. So these considerations on what to do central what not to do Central, what should be a dependency in that case. So that was really one also think was really struggled with are what are the questions that we need to ask? What are the real choices that we need to make? And this was something that was something we searched for at the start? What are the real key decisions in adopting this data mesh? What can you actually choose? Maybe this consulting model was one of the choices, but also what part of your technology stack will be centrally provided? And what is the degree of freedom and in the domains? So really, these questions are very hard to figure out? So having some sort of guidance, there will also really help out what are the key decisions that an organization can make when adopting a data mesh?

Koen Hendriks 22:00
And then is of course difficult to find the answer on those questions, but finding those questions themselves, was also like a challenge for you, as I understand.

Telecom consultant 1 22:11
Yeah. And also, we did another project where it was for a daugther company of, of a larger organization. And this organization would say, we're going to simplify the IT landscape. And we want everything all our daughter companies, they should use the same it IT stack as we did. Unless there's a good reason that this starter company should

have their own choices. So what we really did also per capability, really look at should they adopt the central stack from the mother company? Or is there enough reason to introduce their own? Or maybe, as this daughter company, should they provide the service to the mother organization to discuss somewhat of a similar discussion? And there, it was a bit simpler, we were able to make this decision tree where you would have really have four steps? Like, is this a differentiating solution? Yes or no?Do we have certain requirements that the mother company probably doesn't have yes or no. And in that way, you would really end up at these centralized outcomes.

And you would really have a structure to go through to ask the most important questions, and end up with this capability map, where you would say, okay, that's those 10 capabilities, we don't need to spend more time and effort on it, we will sources it from the mother organization, these capabilities, we can source on the mother organization, but we need to adapt it. And there's other sets, we are so good at it, we should provide the service to the organization as a whole. And as such, the structured decision making structure, we didn't find in any context of a data mesh further, that is something that could really help.

Koen Hendriks 24:02
Do you think that the complexity of decisions around data mesh can fit some kind of decision tree like that, or do you think that you can only for a specific part make a decision tree like that ?

Telecom consultant 1 24:33
We tried to do some so in this capability model, we said there are like five overarching level two capabilities. So infrastructure, and we did infrastructure, you would maybe have CI CD, or level three get abilities. So maybe you can make the specific choices on a level two capability but not on a level three. I don't know but I guess Probably something there to make these structured decisions. It might be helpful, for example, to do it only for the roles and responsibility division or something like that.

Koen Hendriks 25:18
Okay. Also very interesting. I didn't Oh, there's only five minutes left in the meeting, but I didn't. Yes, all of that. So, nice addition. Do you have anything more you want to add to this? Interview? I

Telecom consultant 1 25:35
think I mentioned most of the points I wanted to make.

Koen Hendriks 25:40
Nice. Nice. Okay. Um, let's see. All my questions are also kind of done. There's Thank you very much for the interview and i will keep you updated on the results.

PIT2 Problem investigation interview Telecom 1 Telecom consultant 2

Koen Hendriks 00:00
Can you introduce the project you are currenlty working on? .

Telecom consultant 2 00:04
So the Telecom 1 project that I've only been on it since, let's say end of October, let's say mid to end of October. So my understanding of the project is we're trying to build them a marketplace, so that data producers can publish their products on there. And then data consumers can request and get access to them. And then a bunch of other functions like you know, escalate data issues, escalate data products that they'd like to see, etc.

Koen Hendriks 00:37
Is that the goal of the whole project?

Telecom consultant 2 00:47
I guess that's the goal, and then putting in the processes and governance in place to support that. And then I was like, upskilling the people and identifying the right roles and then activating them enough so that they have the right skills to continue scaling that to the other data domains that they have. So my part in it is like helping to set up the data governance side of that, and we've been focusing on the roles and how they fit into the domain, we started looking at the central data office. The way that they wanted to say, to

do it, because you know, in a mesh, how really the, I guess, like most of the work needs to happen within the domains, because they're the people that have most knowledge of their data. And the central team is really just to enable the domains to be able to do their job, like to produce data products. So to really high level we haven't really deep dive into this with them, would be to make sure that there is like a platform that they can publish and consume products, but also that there are at least, policies, processes and standards in place to set a baseline for how their data should be managed for maybe a process for how data issues should be escalated. How products should be requested, what that decision making process looks like and who should be involved in that. So that should come from the CDO. But really, the data domain should be responsible for taking that and then making sure that that runs as Business as usual. From a day to day perspective. So I think the problem with setting it up in the first place is number one, they don't have a good level of maturity for their data governance. So they haven't really done this before. And usually my understanding of mesh is that it's for organizations that are a little bit more mature, and have some of these. Yeah, I know already has some of these in place. But that's okay. Because they seem pretty determined. They don't like I think data governance is quite new to them. So they don't have owners, and they don't have stewards. Okay. But we started a POC, and we started the pilot. And for those, they started assigning those roles, but those people still need to be on boarded. So they understand, you know, what do I need to do with the data? And what do I need to do as a as a data steward

Koen Hendriks 03:26
Is defining the roles and reponsibilities an interative process?

Telecom consultant 2 03:29
I think it depends on the client. So with them, they want to do it as a bit more of an iterative process, because I've worked on other projects, not for mesh, but for, like, I guess, like standard data governance projects, where the client is happy for us to tell them what are the risks, what roles and responsibilities of a data owner data steward, and then they go and they assigned the people and then we help them onboard them. But with Telecom 1 what they wanted was, they wanted us to, I guess provide a point of view for theoretically, what should the role of a data owner data steward be?

Koen Hendriks 04:15
Have a look at their skills and capability, and then also the language that they use? And then we kind of mash that together so that we come up with a role description that is both theoretical but is also practical for Telecom 1? And how did you do do that? Because that's, that's like my, the interesting part for this thing I want to do my thesis on like the as you call the theoretical roles, the school description and also the responsibilities of that, yeah. How did you come up with those roles and responsibilities?

Telecom consultant 2 04:49
Okay, um, so, there were some roles and responsibilities that were developed before I came onto the project, but I think, um, due to like competing priorities, it wasn't really looked at in a lot more detail. And then I came in, and I had a look at them. And I reviewed them. And I reviewed them a little bit more added a little bit more detail based on, I guess, like, the experience that I had working with data governance projects. And then adding a few more things like, I think what we had was a little bit higher level what we initially proposed, then I looked at, okay, so for the scope of this project, what do they want to do? What skills would be needed in order to do that, in practice, and then pose that. So where we are at the moment is, I've got a template for a proposed description and skills required for each of the roles. And what I'll do next is, because we've already done the POC, and we're about to start the pilot would be working with those people that have been assigned certain roles. So people that have been assigned data or data steward and asking them, does this description resonate with you? Was there anything different that you did when you did the POC or the pilot? Or is there any skills or any responsibilities that are Telecom 1 specific that we need to put into our description? So it's a bit of a team? Collaboration? Effort, almost a co design and the Exactly, yeah, because I think I think there's one part where, you know, you can propose something in theory. But realistically, there is you need to understand what your client requirements are and how they work and then tailor it to them. Yeah, because it has to it has to resonate with them otherwise.

Koen Hendriks 06:36
Is there a reference model already for the roles and responsibilities?

Telecom consultant 2 06:54
I don't think there's a reference model. And to be honest, I don't think that would work. I wonder like, because there's like a massive human element to this, there's like a massive change element to this, especially if they don't have any governance. So you need to work out with them. Like, how are people going to best be receptive to being given a new role, like, for example, sometimes when you're assigned, when you introduce the role of data steward people are a bit like, Whoa, I don't want to do this, ya know, they might

already do some of the data steward responsibilities, because they see it as being more being given more work than

they already have. So then it's a bit of an exercise to work through with them, you know, helping them understand what a data steward role is, but then also taking learnings from the pushback that they were given to adapt the way that you communicate what this role is, and how it's going to work in practice. So it's maybe more like a set of questions that you need to ask with, or explore with the client to then come up with what's the best approach? Because I think in theory, you it's like, you know, what does the data steward do? What does the data owner do? How should they sit within the organization that then these other pilots, like how do we actually implement that in practice, and that's the human element that is difficult to have, like a standard framework that shouldn't be like, there's like, there is a framework that I feel like it'd be super high level. But there has to be said that you need to adapt to the requirements of the people, because it also depends on you know, whether or not there is a champion on their side, who's going to help push the agenda of setting up data governance on the client side, because there's only I feel like, there's only so much that you can do, it's like going to someone's house and being like, Oh, I'm going to redo your whole house with them, not the people living in the house not having a bind, like you need to have someone there to kind of advocate for the changes that you would like to introduce, but also do that change management piece, where they talk to the stakeholders understand maybe their concerns about being assigned any role, or how it's going to fit in with their existing roles, existing forums and kind of work that out together.

Koen Hendriks 09:00
How do you approach the federation of the data mesh?

Telecom consultant 2 09:46
So I think we're gonna go with that, like the central data office is going to basically own policies, processes and standards that the domains will use. But what the domains need to do is take those policies, processes and standards and review them against their own requirements and see if there's any tweaks that they need to do make those tweaks and then, I guess, establish that as their policies, processes and standards for them. So it's kind of like taking what is the standard, but making sure that it fits with their requirements.

What we're also going to test is using because you know how, like, in like traditional data governance, like models, there's like this, you call the executive sponsorship, governance body, yeah, then there's like the data governance councils. And then there's the data working groups. And their data working groups is what we've seen as more of the Business as usual operation of the domain. Like more of the building, the actual doing, like the building the data products, monitoring the data quality, etc. And then anything that needs to be escalated, or that will maybe an issue that impacts cross domains needs to go into the data governacne council for discussion. So having that that data governance council as like the link between the Central data office and the domain. So for example, as part of that exercise, and this is something that we to actually define in a lot more detail with a client is okay, so the Central data office sets up the standard policies, processes, and all that. And then the domain needs to, I guess, like, review them and make sure that there's nothing that needs to be adapted for their own purposes. And if there is, then maybe a suggestion is that they go to that central body to get sign off, like, you know, you've proposed, you've given us this process. But you know, for our domain, we see a little bit differently, because of these exceptions.

Koen Hendriks 12:29
Who are then part of the Central data office?

Telecom consultant 2 12:31
So, in that central data office, they're going to have leads, one for data governance, one for data literacy and analytics, one for the platform, and then one for the portfolio's. So I see those leads kind of being the people that are like the final check, like, if you want to change this policies that we have set in place, essentially, then it should go for example, to the data governance lead to say, okay, yes, I understand why you done that. And I agree, do it. I think from a data governance perspective and a data management perspective, I think that every domain needs to still have the data owner role. And then from a product perspective, who owns the actual final products that are then published onto the platform there should there is a data product owner role, so anything relating to, for example, changing the processes that are set by the central data office, anything that needs to be raised for approval should be raised by the data product owner to maybe the data governance lead in that data governance council.

Koen Hendriks 12:34
And what exactly is the data governance council ?

Telecom consultant 2 13:52
So we've kind of tweaked it a little bit, because we're also trying to, because I think some one danger from mesh is that because you kind of give a lot of control and a lot of power to the domains themselves, you lose some of that central oversight, which is not necessary, which is not necessarily a bad thing, because you still have a lot of bottlenecks

happening, but at the same time, you can stop efficiencies from happening, as well, because everyone's just doing their own thing. Like you don't have people making them. So then what we want to test is using that data governance council equivalent as a forum, or the platform where the CDO and the domains talk together, and then the domains also talk to each other, as well. So it's kind of like a forum to connect people but not everyone within the domain needs to go so be representatives of the domain. So from a data governance perspective, I see that being maybe either the data product owner and the data domain owner to represent, you know, these are recommendations.

Koen Hendriks 14:30
What exactly in your data governance structure is the difference between the GDL and the council's?

Telecom consultant 2 15:41
So I've seen I've proposed that as a starting point, just because it's kind of like the standard. And also, there's no CDO at the moment. . So there needs to be something there. And so I guess, to like, try and delineate responsibilities. So these aren't necessarily like there needs to be exec oversight into that. But it's more of a way for the domains to talk to each other with guidance from the CDO members. That's where it happens, that middle sort of thing. But it is something like it's a good point, because I think it needs to be tested, like does this work? Like once they actually have the CDO roles, then does it make sense that we have, say, that CDO and then the data governance council is two separate bodies? Or should we actually merge them together? Because also, I also split them? Because there might be discussions that don't need that there are more executive level discussions like, you know, funding, all that kind of stuff, like the data domains don't necessarily need to be in those discussions, or they might not actually cut that decision making power to it for it to make sense. So it's kind of like a starting point. But the idea is that we're going to test this as we go. And as we get more roles, on boarded onto it. And also the data governance, that data governance council equivalent, we're not sure how many forums are going to sit in there at the moment. So at the moment, they have one existing forum, which is like a Data Portfolio thing where they want to talk about the value cases, the idea would be then Okay, so let's prioritize this value cases, this means that we need to make these data products, and then that gets passed on to the domains themselves to produce those data products. So at the moment, we have one domain that we're working on. So I think that at this point, a single board should can be fine to touch upon all important issues. But as more and more domains get on board, I think they're gonna have 13 domains, that council is going to be too big to be effective. At that moment you should think about splitting up this council. So I think because it's new, it's okay. But the other thing is that we need to get them into the habit of having these sort of different discussions. So you know, how does an agenda for a data Working Group and the participation and data working group, how is that different to an agenda and a participation from like a data governance council, sort of forum, so kind of getting them into at least that sort of habit or like that sort of, like while you're working, and then being in a position to then adapt it, while keeping I think, like, I think there are certain capabilities and standards that they need to follow to make sure your data, data is managed correctly, there is like, an efficient and clear decision making process as well before kind of changing it up. But it also we have to propose something, because it's also easier to give people like when you tell them, this is your role. So we also give them like some sort of parameter within which they work. So they understand who I work with day to day, who do I escalate issues to.

Koen Hendriks 19:38
Nice, and then, Is the data domain owner then responsible for the performance of all the products in the domain and the data product owner responsible for the performance of the data product?

Telecom consultant 2 23:19
Yeah, so they might be assigned, like a data product that they need to be responsible for? So the whole package of it? So they should maybe be like, this is another thing that we need to work out in practice, how this is going to work? Like how do you assign the data products responsibility to people? And then when data consumers request data products that don't exist, but needs to be made? Who should that who should be looking at those metrics? Should it be, you know, a regular working group of data product owners, who then assign, like maybe particular topics that they are good at that then work together, come with prioritization with the data domain owner and the data domain owner thenadd their input as well. And once they agree, then they go to that middle data governance council and kind of discuss that with the central team, as well, because another thing that the central team need to make sure is that okay, so the data domains want to create these products, is our platform, going to be able to meet the functionality that needs that it needs to have for those to be published? So it's kind of like making sure that there's like some sort of connection

Koen Hendriks 24:38
Are there also data platform representatives in the data council and in the CDO?

Telecom consultant 2 24:46

So in the data councils and the CDO is based on the pillars that our client wanted, and it was like portfolio platform governance, data governance and data literacy and analytics.

Okay. So from a product perspective,I think that in the council, there needs to be some sort of connect between the data platform lead, and the Data Portfolio lead to be like, Okay, these are the products that we want to prioritize. Can the platform allow us to publish that? Does it have the right functionality? Yeah. Is it like, and are there any problems with the platform that we need to be aware of? Because the that central team is going to own that platform? And then from a data governance side, I think there should be like a separate working group where they look at, you know, data quality metrics, look at any issues that have been raised. SoI reckon there should be like more data steward data owner, like, you know, more like traditional governance, and then decide whether or not there are any initiatives that need to be spun out. And if there aren't industries, find out then you know, who was in the domain should take responsibility for that, and then build that initiative team to action that. So, i think that there should be data product working groups in the domain and working groups that focus on the data governance. So they need to monitor the data quality of their domain. Look at any issues that have been raised, and all the issues that have been raised, is it showing that there needs to be some sort of data governance or data management improvement initiative that needs to be introduced? And if there is, then they need to decide, okay, let's spin up an initiative team and decide who needs to be part of that initiative team. To then to resolve the governance issue.

And then it says anything that needs to be escalated further, so they find any issue Oh, hang on a sec, this actually might impact multiple domains, and it should be escalated further, or maybe if it's not possible because of global standards, then they have to excalate this. Because there are things that they can anything that can be resolved within a domain will just stay in the data working group or like within the domain level, but anything that they can't resolve, or that they need extra help from that it should go to that data counselor. But like, Yeah, I'm just gonna say this is like very careful agreement, like because we needed to come with a starting point. So we're going to test to see if this, this works. But I think the main aim was making sure that because this is so new number one, that the central team, or the CDO could help guide the people, but also, I guess, provide that championship and like that alignment, for domains to help them build that capability so that later they can kind of run on their own a little bit more.

Koen Hendriks 28:01
So the communication is also really important?

Telecom consultant 2 28:02
Yes, I think data mesh is good. But I think there needs to be some sort of like alignment within an organization to make sure that people aren't doing the same thing. Yeah. And that there are, you know, efficiencies where people worked together or shared ideas or something like that. Like when you are searching for a data product in the data market place but it is not there, you don't need to create it but ask whether you can create it or whether it is already there. This prevents people from performing double work.

Koen Hendriks 30:01
Yeah I can imagine that it is really important there also to have correct metadata and govern the marketplace. This is part of the Self serve infra right?

Telecom consultant 2 30:10
Yes

Koen Hendriks 30:19
Who is responsible for the self serve infrastructure?

Telecom consultant 2 31:01
The CDO is the owner and has their own team for the self serve infrastructure. At this moment they have a central IT team that gives uit resources. So for the pilot this is feasible but for a business as usual perspective, they need to figure it out. Within the domain there are 2 types of roles, the more data governance roles that manage and monitor the data quality of the data that the domain owns and the producer team to create the data products. Both of these teams should be represented in the data governance council.

This way, there are like a lot of people in the council. Which actually is a real challenge in data governance. Because according to the DAMA-DMBOK it is really straight forward but when you add humans to it, you add a layer of complication especially with data mesh theory. So how to get the mesh of theory and practice together would be really interesting and helpful.

Koen 35:51
Yes, I think I will look at the differences between theory and practice to help with making the implementation of a data mesh more practical.

Telecom consultant 2 36:10
I think one of the main things is that the framework should be a guide and we need listen to eachtother because we as consultants have expertise from the best practices and our experience and the client understands their way of working and their requirements. With this we need to figure out together where we can meet in the middle and encouraging them to follow the best practices, but if they cannot immediately do that, we need to work out a way for them to get there or a way to be able to communicate why it is important. So a guide could work or a list of considerations with what we need to consider while doing this, or while codesigning with the client. Even if you propose a framework it always needs to be adaptable. And also make sure when doing this that you are not too theoretical. That is the feedback we got from our clients, that sometimes when we discuss parts of the data mesh book, we are too theoretical. So what does it actually mean and how do we do this? And how do you approach the design and the implementation. It is a danger that after the design you don't now how to action it.

Koen 39:10
Yes I understand, the time is also up so we have to wrap it up now.

Telecom consultant 2 39:40
Yeah true, I think we have discussed everything and again the sponsor where we talked about is really important because otherwise your project will end up nowhere.

PIE3 Problem investigation interview Data governance expert 2

Data governance expert 2 00:02
I'm Data governance expert 2 I started at Deloitte last October. As a manager with the financial risk analytics and reporting team. It's an financial services and insurence focused data and risk reporting team. Yeah. Before that, I was with a company for about eight years. Very similar stuff as what they're what we do currently in the fraud practice, but it's much more organization. With that experience, I started as an application developer with

Company for a product called product. And then I became an implementation consultant. Then I did a master's in financial engineering, and then I moved into consulting. My first project was BCBS 239, which is a reporting regulation. It's also very important when you're considering data mesh, especially for financial service organizations. So it basically says that you need to have certain level of data lineage certain level of data quality, certain data dictionary things that you need to document. Certain meta data you need to document have those data governance in place, be able to address generate a report in terms of stress crisis.

Koen Hendriks
Is that some kind of standard? Or is it a concept that you use?

Data governance expert 2
BCBS 239 is a directive? Okay, that it's a puzzle regulation. More ECB will have to translate this into requirements, right. But they haven't done that yet. But banks are still required to comply with BCBS 239, though it's not a requirement yet. That plays a big role when it comes to data governance and having those measures in place. So that was my first project as a consultant. Before the data migration client. I did that for two years for liquidity risk and capital risk. That was my first exposure to data governance as well. And then I went to financial company to set up their regional CDO office. The regional CDO Office Chief Data office. Right. So, you have global CDO office, which is a CDO himself and they have like the Data Quality Officer, data governance lead in our head of data management, different roles. So they are exposed sponsible to, you know, to implement the data strategy for the organization. So you have the global city offices, but we needed a regional city office, which we helped set up forFinancial company 1. And then we also had to appoint data owners and data stewards for all their regional offices like Korea, Taiwan, Japan, right and Have we had to translate requirements from from EU to what it means to Asia? Because everything is not as-is taken to create local regulations into consideration and come up with new requirements, new minimum standard.

That was for a bit. And then I worked on product again, developing product for managing external data. We just called it external data management system. It was like a marketplace for external data. It's a marketplace itself. So for example, Bloomberg supplies, data, to start chat with about 18 Digital licenses, for example, it's way more than that. But then if you consolidate it as a group license, you have added value, right. And these 18 different teams work in silos. Doing their own cleansing operations, did their own aggregation operations and everything, we brought a common system where they can share that new version of data and consolidate effort, maybe save license costs. And this was not just Bloomberg, they had about 150 data vendors, from whom they were subscribing data. Every time they go for new data, when the time to market was like three

to six months. First understanding what the data is going through the documentation, seeing if it's useful, then contracting and negotiating a contract. And then setting up the infrastructure to receive the data and then stage the data. And now it's very long process. And even with a lot of effort, it's still three to six months. So making all that streamlined. Having a data owner for every data that is subscribed, or next time, you want the same data for a different team, don't go to the vendor, but go to this data owner so that he can expand on his contract, renegotiate what they're doing, like a very, very, very successful useful products project as well. Yeah, indeed. Interesting project as well. We did that for about two years, before I moved here, and something entirely different for a different client here. More on product migration. And then I worked in LeasePlan for a bit, setting up their data management policy. So lease plan, implemented data mesh 2020. And they didn't officially go with the data mesh label. Right, but they have the same governance structure. However it was not a strict implementation regarding al the other principles. They initially had that discussion and lease plan wanted to become a data driven organization, and then they had a few projects around it. I don't think they loaded it. For them. It was a different documentation, which helped him get that. Right. But then they loosely adopted the principles and moved on without focusing on being data mesh. Where I went was to redo the data management policy. When they did the audit, they found that they don't have central policy to cover the BCDC 239. So they wanted a data management policy, but when we drafted it, we help them cover the entire aspects of the DAMA wheel.

Data governance expert 2 09:44
Now the policy had 12requirements. So you know, winter data requirement data dictionary quote, I'm in detail in his requirements, data architecture requirements, data quality requirements, all the 12. Yeah. And within each section we had minimal requirement and a guideline requirement.

Data governance expert 2 10:18
And who did define those requirements? It was like you were done the external team that defined them for them.

Data governance expert 2 10:25
Correct, we help them we facilitated those discussions, sat together with their own data management team, their domain architects, right. So they have about 20 domains. And for each domain, they have domain architect assigned, that they're responsible for getting there. Similar to it, is the damage not similar, right? They're responsible to get the data into the next generation. So they call they have something called next generation data architecture (NGDA). And they had 19 domain owners who are responsible to get their data into the NGDA system. So we sat down with them in those meetings and then we had the data to Enterprise Data Architect, we had the AI Head, head of AI, we had the data privacy officer, we had the second line wich was just risk. So we had non financial risk, but we also had risk control, functional head.

Data governance expert 2 11:43
There was like one big group of all different architects, all different domains, or domain architects, and also this data governance lead and people outside of the data governance capabilities like risk and things like this.

Data governance expert 2 12:05
because most organizations, they have first line second line third line, are you familiar with that? Like strategical tactical operational? That no. So the business is the first line. And within the business itself, then they're supposed to take care of data privacy data governance, so this business and support functions to run the business are the first line and then there will be a second line which is compliance and risk. Their only job is to challenge first line right, slow down the process, but make sure that whatever they do is foolproof and you know, has all the cuts has all the considerations taken into account and everything and then he had a third line which is audit, which randomly picks different functions within the first line and audits them basically and use their findings. And whenever you have like a federated governance structure, we say just simple model validation or something like data mesh central processes, like data management policies, right. So whenever you have a group wide policy or something second line needs to sign off as well. Okay, you need to convince them that this is the approach you're taking business models based. And then how this is the risk of that is how you're managing the risk and these are the controls we are putting in place to manage these risks. So secondly, needs to sign off. So this is still centralized. So, we helped them draft data management policy and we had all these people in one room right. We started with the DAMA framework is the baseline for this policy. Right. And on top of this, we have taken the six regulations into consideration and other requirements set by the six regulations into consideration. And whatever findings we had from different issues, different audits, different control checks, which has all happened in the past, right. So Taking all this into consideration, this is the policy we need to have in place for federal governance, for centralized data governance, for data management. Okay. Yeah. But we I haven't been an official part of data mesh implementation yet. Right? I've been working on the theory, I've developed some point of views been working on a central COE within risk advisory for data mesh. Yeah. And I'm well versed with the concept. But we haven't done full data mesh implementation yet. We have some sessions coming up. But nothing as of now, we have a project where we are requested for an inspiration session of data

mesh, like a half day workshop, and then we will provide them with a high level operating model, so that they will discuss it within their organization and come back to us if they want to go to the full implementation.

Koen Hendriks 16:04
Have you already designed that high level operating model?

Data governance expert 2 16:09
no, we haven't. So the process for that is we'll have a standard practice session to understand the current data architecture. In terms of processes, like you have current policies in place, do you have current controls in place? Whatever. Right. And then people, what is your existing governance structure? You know, do you have already had it on data stewards appointed? Are they functional? Are they exclusive? Or their renewal role with the current business as usual responsibilities? . So we have questionnaire like that and then technology wise, do you have data warehouses? Do you have our data Lakes? What technology? Do you use? Are you on cloud, are you on premise? So based on that, we will go with the inspiration session, which talks about we'll start with the four principles of data mesh, right. And then translates it to saying, this is where you are and this is where you need to be to get your data mesh approach.

Koen Hendriks 2 17:17
That is really interesting. Because you said that you first analyze the current processes, current responsibilities, the roles, the structure there, then you to take like the full principles of data mesh as you call it. And then you combine those together to make like a structure that fits their organization while being aligned with the data mesh principles and relational database capabilities.

Data governance expert 2 17:56
Yes, In some organizations, they already doing data mesh without knowing it as just a good practice. It doesn't have to be a complete shift. What I mean, and some might not apply. And we'll also tell them, what are the challenges if you're going to save to this? Right, from what we see? Right? What are the hurdles you will run into in terms of the technical implementation? And possibly see that these are the solutions you need to start from day one. So that's the exploration session, right, where we just discuss these things. After that we have another thing called data mesh readiness assessment, which is a much more detailed questionnaire. Then we combined this with a maturity assessment framework, which gives us a data mesh maturity score. And then we discuss that with the client and propose a high level operating model. That describes how it would add value to your organization.

Data governance expert 2 20:19
Before the readiness assessment, you put together the current processes and principles of data mesh. What kind of what do you define as the full principles of data mesh, that you're going to use? To better with the current processes and roles? Do you have like, resources for that? Or is it just the theory book of Zhamak?

Data governance expert 2 20:50
The Theory book of Zhamak, but we have created resources from that. This resource translates the four principles into organizational shift, architectural shift, technical shift, operational shift, and principles shift, which is already the first shift. Right, and then defines it into what technology will need and what will change? And for each of them pros, cons. So yeah, that's the starting point for us. And in terms of the questionnaire as well. We haven't zeroed in on the exact questionnaire, but we have a current draft, which I can send it to you. Right, there will be also nice. Yeah. Which is just to understand their current data architecture. Yes. And then those questions are also designed to get specific answers that are needed for a data mesh implementation right. Not, the question that I'm going to send to you. That is to understand, which is relevant, you know, we don't want to just keep talking about Zhamak, that I don't want to just keep talking about Zhamaks theory to our stakeholders, right, that .they

can just go online and see, our role is to make it more relevant to their infrastructure. So what we're trying to do is based on our experience in working for multiple clients, trying to capture the essence of the whole organization, just 30 questions. which are just the basics questions. Like: do you have data leaks, or a complaint from Do you have things on the cloud? I plug compliant? What tools they use? They use data lake houses to use data bricks, right?

Koen Hendriks 2 27:01
Did you run into accountability and responsibility issues during the design of the data mesh (like) architecture? Sometimes you hear that the central team is accountable for responsibilities where they do not have any influence on.

Data governance expert 2 28:40
We changed that already. So, data mesh doesn't keep the central team accountable, right. Data mesh, makes the domain owners accountable as well. it's an obvious pain point that has changed. So for example, what we do is with Financial company 1, you have the CDO and the regional CDO, right. So the regional CDO is supported to get there. We will give you the resources tooling and things to get to get you to the desired state of data governance. But then we have a data protection executor who is also the country head.

He is the legal representative and every document. This is before data mesh. This is just how federal governance works. So the DPE is actually fiscally responsible for the organization. He's normally the country head, but he's also given the title of TP. And all the data owners are the department heads like the tax Tax Department had the HR function, return regional head of HR. So those department owners were the data owners. They were accountable. And they report in the data organizations to the DPE, who's the country head. Right. So now, the problem is back in their plate. What the CDO does is only support them. Right, we'll give you the tooling, we'll come up with the requirements, we'll come back to the regulation, we'll come up with the drafting the policy and putting those things in place, but you are responsible. And, for example, they are also responsible to get the data into the central data lake, for example, right, to the common platform. And they have KPIs on how much of the data is, is already in the lake. Right, they have KPIs on what's the data quality within the domain and the lakes, but they're not responsible to do it themselves. It is the CDO, who still helps them do it and defines the requirements of the KPI? But they had an executive data Council. Yeah, right, which was all the C-suit employees, CRO CIO CAO CEO and country heads and, everyone. And from there, we define these KPIs and then trickle it to the system and get it approved and implemented. Okay. Yeah. And this same happened standard Chat started off as CDO organization with like, 60 people in the CD organization. Right. That was in 2017. By the end of 2018, they had dissipated the CD organizations like 10 people, and the responsibility was implemented in individual teams. Yeah.

But they're still, you know, they're not data people. So it's a cultural change. So the first time they're like, We sent them an appointment letter, saying that you are the data owner, right? These are your responsibilities, these are your KPIs, and then you will have a big session, where we'll give you a crash course on data management and all the responsibilities of data and then we tell them, so you need to actively engage your team, the IT counterparts of your team to to do this. So you take Tax, for example, which is a very simple function, you will have 16 or 17 registered IT applications, which they use the team. So, for each of this application, they'll have an asset owner. Who will be the business will be the taxing. But each of these applications will also have an IT custodian. Who will be it that's, that's just like a data custodian, but then for the it itself, for the application, you know, of the application itself. So he is responsible to maintain the application have to upgrade the applications that support the application, in order to keep the application live, he's responsible for this issues in the application. Right, so each application has an application owner, sorry, as a donor, and it custodian right now is totally apart from the from the as Zhamak calls it, the self service platform. That's only the self service infrastructure. That's like owning the infrastructure, and then the it counterparts. I like to applications itself.

Koen Hendriks
So this is a little before zamak, right?

Data governance expert 2 02:21
yes, data is very important. But organizations were not built around data, it was built around the business. Yeah. Right. So they had this infrastructure between this this law organization between IT and business already.

Yeah. So back then we didn't have that self service platform, it was still a central data lake. So they were responsible to get it into the central data lake. So you, you tell them, you know, so these are the 14 functions, 1515 applications, IT assets within your department. And three of them are already planned to be in the next cycle of data lake projects. Right? But these are not. So that should be in your agenda for the next cycle. Right, you need to push it as a custodian, I guess it custodians, it your storage to to get their data into the lake. But it can't be a one side initiative where the central team, for example, for data mesh case, the self service platform team contacts you and pushes you to publish your data, you need to take the effort from your end, because that's a KPI for you as well.

Koen Hendriks
And that was difficult for people that were not like in the data culture?

Data governance expert 2
Correct. So when we start those discussions and have those workshops, it's a surprise for everyone. Like, you know, I have the tax department, I'm not a data guy, why are you talking to me about meta data and data dictionaries and getting the data into the data lake and all those things? You know, do I get a special headcount for this? Am

I gonna get extra budget for this? And those are the discussions.They start with. We say that No, no, we will support you. You know, yes, you will, if you if you think that is the case, but

then you need to first decide what needs to be done and create a separate called Project kickoff cycle, logic CAC and then get the budget and that releasably discussion, right, but then you educate the current people who are already part of the business to business requirement.

Appendix E

Case study method for practical knowledge extraction

TABLE E.1: Additional observational case study methodology

Step	Activity	In research context
Getting started	Definition of research question Possibly a priori constructs Neither theory nor hypotheses	Which design decisions are made, what influenced these decisions and what were the results of these decisions? Team structures, roles and responsibilities and allocation of decision making. Because of the limited availability of theoretical resources on design decisions in the data governance of data mesh, the bias from theory is limited. However, it is important to take the bias into account that could result from the comparison between the practical and theoretical structures. Therefore, the mapping of these structures must occur naturally with limited effort
Selecting Cases	Specified population Theoretical, not random, sampling	Organizations that adhere to the domain-driven design principles, decided to migrate towards a data mesh architecture, used the theoretical data mesh structure to design a data governance structure, cases from different sectors to avoid a bias towards a single sector and have available resources for analysis. From this specification, the availability of resources is considered the most important. Even though the two cases both follow the full specification. Cases sampled that have made decisions to deviate from the theoretical structure
Crafting Instruments	Multiple data collection methods Qualitative and quantitative data combined Multiple investigators	The data collection method used in this study are archival sources of organization structures, and semi-structured interviews that discuss the archival sources and answer additional questions about the design decisions to come to these structures. Because of the complexity of the organizational structures and the lack of availability of quantitative data, mostly qualitative data is used. From each case, multiple team members are interviewed to include different perspectives and enrich the data.
Entering the field	Overlap data collection and analysis, including field notes Flexible and opportunistic data collection methods	Data collection, is combined with data analysis to Improve the understanding of the case. Important findings during data collection are analyzed and implemented in the data collection method. This was possible due to the semi-structured nature of the interviews.

Table E.1 Continued:

Analyzing data	Within-case analysis	The interviews are transcribed and coded through open and axial coding. The cases are anlaysed based on the structural mechanisms of Abraham (2019). Statement about processes, roles, responsibilities, and team structures are identified and grouped together to form descriptions of these structural mechanisms. These descriptions are related to the theoretical structure to describe the made design decisions.
	Cross-case pattern search using divergent techniques	The similarities and differences between the cases are listed to form new categories of design decisions and their resulting structure.
Shaping hypotheses	Iterative tabulation of evidence for each construct	Important similarities and differences regarding the design decisions are generalized and internal constructs are explained.
	Replication, not sampling, logic across cases	The generalizations are validated both with the provided sources as with a validation session that gives the case respresentatives the opportunity to confirm, extend or sharpen the generalizations.
Enfolding Literature	Comparison with conflicting and similar literature	Compared with the data mesh book
Reaching Closure	Theoretical saturation when possible	The iteration process stops when the incremental improvement to theory is minimal. Due to the availability of resources, the research will not be extended with additional cases. The saturation within these cases is approached through additional interviews with the possibility to add more design decisions.

Appendix F

Treatment design and development interviews

F.1 Treatment design interview protocol

See the next page.

Treatment design interview protocol

Introduction
In some cases, the data platform architectures implemented in organizations are reaching their limits. They are not able to manage the increasing scale and complexity of the data used within the organization. A new socio-technical system called "data mesh" aims at increasing this manageability by decentralizing the organizational architecture of data platforms through the division of data responsibilities across different business domains while ensuring cross-domain interoperability through a governed self-serve infrastructure.

The migration towards this new architecture is often complex and requires a well-formulated strategy to increase the odds of success and potential profitability. The design method for the strategy, and the strategy itself is likely to differ per organization. One component that is commonly part of the data mesh strategy is a roadmap from the current governance structure of the data platform towards the new governance structure that incorporates principles of a data mesh. Before being able to create this roadmap, it is important to design the target governance structure to know where the roadmap should lead to. This process can be done in many ways and supported by many kinds of models.

This research investigates governance structures within a data mesh and tries to improve the process of designing the target governance structure within the future data mesh of an organization. This will be done by developing an operating model that can be used as reference model to describe the governance structure in a data mesh. This is defined as the framework regarding rules, procedures, roles and the division of responsibilities within the decision-making process within the Data Mesh. This framework addresses important components in the governance structure of a data mesh, the relations between these components, and the processes in which these components and relations are reflected. Examples could be descriptions of specific roles and responsibilities (example: the difference between a data owner and a data product owner, or the difference between local and global responsibilities), technologies that are used in decision making (example: definition of data products, data catalogues and self-serve infrastructure), or processes (example: the decision making on standards and policies or the usage of data products) within the data mesh.

Goals of the interviews
The interviews have the following goals:
1. Investigating the process of designing the target governance structure
   a. Formulating this process
   b. Explaining the challenges in this process
2. Investigate the model usage during this design process
   a. Describing the models used in this process
   b. Formulating this interaction between reference models and the design process
   c. Define assessment criteria for the models
3. Analyse important components of reference models in this process
   a. Define important components of reference models
   b. Formulate improvement opportunities for current models

Practical information
Date & time: To be announced
Duration: 45-60 min
Type: Semi-structured Interview

Interview Script

1. Could you introduce yourself?

2. What kind of data mesh project have you done?
3. From your experience, what are important design decisions that need to be made on organization structure?
Teams structure:
Roles:
Responsibilities:
Capabilities:

    1. Where are these design decisions based on?

4. What are organization structure components from the data mesh theory book that are required?
Teams structure:
Roles:
Responsibilities:
Capabilities:

    1. Why are these components required?

5. Case specific
    1. Do you have a specific case which could be interesting for my research?
    2. Can you tell me more on how the roles, responsibilities, accountabilities and team structure is of this case / share the resources?

6. Principles and best practices
    1. Can you give any examples of solutions or best practices?
    2. Can you recommend any best practices?

7. Do you have any other important additions to the research?

9. How would you describe your experience with data governance operating models for data-mesh like/ federated architectures?

10. Project status
11. Can you describe the current status of the project?
    1. What was the starting point?
    2. What is the status of the project?
        1. Finished → How long did this process take to finish?
        2. Not Finished → How long did it take to come to this state and what are the expectations of the remainder of the process?
        3. What was your role in this process?

12. Result:
    1. What models did you design for the operating model?
    2. How do these models look like?
    3. What are the most important components of these models?

13. Design process
1. How did you design these models?
2. What components were difficult to design?
3. Did you use reference resources for this model?
4. NO--> why not?
5. Yes --> What kind of reference resources did you use for these models?
6. Did you use design principles for this design?
7. What were important components of the reference resources?
8. What were the key learnings of the design process?

14. Finalizing the Interview:
1. Do you have anything to add to this interview?
2. Do you have any other questions about this research?
3. Do you have any recommendations for contacting an expert on this topic?
4. Are you interesting in the results of this study?
5. Would you like to have a review possibility of the results of this interview by receiving the interview notes before publication?
6. Can I contact you for the evaluation of the developed model?

This is the end of the interview, thank you very much for the time and effort.

F.2 Treatment design interview protocol

Design and development interviews

TDIT1 Design and development interview Telecom 1 Telecom consultant 1

[...]
Koen Hendriks
Yeah, let's. Just try a question per question and we. Can see how far we can go. OK. So did you use design principles? For the data governance design.

Telecom consultant 1
So specifically for the operating model, we use the design principles because I think I share it with you. As well, yes. There are a couple of statements. That, yeah, brought this new ones between certain topics, centralized, decentralized, supply driven demand driven it driven business driven coordinated self organizing where you have the full list and that is what we first used to establish a. So in that way. In general, we use those design principles. The first step to agree upon. A set of basic principles that everyone believes in as a basis to make the choices later on in the process, and those social later on where for example the RACI or exact responsibilities and accountabilities those kind of choices. But we started out with the first workshop where we discussed those topics and basically scored these principles that we don't want that would be contradictory to each other. We score them to either one side, the middle or the right side, and also the future ambition there.

Koen Hendriks
Where did those principles come from?

Telecom consultant 1
So we used a workshop for that. We did create them. For that workshop. I'm not sure where they came from so something like centralized decentralized would make sense to have that kind of choice. Bbut maybe I'm not sure if that comes from a model. I can imagine that we use the data mesh book for that. So the idea we already had data mesh in mind. So I think we used some of those. Maybe augmented them, maybe with some existing models. To to come up with these choices that we had the audience vote. On the clients.

Koen Hendriks
And do you think that those principles are organization specific? The principles or the the to actually the horizon like from one one point to the other. If any other organization would have different horizons on that.

Telecom consultant 1
So I think the choices are similar. For example, choosing how much you do centralized decentralized. But the preference can differ per company. So whether you are very centrally oriented now or not, or people working in silos and have a lot of autonomy within their own silo, and then they have to work together, they don't want to give up this autonomy or companies that are very open sort of like a ecosystem on itself. They're maybe more. Comfortable with it, so I think. The choices can be applied to multiple organizations, but what choice they make can be dependent on the organization.

Koen Hendriks
OK, nice. And you already mentioned like the centralization and some other components that should be covered by the design principles, but if you would come up with like? Four or five key. Components of design principles. What would be the focus of that? Only looking at the data of governance and.

Telecom consultant 1

Central versus decentral on a lot of levels. So responsibility, accountability. Who can make the choices. And how that cascades into the different domains? Also, I think on organization design so. There can be a lot in a central IT organization. So the size or more or less of the central IT organization or it can be Federated? Or you can have a choice in having a lot of central resources in sort of a consulting model. So and maybe a good principle on that would be sort of business versus it. Or so. Maybe you can have ownership at at the business the different businesses that can be data domains, but you would have a separate IT organization supporting those businesses. So what your operating model be business focused or it focused? I think also an important choice that we had is sort of this. Variations of having a center of excellence, having a consulting model. Having all the different skills and roles within the domains. So this is also a choice that you can. You can have. There might be. Maybe sounds vague, but so you can have. Also in line with. What we discussed on responsibility and accountability, so you can also have an organizational structure that is. As many layers versus few layers. So I guess there's. Also a choice to be made there.

Koen Hendriks
And then you mean like domains sub domains, sub sub domains and then like a pyramid of? Accountabilities and responsibilities then.

Telecom consultant 1
Yeah. So you can have teams. So I think there are a few choices there. So it's also on. How big those teams are for example, so how large? Is your it or is your IT part? Of the business domain, does IT support a business domain? Then there's also. Add the amount of layers and also do I have a CIO with underneath? A center of excellence with underneath the different themes. So it's not only the size of the different IT and business teams, it's also the degree of structuring those and who reports to who?

Koen Hendriks
And also some that's actually the combination of the responsibility, accountability and the interaction between the teams? About your case, you said that. Your business and IT was kind at the same place when they started. So did they did not have like a single IT department that supported them?.

Telecom consultant 1
Correct. Yeah. [...]

Koen Hendriks
What is design principle that is the most important to you?

Telecom consultant 1
Yeah, there will maybe be central versus decentral in a fairly generic way because some things you need to keep central such as your self-service infrastructure. Some standards that you pose, there are some choices that you need to make in order to provide freedom to the domains. So start with those choices. In our case it was which cloud provider? Would be easy if you pick one. You could also do 2, but no more than two, so simplify your landscape so you can make it interoperable and you can have to also make their own choices. So I think that was a very important one. So where do you leave freedom and where don't you? So we said from a data producer perspective. We want to limit the choices. So that we make it easier to make it interoperable. So to all use the same cloud provider, use some sort of landing zones or preconfigured set of services on those cloud providers to make it easier to produce your data and also make it available to others. And then we said on the consuming side, there might be very specific use cases. So one domain might meed tp run their own scripts, so there we don't want to limit them. So pick the areas, the key areas that were important for us with producer, exchange, marketplace and consumer. And define the degree of freedom for each of those, and I think the same holds for an operating model. Like there should be someone that guides certain choices. It can be on producer, consumer level, but there you need to to start with. What do we need to have central. Where do we need freedom and where don't we?

Telecom consultant 1
Yeah, in I think then in. The governance operating model so that you would, for example say that. If it is really centralized. Then you have very strict data quality. Standards, for example. And when it's really decentralized then.

So from a central point. You would, for example, say all the data needs to be in a certain format. For example, to make it interoperable, so not necessarily data quality standards. More on format and those interoperable requirements. And and then you would also say. So maybe you would enforce to have some information on data quality. So there could be some basic checks. That you would define for. For all your different data products. A complete list would be an easy one that you can check. Or you could leave a bit of freedom, but at least make it mandatory that you share the data quality requirements that you set yourself and as it's clear to the rest of the organization, this is what is what is

important or what what we checked for in our data quality. So this is more like a checklist that we had for artifacts for data products like what does the data product need to have needs to have information and quality metadata, data sharing agreements and when can you use it? When not stuff like that.

Koen Hendriks
Where those things defined during workshops or on yourself?

Telecom consultant 1
Bit of both. We used the book data Mesh book, but also some reference. Materials from other clients. And then check this with the architects. And engineers data engineers. So we had like maybe 20 potential artifacts and then we picked the most important ones. Maybe from the audience we added. Two or something and. That picks from the list of 20. And end up with 10.

Koen Hendriks
And before jumping into like these specific problems. If you look at your clients case, they did not have their governments in place already. And and one of their aims was also to get, like a central or get a global data governance to make it less tight and more interoperable and. Ensure data sharing. That's like if you could also have implemented like a normal Federated governance model for that. What is in your opinion the difference between a normal federated governance model and the Federated governance model of a data mesh.

Telecom consultant 1
Yeah. So. Normal Federated governance model. Yes, I would say the data mesh is a fairly good described version of the Federated governance model. I don't know where a normal Federated governance model would be different.

The most important difference is that. There are new roles like data product owner. That, like, incorporates the other principles into the governance principles. Like product thinking and domain driven design. That's like at least what I understand of the difference, and it's just not a form of a federal government model, but at least with more business allignment. Yeah, I know that's true. That's true. So the current situation could be seen as a very, very Federated governance model where everyone could make their own choices. So what they had in their current situation is that people are using data which they don't know where it came from and. That was also a huge problem for for them. So this product thinking made a lot of sense to them to have this single source that you can trust and to use the data from the source that you know where it came from. And if you have any issues, you can go

to specific data product owner and address those issues. So that was something that they really liked and. Was a choice for them to do so. Because this would be still a problem if they're just only followed a Federated governance model. This problem with data lineage like knowing where the data comes from and knowing who owns it and. We can fix your issues, but still something that they need to do. So this was like two wins. In in one model.

Koen Hendriks
OK, so let's go to the specific problems. We have 13 minutes left. So you can go quickly through them I think. Were you there while they were defining? The roles, responsibilities and accountabilities.

Telecom consultant 1
No, no, I wasn't there, but I share with you the workshop deck that they use for that. So I heard something about it. Maybe Telecom consultant 2 was there. But the large problem that they had there. Was what we discussed before, so the responsibility and accountability. So I have peoplein the C layer that were accountable for choices that were made in the business. Without being in their reporting line, so there's also if we discuss. So if your boss is responsible and you are, if your boss is accountable, you are responsible. That can be fine because you report to your boss anyway, but it will be totally different domain, so one will be in the IT domain. That would be accountable and the business would be responsible. So the business makes all the choices and IT has to. Fix it. There will be something that will be unworkable because there needs to be some. Agreement there between the two, of course.

Koen Hendriks
And how did you? Solve this problem.

Telecom consultant 1

So we had the service catalog workshop where we had all these services and we would say where would this service reside with the CDO? Would it be within the domains or would it be within the IT organization? And after our first round, we identified which one of the combinations were strange. So like we just discussed. So our responsibility is somewhere else account with somewhere else. Then played it back to. The leadership architects that were involved in these choices and then re. Visited choices and and made them differently. So maybe maybe. A two step approach in that way standardized wouldn't be that bad, right? Or you need to limit the choices that certain choices are combined. So if a set of services has accountability at one team, it automatically means that the other team that is responsible. Must be within the same line of reporting, for example, so there can be choices that you can limit up front.

Koen Hendriks
Does it also mean that? In the first workshop, you. The lead architects weren't there.

Telecom consultant 1
They were there. They were there. But so those services, they came from a general reference model that would build on a central IT organization. So in the central IT team organization you don't have this issue because everyone reports to IT. So responsibility accountability is closely linked. But if you split this over business and. It and you. Still have some separate it separate business teams. That make different choices and are allowed to. Then those services cannot be separated sometimes.

Koen Hendriks
And during the workshops. Did you get an output and the output you checked with people from the same workshop and then changed it?

Telecom consultant 1
Yeah, because if you play it back, you just reiterate, you will say an extreme case. In this extreme example, they would recognize like indeed that this may cause problems.

Koen Hendriks
OK. And something new from the previous things that we aren't going to rate this complexity or importance? If you look at the. Solution that you had for. Then only defining the responsibilities and accountabilities. How complex was this problem for you to solve in this structure?

Telecom consultant 1
I would say it is a 5 so. It's relatively easy to propose a different set of responsibilities, accountabilities, but then to make it fit within the organization can be very hard. So that was a real challenge. So we only. Did a follow up workshop. But then, if this organization is really used to a certain way of working, also convincing them and dealing with that can be very, very hard.

Koen Hendriks
So actually defining the roles was really easy, but fitting it to the organization was mostly difficult?

Telecom consultant 1
Yes.

Koen Hendriks
Nice and the importance of. Solving this problem.

Telecom consultant 1
5, what you would see now is that it wasn't defined who was responsible. So in the end. The CIO would only make

a choice if everyone agreed, so it wasn't really defined. Who should make choices? They have a formalized decision structure. It's very important if you want to make any decision. So in that way, this is really important.

Koen Hendriks
OK.

Koen Hendriks
Yeah, this is more of an organization fit type of question that you identified probably capabilities that you want the data governance and people to have. You need to make people responsible and accountable for tasks that they normally don't do, so it can fall outside of their normal set of capabilities. Did you encounter this problem and how did you solve this?

Telecom consultant 1
Yeah, definitely. So in the end, we propose to do the training. So the DAMA-DMBOK training. With the most important stakeholders, so that would fulfill a key role within the future data organization so that they know what is meant within the governance, what is meant with metadata management. It's internal training that we also want to give externally now for the first time, so to really. Make sure that that it fits also, then part of it sort of the train, the trainer model where we would just give it to the. Most important people at the. Client side, they would adapt the training to the organization, so make it more specific. Also adapt it together with us to data mesh because for every chapter of the book you can make a specification for data mesh because things won't be as described in the book. So we need to adjust that as well, and then they would train other people with an organization and that way we wanted to improve the maturity there. So it was also part of it. So if you knew, roles were created. Specifically for this, so that is true. But also they had a lot of roles already which were doing something in a more central way, so they did have an in the specialist, but this this person was good, but they just need to adjust to data mesh situation. So this dispersion known in the organization already is used to his or her own silo, or current way of working. So if we can gather those people on board, we'll be better for ahead to to have an ambassador of change because. This person knows organization. I am. And also would be then more suitable and quicker than finding someone externally.

Koen Hendriks
But if you do, you have then for example look at master data. If you have one master data specialist in the organization in one domain. He cannot be the master data expert in all domains.

Telecom consultant 1
No. So that is true. So we wanted to start with one domain and then scale to more domains. So in the end, we envision that in three years when the whole organization would move to data mesh. There will be 200 + data related roles within the domains so. A large part of. This probably would be new roles. Or some roles in smaller domains that people would pick up part time. So maybe a data analyst that would also be MDM specialist for 20% of his or her time

Koen Hendriks
And the training of those people is done within the organization?

Telecom consultant 1
Yeah, that one would be train the trainer. So we would have one master and the MDM expert in the Central Data Office. And with the onboarding of every new domain, he or she would do the training or someone from that team would facilitate the training.

Koen Hendriks
OK. And and how complex was this? Problem to solve.

Telecom consultant 1
So in the end it's not really solved because. We didn't do the training yet. The idea was good. Yeah, but yeah, then again, to make it fit, to make it work can be very hard. Again, say. I mean, it's working with people, so it's very hard. So I would. Say 5. But, it's kind of difficult to scale complexity because in the end, what we also thought was that we were doing this vendor selection for the marketplace and then implementing the solution. But the technology is never the problem. It's the people working with it, so training everyone and getting everyone on board is actually the hardest part of this whole transformation. So, for example, the vendor selection would be a 2 or 3 and this will then be a 5, just to make sure that not every problem is complexity 4 or 5, but that the more easy problems are also accounted for.

Koen Hendriks

And what's the importance of solving this problem?

Telecom consultant 1
Also like a five

Koen Hendriks
So about the interaction model, or at least the different kinds of meetings that there are within the organization. Sometimes people have problems with the interaction model, for example, coming up with who needs to talk to who for decisions on standards or the prioritization of data products. How did you define the interaction model and did you run into problems there?

Telecom consultant 1
Yeah, like what you said, we proposed a regular interaction model that includes a governance council or some regular check-ins with the data platform owner and domain owners to have that alignment. We also discussed a consulting model, which might be part of it. They had a large sense of excellence in how they could lend resources to different domains, so that was an important part of it. It's also part of the service workshop where we just said, "Okay, now they have a large central IT team. What to do with it?" And we can go a couple of ways. We can dissolve it into the different data domains or we can do this consulting model where we keep them all in one team but internally consult them within the domains.

Koen Hendriks
And there are a lot of choices there, as you said, because of the model or send directions. Like all these choices that you made to eventually end up with that structure, but making those choices was that difficult? And how did you make those choices?

Telecom consultant 1
Yeah, if one agreed that there should be a central data platform, there was some discussion on ownership there, but everyone saw new things coming, such as a marketplace, this data platform. So it would be easy to have the central organization, this new team that would own that. The new data mesh specific products, which aren't actually that specific in your case, came with the data mesh implementation. But then, and this might be slightly different topic, if you have an infrastructure team currently, it was harder discussions. What will be the future of the team? Where will it be placed? But it's still be the same team, will it be a consulting model for those people as well or will they move to the domain, so that was a tougher discussion.

Koen Hendriks
Didn't they move to the infrastructure, like that? Didn't they move to building the self serve infrastructure then?

Telecom consultant 1
Yeah, true, true. But currently, they also had all this data engineering skills, which you would rather have in the domains. Purely theoretically, there should be only infrastructure, but they did much more. They were also creating data products. So you would have preferably a split there, so either the people that also create data products would be in this consulting model that would support the domains, or they would be in the domains. And this is a tough discussion because people want to keep their teams in influence, yeah. So that's still not really tackled, I believe.

Koen Hendriks
Okay and how would you scale this problem?

Telecom consultant 1
No one really disagreed. So you will have a central team that will support with this and this, okay, you would have a governance team that sets standards, okay, that was quite agreeable. So I would give the complecity and importance both a 2

Koen Hendriks

22, then I will do that. Also talked about Telecom consultant 2, about the interaction model because she. I think she was. She just made like a slide to present to the client. This is one of the things we already talked much about. The story about the equilibrium between centralization and flexibility. How did you solve this problem?

Telecom consultant 1
Yeah, I think this goes back to the service catalog. I think to make it more explicit what the consequences were of certain choices. But I think what really helps, so if you say, 'Hey, you want to be central or decentral?' to really specify what you're then talking about, so to make it smaller and more explicit, I think that was the key choice.

So what to do with per service and who makes this choice with who and what team and who is impacted by it. So basically, this role, I think if you make it really explicit, that would make, yeah, come to life because everyone might agree on a certain statement. It should be decentral. OK, all right. But then to make it more specific would be really the challenge and... Yeah, and that's also what makes it then challenging and complex because then you get to the details and you get certain choices that are not viable but.

Koen Hendriks
How complex was solving this problem? If I say like this, it kind of sounds like the same complexity as question 6, like defining the responsibilities and accountabilities.

Telecom consultant 1
You can use the same scoring as in question 6 so 5/5

Koen Hendriks
And if we then look at the prioritization, so prioritizing use case set of products.

Telecom consultant 1
Prioritizing use case and data products, I would say the responsibility would lie within the data domain, so they can prioritize their own. So that was relatively easy because they would tackle the biggest problem. They usually

sell the products not only for other domains, but also as a way to improve their own data quality. So they would say this is the most important data to us. Let's fix this first and also make it available to the rest of the organization. So they would pick priority based on their own data product. Can be harder if there would be another domain that would say OK, I want to have this data from you and it would not bring value to the domain that should produce it. Then it can be a struggle.

Koen Hendriks
But you chose to go for a push kind of data product development then instead of a pull one?

Telecom consultant 1
So for for, for some. Yeah, for most domains, they had certain data sets that they were producing that they could manage better. And use this as a way indeed as sort of a push. Meanwhile, there are also projects that would use multiple data products from different domains, so then it. Would also get prioritize. For example, I wanted to predict future sales. I need R&D data. You need product data, you need financial data, so then they would combine the three of them. So then it would. Be on a project basis. And OK, so prioritization has two axes. One will be projects and it could be cross domain. So they would have. Multiple data products.

Koen Hendriks
So there wasn't like a Global prioritization of use cases and data products?

Telecom consultant 1
No, because I also believe that it shouldn't be like that because the domains produce and consume their own. Data products. So the. Prioritization of their data products should be theirs. They have all the capabilities within their domain. Platform capability development might be different, but that is a central theme. Was in our case. So there is no prioritization between data products and building the platform because it was a separate team maybe in the beginning. If you're not really mature, you have the same people that can also produce data products also built in the platform. That could be true. But then you would prioritize. That's from in the beginning. And once that is up, you would maybe start to to grow and. So ideally every data product is consumed by another data domain, so there should be a demand for it. But also we saw that there might be. So there might

be use cases but not a strong demand. And also domains wanted to organize their own data. So they created these data products. I agree ideally, you would only make something once there's a strong demand for it and then create the data product, but they also use it to set up their own data management data quality. And to to have a use case within their own domain. So theoretically that's not a good use case. Yeah, I can imagine that. You said like the use case within the domain to enrich and. Every should data and make it higher quality than within the domain. You can use it better.

Telecom consultant 1
And this is telling me because I weren't really mature, so some. The by the domain wanted to participate because they knew that some of their data Was very badly managed, bad data quality. So they saw this as a way to mature the whole domain, but also some of those data products were consumed outside of the domain.

Koen Hendriks
OK, OK. So how would you? Rate this problem and complexity. And do you perceive it as a problem?

Telecom consultant 1
So no, not really. Not really. So I can imagine that it can be a complex problem when you have a lot of demands from a certain domain. And then you would have prioritization. Which use case or rich data? Am I going to do for? If those related. Persons are only consumed outside of the domain, right? Because then you have nothing to really gain from it as a domain. Yeah, you still want to help? Maybe the other domain, so that could be an issue, but we didn't experience that. So in that way for. Us it would be. Yeah, wasn't that complex, so a 2 maybe. So there can be case it can be. More complex but.

Koen Hendriks
And the importance?

Telecom consultant 1
So we didn't recognize this really as a problem. Yeah, if it would be a problem can be important. But we didn't recognize this problem. So maybe just. Do not applicable on on these two, because they can imagine that another case would be very important.

Koen Hendriks
Hmm, interesting. I talked to Telecom consultant 1 and, maybe because she entered at a different moment, or because she's still busy with it in a later stage. But she said that prioritization is like really important problem.

Do you think that on the 1st first glance and with the start of a project, you'll have one domain. Then it's not a problem at all, but then after that, it starts to become a problem?

Telecom consultant 1
No, definitely.

Koen Hendriks
That's also just something to think about. But for now I will do this. As non applicable. Then two questions which

I don't. Which are normally implemented in a data mesh which is. Very mature. But still, maybe you implemented it? Did you already try to realize? The computational governance.

Telecom consultant 1
What is computational governance?

[...]

Koen Hendriks
OK. Any other design decisions that are? Not discussed yet that are important.

Telecom consultant 1

To have like the roles and responsibilities defining them and realizing them, then the part of realizing which is the capability flexibility with the interaction between those roles and between the teams. Then the flexibility, the centralization, so the coordination of the teams. Think one thing that we did discuss, but it isn't maybe explicitly mentioned. So you have roles, responsibilities, accountabilities but you also have the organizational structure. That's what we discussed there. So those are two things. And I think in designing. And operating model. Or structure is also a very important one.

Koen Hendriks
So that's like, do you think that you can define the organization structure with the interaction model, the roles, responsibilities and accountabilities together?

Telecom consultant 1
They can be a part of it, but if you have those three, you don't necessarily have an organizational model. So you can say. So this is an organizational model who reports to who and is under which department. This organizational model, then the interaction. Model could still be like. OK, you send me an e-mail if you. Need my help? Something like that. And we have a formal meeting every month or something, so that doesn't necessarily mean that you are reporting to someone just means how you interact. And then also with roles and responsibilities. It is. Then it defines sort of maybe yeah. Roles responsibility. Sort of like reporting kind of structure. But still then then you have the discussion if. Responsible and accountable. And those are two different parts of the organization. There still can be a problem and it acquires the same organization, so the work structure is different. It might not be a problem. So yeah, you can. Center of Excellence with underneath an infrastructure team. Doesn't need to be. You can also have a CIO with an infrastructure team and a center of excellence that. Don't report to each other. Only to the CIO.

Koen Hendriks
Thank you very much, for this addition. I will stop the recording now

TDIT2 Design and development interview Telecom 1 Telecom consultant 2

Koen Hendriks
What is kind of your start and finish when looking at this process?

Telecom consultant 2
So the way that we've. Done before is like I think you definitely need a strategy and a vision. So what is our North star? What are we trying to get to?What are our priorities? And then once you work out the priorities, it's a lot easier for you to work out how you then operationalize, like what's realistic, how many people do I need? What people do I need? To help operationalize this, so as part of like, you know, developing the strategy, it's kind of ask the questions of why do you want it. For example, why do you wanna move to a data mesh? What problems are you trying to solve? What are your key data challenges? Because based on those outputs, when you then go to operationalise, for example, when you need to start looking at policies and processes, I mean there. Are so many policies and processes. I don't think that realistically it's feasible to do it as like a one bang kind of approach. It's also not sustainable because what you want is you want people to get into the culture of seeing data as a product, well as damaging you see data as an asset, but here but it mesh you need to see data as a product and that's like a cultural and a minor change. So that takes time. So by being able to have a strategy and prioritizing and understanding or understanding what their key issues and challenges are, then it allows you to prioritize the tasks. So for example, from process policies perspective, they really wanna look at data quality. So in the first instance, let's focus on getting in data quality processes in place policies in place like life cycle management and stuff. We do that like maybe in the first Sprint, second Sprint. Let's look at the next set of priorities, third Sprint. Let's look at like the. 3rd set of priorities, so you kind of keep going and then that way you kind of build that. Culture and you kind of build that expertise.

Koen Hendriks
And you design which kind of processes you need?For, for example, data policy and then you implement those and then you continue to a new policy?

Telecom consultant 2
Yeah, well, you can kind of do it in streams so. Maybe you have like. How have we done in the past? But this wasn't for a mesh, it was more for like a normal governance one, it's like you have a stream that's more of like a business functional stream that. Gathers the requirements and then builds the processes and then they hand it over to implementation team. Because the for example, the implementation team might be implementing the processes.

Via a tool so. You kind of hand that to them. So when you hand to the implementation team, the functional team can move on to the next set of priority, prioritize processes, and then keep going. So you kind of had this like. Nice little

handover. Face as well and it kind of keeps it going. It also makes means that. Because I think sometimes the problem with governance is people think of governance and they're like, Oh my God, it's so much work when you could actually break that down into more easily bite sized chunks. And that way you kind.Of slowly on board people or. Get them to see the value of what you're doing as well.

Koen Hendriks
You don't first look at all the things you want to have of all? All the for example the data quality and the data lineage and things like that. First, look at that and then say OK, what do we have to do to realize those capabilities?

Telecom consultant 2
Yeah, you can do it as kind of parallel streams like for example, because if you have the strategy and the vision and you understand what the key priorities are so like you know, what are the problems that we want to address? Then you can start to prioritize the activities that need to go. So for example, first you have. So you have two streams like a function like I guess like. Requirements gathering show you have implementation stream. When the requirements are being gathered and the process of policies being developed, the implementation stream can be onboard. To learn how to use a tool or understand what their role is once those processes are developed, they can be handed by the implementation team and they can start implementing. When that implementation team starts to implement, then the requirements gathering can then look at the next set of priorities and start doing that. So you kind. Of have a continuous flow of. Is that what you mean?

Koen Hendriks
Yeah, I'm especially interested in the process of alligning the roles and the people.

Telecom consultant 2
So I guess like there's two ways that you can do the this. From from how we've done it in the past. When you're doing the processes, for example, there's. The people that need to help. Provide the input so that they can develop their processes, and then once the processes are in place, there needs to be ownership of the processes and usually that. Might be like new data. Or no data steward. However, else they want to want to do that. But when you're developing the processes, you might need input from like multiple other people like business, SMEs or people that will consume the data in the end so you know. If we put. This process in. Place because you're the end user. How do you see this work? Does this work for you? Is there something that we need? To add to the process, OK, so like you've got two groups of people when you're making a for a process, you've got the people that need to be involved to create the process, and then once the process is complete, then you need someone to be responsible for the execution of the process. Authorize this as the first people that build the process and also assign accountability and responsibility of the process. But what we usually do is usually there's a template of like a process. So what we would do is have a workshop. With the required. People that need to give us input and we say look, this is a skeleton. Process start to end.

For example a data quality issue. And then try and tailor it to say for example a Telecom 1 context or the client perspective or client context where it's like you know what systems are in place currently, what are. Like what happens currently like is it very manual, what teams are involved, so you kind of have all these bits and pieces and then you make this client specific. So you I think it's important that you have a skeleton in mind about what you think is. And a good understanding how they currently work, what elements they want. But as long as you have an idea of what good looks like for that process, it's a bit easier.

Think the the important thing is understanding the problem of the client and then coming with a solution, because sometimes you might implement a process and they're like, I mean this is good in theory, but it doesn't work for us because of reasons AB&C. However, when talking about data mesh and the processes. And data mesh. Those skeletons and those reference. Processes are sometimes not in place. I think governance in general, sometimes people have. A vague process, but it's never really documented or it's never really official. It's people trying to find their way around, so that's one problem.

And then the second problem with data mesh is i think sometimes it's quite new, so they kind of rely on us. To give them a bit. Of direction about what a good process should. Look like within a data mesh. So we wouldn't really expect them to have. Those things in place, So what? We would then instead. They show us they walk us through the key parts of it, keep key roles, key responsibilities, any system. Any other things that you need to know of? And then what we would do is probably go away and have a think about. What do we want and what do we need from a process for data mesh for that particular topic? Like does this fit? Can we try and reuse any of this? Because I think the other. Thing is. People don't really like change, so anything that you could reuse is better, and it's also easier for them to accept as well because they don't want to do. So it's kind of like seeing what they. Have looking at the concepts of data mesh and seeing how we can fit that in. And if we can't, if we have to redo their whole process and it's kind of like, OK. So how do we do it? And how do we get them to understand why we're doing it, yeah.

Koen Hendriks

You said in looking at the concepts of data mesh to kind of combine the process of the client and to what kind of concepts of data mesh do you look? Do you have any reference processes for data mesh?

Telecom consultant 2

I think it's more like for example. This will not find out, Because I think the processes that we have outside of a data mesh are still really handy, but then it's like how do you adapt them? How do roles and responsibilities change? Because now instead of having processes central to go into it like say, leading to a central team, it's actually managed within a domain. So it's kind of like looking at how do we then adapt processes central process that they might have to a data mesh. How do we make sure that the domains start using that process or the domains to understand their responsibility like they have a little bit more responsibility now for data quality because they're they're responsible for their own data. So they need to do a lot more of that root cause analysis. For example, if there's a data quality issue. So there's that kind of like taking a central process. And applying it to like a more Federated model. And then on top of that, it's also then taking a step back and being like, OK, in theory, each mesh is more efficient. Because you don't have the bottleneck from waiting on a central team to solve issues. But then at the same time, because you don't have that centralization, there's the risk that people don't talk to each other, that people work in silos. People might be duplicating work effort when they maybe they could. Work together. So it's like, OK, so then. How do we? Kind of introduce. A process so that they talk together. So it's kind of like taking bits and pieces that you know and applying it to how a mesh works and then making it practical and realistic as well.

Koen Hendriks

What I can imagine would be a difficult choice is, which sometimes you have to stand a little bit back from the theoretical data mesh.

Telecom consultant 2

Yeah. Yes. You have to say, well, in this organization it just doesn't work like this even though it's like nice theory. Some theoretical data mesh processes need to be changed. And also there's also responsibility aligned with that, but some processing processes are welcome. So you have to make some kind of choice. Which one you are going to change and which one will be directly "copied" from theory? How do you make choices? Like that or how do you assess that? So for example, the domain is getting set up so like OK, the domains are different process and then there also needs to be a way for them them to communicate across domain as well. So, you know, in theory how you haven't got out of governance council and you have that so. But we spoke to the client about it and they were like, we don't really want to introduce new meetings at the moment, because people are not gonna come. People are just sick of meetings. So like, OK, So what do you what? Do you have? Existing that we can use to adapt to that. And they do have like a specific board meeting that happens. And I think it's cross domain. So we're like, OK, let's maybe try and use that as a starting point for a Governance Council. To start managing the domains.

Koen Hendriks

So that could be like a loop in for them to escalate?

Telecom consultant 2

Yeah, issues like so I can. That that body can be like a part of the process for escalating issues above the domain. At the moment we're kind of like a bit in this weird spot because they don't quite have a domain set up yet, so it's kind of like we're trying to do 2 things at once. We're trying to set the domain we're trying to set up some sort of central organization so that there's that like connection in and also like. The way to for the domains to connect with each other. So like really what I'm finding is it's kind of understanding like the data management and data governance concepts, understanding data mesh concepts, understanding what the client wants and what the client is capable of doing at this point in time and then combining. Those three things together to help them. Come up with like. A reasonable solution and the other thing is.

You can't move to a mesh straight away. So you have to also understand in terms of clients capability like how much they're able to change over a period of time. So what we've done is we've introduced like intermediate targeting target operating model as they start to set up their domains and get their data management maturity up. So that you know we have in the beginning for example we have. More of a central data organization to help. Bring everything together and help drive data literacy. Help drive that communication that we're going to move to a data mesh and then overtime for them to then transition into a data mesh. This is kind of like breaking up the the transition, because it's a pretty big transition. And if they don't have the data maturity, it makes it even harder. To do that.

Koen Hendriks

And are they also specific processes is in place that you can immediately change and processes that need like a longer period of time?

Telecom consultant 2

Yeah, I've been talking about data quality low because that's like the main focus at the moment They have data quality processes. So it's kind of like understanding what that is. And then trying to fit that into the tool and like a new way of working based on the target operating model that. We proposed for them. So I think that one is going to be like I don't wanna say straightforward, but that one's a bit more like. Take what you have. We'll apply. It to mesh and data governance principles and imposing make it a bit more streamline process because at the moment it's like people have this many people have this. Many people have this meeting, then they come together and like it's. I think it's pretty manual. So it's kind of like using that manual process. We know that they're going to be given a tool and helping them refine that process via a tool automate parts. I think the thing that's going to be one process that we're going to need to. Introduce to them like or. Probably like write from scratch is. Requesting data products.Because they don't have a mesh and they don't have products yet. This is in process, you know, for how do we request data products as long as we get the marketplace up and running, I think that should be pretty

straightforward. I think the problem. Going to be when they request a data product that hasn't been made yet, then what's that process of getting that request recognised and then actioned. So we've kind of proposed like a high level process that needs to be probably that needs to be worked through in a bit more detail with the client and then signed offf as well.

The other problem. Is that because we're implementing a tool at the same time we're trying to like do this? Target operating model people sometimes want a quick fix and so I think the tool is going to fix the governance issues, but it's not not enough. Like it might mask it for a bit longer, but it you still need the governance in place, so it's kind of trying to articulate that. I think the main. Thing is to. Like keep a keep in the back of mind like key concepts, but I think what is? More important is understanding the client.But I think the non-negotiable is like you need to have some sort of data quality process in place.

What that looks like is going to be dependent on client requirements, what they what systems they have, what people. As a minimum, they should have you know, from a data governance perspective, they need to have owners and stewards. That's in the processes within each domain. So I think how how we kind of explained it to Telecom 1 is like. There's two types of teams in the domains. One is like the data governance, so making sure that the data is good quality and can be used by other people and so that the domain can be like proud of the data that they that they managing. Then the other team is the people that actually produce the data products, so they need to make sure that the data that they use to put it together is good. So that's their responsibility.

Koen Hendriks
Yes, that is like according to the governance, right.

Telecom consultant 2
Yeah, but they need to make sure that. The product that they create for other people to use, is good quality. So how we try to simplify it is that anything relating to data governance should be owned by maybe the data owner data. Or like owned by the data owner supported by the data Steward. Anything related to data products. So like for example the process of requesting data product. The process of prioritizing the data products that need to be made by that domain should be owned by the data producer. Team and then the data domain owner kind of overseas both.

Koen Hendriks
Do you think that there is still some kind of flexibility to move from that approach, right or?

Telecom consultant 2
I think, yeah, I think anything's possible because every organization is a little bit different. So if would have like blueprint then you can adapt the blueprint based on your client requirements, which I think is why I'm not sure like just from what I've seen today, I'm not sure that there. Are definitely like. This is the way that you have to do it because I think there are like checklists of things that they need to have in place, but the items within the checklist themselves. I think there can be a skeleton or a blueprint for each. One, but everyone is probably going to be different compared to another organization.

Koen Hendriks
Yeah, that's that's like the. That's the difficult part because it can so different between organizations

Telecom consultant 2

But there are key parts of the process that must be there. Like ownership, ownership, accountability, a decision making point like start and end, and I think there are. Key concepts that we can apply to that. But I think there's a human element to it that the process needs to be realistic to an organization. So that's why I'm kind of not sure that there's like a set blueprint for how the processes should be done, or that there's like one process that everybody has to.

Koen Hendriks
Do you think that there could be some kind decision tree that explains the theoretical part of a data mesh with some recommendations to make it more practical?

Telecom consultant 2
I think so, because I mean, a lot of it would depend on like their current arcitecture, their people, the way their teams are organized. But I think the processes that we're looking to implement at the moment are pretty standard ones around like. Really, it's really data quality management because I think everyone's focused when you first look at these things as improving data quality. So a lot of it's around data quality management like so. That's pretty straightforward I think. But what kind of processes are you thinking about when you talk about this?

Koen Hendriks
It's an interesting one because data mesh is meant to be more efficient, so I think it eliminates some processes as well because I think that mesh is meant to be less. Like traditional data governance, data management heavy, But for example, the data quality support, so that's like a capability that they want to have.

Telecom consultant 2
But the process I think in itself is the same. It's the accountability that's different, but I think the process itself. Is the same. Yeah. And then maybe the KPI's. For that is going to be different between the domain and the organization, because for a domain they could be like OK, we can't. Yeah, we can't have anything. We can't have completeness lower than say 70%. But from organization perspective, they're looking at whole thing like like it

could be 60%. So yeah that would. But i think it depends on my understanding is that it depends on the data, like what the data Quality is, so if it's data that's not shared with other domains and it only stays in the domain, that's fine. But if it's shared outside and other domains are gonna be able to use it, then that's a bit more of a global thing. But the process of profiling that doesn't change, you know what I mean? Like it's it's the it's the number that might be different. So within a domain they might have said. We want. I want to see a KPI. This, but globally they might be like look your KPI for this particular data set. Is quite high. We're OK with it being like 5% lower or something.

Koen Hendriks
Another example of not specific data quality, but for example the data format. So let's hypothetically, the global committee says the format should be JSON, that's that's our standard. But then when the domain says well within the domain, we would like to share some Excel files, yeah. So we neglect the format and if we want and we do it, Excel falls within our domain. If we do it to another domain, then we make sure it's JSON, but within the domain it's excel. So in that that example, you have people on top that define the data format. That's the process defining that that data format. But then also. Within the domain you have people deciding you can also use Excel. Those are both processes, like defining the data format. Where the one is globally and one is locally

Telecom consultant 2
I think it comes down to prioritization. So again, it's like the way that we thought. The data domain needs to should have a say about the data products that it produces. So you know what, like, what's the topic of the data product? What's the format of the data product? But at the same time? I don't think that a pure data mesh in. Reality really works. Because you have that risk of people working in silos and duplicating effort and all that kind of stuff and like no cross collaboration. So then I think that between the global and the domains, there needs to be like some sort of platform or some sort of mechanism for domains to come together and talk about their priorities. So that they can make a decision about what. Data products. They should make safe within a cycle like in this month. These are the data products that we're gonna prioritise making and as a part of that mechanism where they talk across domains, they should also take into consideration like what this global team wants. So it's kind of like a forum where I. Would think that. The data domain owner comes and represents their domains and say you know for this period of a month, we've received X amount of data requests for these top three products. So we think that we should prioritise producing these three products for the next month and then the global team maybe represented by like some sort of lead who is responsible for like data product prioritization like globally can be like, OK, cool, I hear you, but from a global perspective like or as an organizational perspective, we're getting particular request to provide this in this format. So what I would suggest to you is: I'm OK with your top three, but I would like you to add a priority for that, and this is due at this time, so then you come to a decision together about what the what prioritization should look like for that domain for that month?

Koen Hendriks
So OK, so if we go back to the example of the data format, JSON against Excel then. What could be done is that you say to the global well, we would like to use Excel within our domain and then the global says OK we define as standard between domains you contact each other with you you you use JSON but within the domain you are free to do anything.

Telecom consultant 2
So, data that doesn't affect another domain, you can do whatever you want. But for any data products that need to be shared and used by other domains, it needs to be in JSON format because that's the thing that. So the domains don't set the standards for their domain. They just communicate with the global to make sure that all their needs are met. Because I think the global team needs to have that Cross domain view because if other people are using their data products then? It has to be workable for everyone. So it's kind of finding that balance between what are the domain priorities and what's the what is, you know, the organizational priorities. And like prioritizing that together. So I think from a process perspective, it's not like necessarily the individual process inside the domains that change. I think the change is how the domains then work together with other domains and work with the global team. Like aligning the different teams is totally different because it's more Federated and there'st he connection with requesting the flexibility right, we're talking about the flexibility within domains. So, like, I think the data domain owner should advocate for their domain like you know, these are the priorities that we're getting from our data consumers because I think with the mesh, we need to remember that it's product focused, so the consumer gets through. It's like you know customer service. And so the focus is on what does the customer want. So I think the data domain owner needs to be like the most for their customers as well. But also there is that organizational perspective because the domains work together. For an organisations global team that comes together with what are the organizational priorities for us and then they kind of come together to find. To create that prioritization, it's almost kind of like Sprint planning. I think of it. So it's kind of like coming together to to discuss and align on priorities. It's it's more a question of collaboration.

Koen Hendriks
So actually the the processes that are important in the data mesh are the prioritization processes and also the communication between the domains and the central data?

Telecom consultant 2
I think so, like I might change my mind in like 3. Years time. But I don't think that. It can work without collaboration across teams. it's the teamwork between teams instead of teamwork within the central data team because I think definitely the data mesh is good for removing bottlenecks. So I think domains are empowered to make more decisions on their own, but I think as an organization, they will still need collaboration for efficiency.

Koen Hendriks
So how did these teams look like?

Telecom consultant 2
I feel like it's a giant collaboration. So this is like the Data Working group are the domain people, and then you have a data council which is like a global steering community. And the data working group there needs to be like we suggested, a Council or like a forum where where the representatives of each domain and the platform, can come together and discuss the priorities.

[The organization chart is created in collaboration with Telecom consultant 2]

Koen Hendriks
Yeah, well. It's pretty interesting I was trying to come up with some key design principles for the data governance structure. And make sure that a tool can be created that helps with the design of the data governance structures like these.

Telecom consultant 2
I think the most important thing is the voice of the customer, so it's almost like because we're moving to like a product mentality, it's like the customer's kind of always right sort of thing. So making sure that when you're designing this stuff, you listen to what the client wants, and then also reflecting that. In the process is because the data products need to reflect what the data consumer wants to see as well. And also, Thinking back like a lot of the. Feedback from the client and the reason why I did the process like this proposal

process like this was because they didn't want something complicated because I think if we overcomplicate it, they're like nice, it's too high a window it. So it's kind of keeping it simple? I would maybe suggest going more. Towards guiding principles for the roles and responsibilities as opposed to like defining a set roles. So like have a start of some role an responsibility ideas and then working with the client to then build that out.

Koen Hendriks
Oh, that's that's nice. That's good advice. How would you describe like the quality of such guidline that help with the design? How would you assess some model like this?

Telecom consultant 2
I think first you need to understanding your customer like understanding the client. Like you know why they're moving to mesh. What are the key challenges and the key data issues that they're trying to address? Maybe even also their data management maturity. And then based off that, you should have guiding principles that can be used to collaborativly with your client come to a new roles and reponsibilities. Like which which combination of roles is going to be the path of least resistance for them to adopt?

Koen Hendriks
So let's assume that we have 2 kinds of reference model that are both not shaped to the organization. How would you choose between one reference model or the other reference model.

Telecom consultant 2
Which one allows for better cross collaboration cross domains? And what is the time it takes to build the organization specific model is also important. I would probably look at both because I mean I think both of them have merits, but I would maybe lean towards. I would prefer a model that is developed from practice because the theory is good but there's a human element to implementing something people don't like change. So that's also why I would suggest choosing the model that's more simple and adaptable. Another important point is the impact on the design and implementation time at the client. Because people want to see results fast, but you can't see results until you implemented the whole structure.

Koen Hendriks
And how do you think that you can make the design and implementation time faster?

Telecom consultant 2
I think less theoretical, more practical and more simple. So which model needs the most people for the implementation. Because, the efficiency of the design process depends on the resource usage and the resource usage is time, cost money, but also the people experience. So needing less experienced people to achieve the same effect would be better. So starting with key concepts in a model, seeing how it works and then like, you know, applying key learnings and everything like that. And then also with the benefit of like doing it again, is that the people that you've done it with have done it before. So they end up being champions and they can do take more hands on work. So from us from like a consulting perspective we can. Step back a little bit more and instead of like having to do the work. We're instead more guiding them. I don't think it's like as important to have an exact model because I mean. People are unpredictable, so no matter how precise your model is, it's probably gonna have to change. So you should find out what these guiding principles are and apply them

Koen Hendriks
What kind of guiding principles you use and work for you?

Telecom consultant 2
Well, I did a combination of things I thought about. Like, DAMA DMBOK, data mesh bookh and customer feedback And they thought about mesh and then I thought about what the customer feedback was about. Like I

want something simple. I don't want something hard so it's like OK, combining all these things together, I created this one slide.

Koen Hendriks
And if you look at the operating model presented in the data mesh book by Dehghani, do you think that that model could work as a good reference model for organizations to design their data governance on?

They have like this operating model.

Telecom consultant 2
Well, the model itself wasn't helpful, I didn't understand like some of the arrows and directions and stuff I was. Like so, it's very confusing.. Also, like, the way of drawing is so vague that it is difficult to make something useful out of it.

[...]

TDIT1 Design and development interview Energy 1 Energy consultant 1
Koen Hendriks
So, just as a recap was for recording and I can remember, I just told you about that the operating models like a reference operating model is way is difficult to create, because organizations, organizations are so different.

Energy consultant 1
Yeah, totally agree it's better to come up with design decisions or design choices, or at least specific points during the implementation of data governance, where you have freedom of choice. And as you can choose for for the implementation. From my point of view, this also provides much more value to project teams than just having a reference model. And like trying to apply a reference model to a company, which then after two months, you identify that maybe it also doesn't fit, therefore provided a view, it's rather helpful, also to provide design principles, best practices, recommendations, and a list of topics that should be considered when you want to implement a data governance organization, focusing on data mesh, so that you don't forget anything, but the individual implementation, from my point of view is really, really specific from company to company. And there are different operating models that were better in a certain company, which also does not only include maybe the organizational setup of a company, but also the organizational culture. And this is, in particular, something you won't find on any slide. This is something we can only investigate by talking to the people and then finding the respective operating model that fits best.

Koen Hendriks
Yeah, so also both the organization structure but also the organizational culture is then really important. Do you mean then the only the data culture like the data literacy, or also other kinds of cultures in organizations

Energy consultant 1
Yeah, yeah, of course, when we talk about data, I believe data literacy is a foundation, or at least if you have good data literacy, it speeds up the implementation of such a governance framework, because people know what you're talking about. However, it also depends on the on the process or culture did say like that, like, Vedic people only work together in their certain departments, or better people work together in a processor setup. How communities are, for example, available in certain companies, like how do people exchange actually, and maybe also exchange across their area of responsibility? Like, how, how big are such communities? Are they really live in communities? Or are these just calls where people are just listening to and so on. And this does not only need to be about about data. But this can also be a business background to pure suspect, count, and whatever. Because in the end, what we want to do with data is to provide business improvement, return on investment, but when your business models, etc, performance efficiency increase, and therefore, it's not about the cost share, which is about data literacy, and a way of working with data, but also about communication, collaboration, and so on with any company in any regard.

Koen Hendriks
Yeah, I think that that's also something that I found a lot. And also I found the problem investigation interviews, like the interaction between people, how the processes flow between roles and between teams. And designing that is also difficult. And I think there's, there's a question in an interview about that. So I think that we will come to that later. But for now,let's just, let's just begin with the questions of the interview itself, to make it some kind of structure that I can use that for my analysis later. So can you I know you for a bit now, but can you briefly introduce yourself and your experience in data governance design projects? Yeah.

Energy consultant 1

Of course, yeah, I'm for since more than four years now with Deloitte Senior Consultant in the AI and data practice of Deloitte, Germany. And I have like around about one year of experience in the area of data governance. And in particular also focusing on data mesh, where I was part of a big project team,which implemented or introduced data mesh.

at Energy 1, which is a global companies spread across different regions and countries.And division was to implement data mesh as their operating model as theirway of working with data way of ensuring good data quality, etc. As part of this project, we have covered certain aspects, we have covered the infrastructure aspects, meaning building up a target, target architecture, introducing a data catalog where we have decided to go with colibra, we had a dedicated work stream for culture and change. So really introducing the new mindset manifests this new mindset in the in the business units, having communication materials, like comics, for example, explaining the different roles

that we want to introduce, and so on. We had a dedicated work stream for human resources, because not all the resources that you need, and all the skills that you need for that are already available. And there's also certain demand that you have to hire. For that you have to ask from your existing existing employees. That that's, that's a dedicated, dedicated work stream as well. And last but not least, the data excellent work stream, which was about defining the governance roles defining the data literacy, also, then rolling this out into the company having a training curriculum in place. And of course, after setting up and defining all these roles on the paper on all the slides, bringing them to life. So dedicated teams, we have ramped up for business areas, to really onboard the Data Domain managers, which we have identified, building up their communities building up their network, identifying the data products, they want to work on doing prioritization, doing data strategy, both of them and so on, so forth. So really bringing them the tools that they need in their daily business as a Data Domain manager as a data steward as a data custodian in the setup of data match. That's pretty much the background I have. I also have some experience in the area of the infrastructure aspects over my focus area was the data governance part.

Koen Hendriks
Nice, well, that's really a really extended experience. So that's, that's really nice. Like all the all the different sets of data mesh, and I'm already looking through the questions, and there are a lot of components that you talked about that we're going to discuss. So that's nice. Before like, starting with this whole project, you said that you have to define everything? What kind of models did you use to design the data governance structures?

Energy consultant 1
Yeah, yeah, with regards to the data governance structures,there hasn't really been a reference model. So what we've used is a collection of best practices for example, so the DAMA DMBOK, so data management book of knowledge was for example, one one resource that you like, whenever you deal with data and data management, data governance, we can come to DAMA DMBOK, which was one of the most valuable resources or references we have considered for example. Other than that, there is a certain like standard roles that you see in each and every data governance implementation, and not sure where they have their origin but it is something that is colibra or has been invented by colibra and Informatica but certain roles like having a data steward a data custodian and stuff like that. I'm leave them was already standardized. And this is of course, something also we have we have used and we applied of course, however, also keeping in mind that there might need to be adapted considering the required a data mesh has also considering the environmental setup in the company you're working. Therefore, of course, the definition on a high level is quite was quite clear to us, we knew that we need this kind of responsibility somehow the the specific implementation of definition of these roles has been developed and has been discussed with the business stakeholders in order to come up with first proper role definitions, which we could then also reach out to the to the company. I believe these are primarily the models we have used as a reference. I'm not aware of any other models we have considered for that.

Koen Hendriks
And so there were no data, you did not use any data mesh specific reference models for this?

Energy consultant 1
No, no, as per my knowledge. Back then, there wasn't really a reference model available for what am I actually this was really, really new. Like we I started or I joined a project team beginning of 2022. Back then the project was already running for one year. So you can imagine it's beginning of 2021, and data mesh more, or the papers were released, like 2018 2019. Sounds like that. Right? So this was really one of the first data mesh implementations and the work we had to do back then was was really pioneering work and going through the jungle and setting what's what's working, what's not working.

Koen Hendriks
Yeah, I can imagine well, hopefully, after a few years, there will come some reference models that can use digital Did you also use design principles during the design of both the roles, the interaction models, the structure of the data governance?

Energy consultant 1
not not really, actually the same answer as already before with regards to to reference models,the design principles, of course, best practices, such as certain governance roles you need, which we used. However, I would say that that's it. That's the that's the design principle of choice. The we hadn't had more more at hand,looking at the the technical and the infrastructure aspects of curls that we had more references like target architecture models, which we have implemented other clients, because usually like the target architecture is not dead specific in a data mesh setup, the only thing you have to ensure that you provide the capability for self service enablement of the of the business users, which is one of the core components of data manager, of course, however, the architecture aspects are pretty much the same as in any other data mesh or data management project from my point of view, because nevertheless, you need certain pipelining tools you need data storage, you need data consumption and access, you need maybe an API catalog and the data catalog in the environment, that events energy, they have also used an ontology editor to model like knowledge about data, not transmitted data, but actually really knowledge about data, hich was quite quite heavy.But that's pretty much the design principles, we have applied so far.

Koen Hendriks
It's interesting, not for my research itself, but that you also use, like the semantics and semantic modeling for that for that?

Energy consultant 1
of course, you can model certain semantics in a data catalog. However, that's primarily a description and stuff like that. And for actually all the deeper semantic stuff used, the ontology editor basically could like enrich your data

with all the contextual information that you need, for example, saying, Column B is only good if the project is already in state delivery phase or whatever. And then you new case, it must always be empty until the project isn't that certain phase and stuff like that. You could model I didn't know the probation that in Japan, the currency doesn't have any decimals and stuff like that, like all semantics, were part of the ontology back. Basically.

Koen Hendriks

That's, that's nice. I talked to a lot of people but none of the projects I've talked to yet that went through the step of semantic modeling and ontology. So that's like the first that it's it's nice that you went to that face. Okay, and if you look at the outcome of your data governance design, what are the components then that were the most important in your case and to design the operating model.

Energy consultant 1

Here was one of the key aspects of the of the data governance model. Of course, we need the definition of the roles and so on. However, in order to come up with certain roles definitions and stuff, first of all, you need to define certain certain processes, because then you can identify what kind of responsibilities do you actually need in order to complete a process. Therefore, we had like to keep your associate with processes which have been defined like one month as part of this project project, and one process is a value management process. Meaning, if you have certain use case of business case ideas, you address them to the value management process, then they will be prioritized considering the effort and also considering the business value, you'd investigate what kind of data do you need and stuff like that. And actually, all of the value management process, you can already identifies in the market requirements for the data that is most valuable to you, what are the use cases that are strategic from a strategic importance for you. So, that's the first process. The second process was the data product creation process, which actually is kind of connected to the value management process, because the requirements are coming from your value management. And this was a key process we were working on as part of the data governance team coming up with a data product creation process, meaning how do we get an order requirements and who should be involved in requirements gathering, when or how does the touch point happen to the development team for example, this is a centralized development team, they may be even the request to take over the development, because you can use the self self managed infrastructure and so on.We have considered aspects like security, data privacy, which needs to be involved in such a process, whenever we implement or design a data product, we have touched all these these aspects, which are important when you work with data. And by doing so we could ensure that all the relevant parties in Energy 1, are involved. And this is quite, quite tricky, actually, to involve them at the right point in time. Also consider really RACI, who should be involved, who's also accountable for certain aspects was just responsible and stuff like that. Yeah, and actually, at the end of this process, a data product is ready and can be consumed by the consumer again. Which means we have thought about stuff like can we maybe split these data products into certain sizes to have an estimation of how long it takes until the data product will be available, then we have considered certain dimensions and said, Okay, this small, medium large data product, for example, and then we could already get haven't provided an indication this takes at least 10 weeks, 12 weeks or longer, until it's developed and available for your use case and stuff like that. So, one very important point are the processes and all the processes you can identify the roles, the responsibilities accountabilities and so on that you need as part of the governance framework. Another very important dimension, in particular, when we talk about data mesh is the definition of data domains from my point of view, because later on everything, and the data measured may be organized by certain data domains, which is a collection of, of people a collection of data, which has a certain similarity, and where people collaborate, where people are accountable and responsible for

certain data, and provide actually this data as a data product as a verb describe data product in the marketplace,so that it can be consumed by any other party in the company, who should have access to it. And therefore, the definition of data majors data remains is quite crucial, because it's not that easy to identify the right data owner, usually from my point of view, and also from our point of view, the data owner should be the party who actually curates the data in the operational or transactional process. And if for example, your customer records are shipped because and there's a free text field and Salesforce available, and sometimes it contains the address, sometimes the city and sometimes the phone number, nobody can really use this, this information because nobody knows what's in there. Therefore, usually the data owner should be the person who creates the operational detail record because this is also the party who can decide who should have access to the data, how to maintain a good data quality to enrich it with all the contextual information. Now taking this example of the customer, for example, the sales team creates customers and sales force.However, usually also the financial department is quite interested in in customer records, because you do certain controlling on customers, which is then part of the controlling department. And then controlling is usually reporting to the C-level because they want to have certain questions answered, which is why patrolling departments usually also say I'm responsible for a customer and I'm owning it. However, the controlling department never creates a customer record. At least not in Salesforce. But controlling department could for example, maintain certain information in ERP, SAP ERP systems or generally ERP system. And now you can already see that you have two parties and both say and this is also what happened and Energy that both say we are responsible for customer. So usually you cannot resolve that. But what a good data domain should do is a data domain should bring both parties together and make them collaborate. Meaning actually, a Data Domain definition which purely focuses on a department structure, and an org structure on a process structure actually doesn't really work from our point of view. And that's one of the key learnings actually, for us. And we should rather consider the context or the data object itself in order to identify and shape good data domains. Because in best case, within your data domain, you build up a new community of people who have never spoken before to each other, and make them collaborate to provide value added data products, because the most value you can provide to the company is if you're a customer it gets from the Salesforce and from the ERP are comparable and have the same keys and have half an hour with applicants and stuff like that. That's the most visual is what triggers you turn by the Data Domain definition

Koen Hendriks
Would you than also combine the controlling ERP and the Salesforce people into one data domain, Data Domain customer, or would you still approach it as two data domains which are tightly connected?How would you describe it?

Energy consultant 1
Oh, both options are possible for from my point of view, it really depends how your organization is set up. At Energy 1, we had both and dedicated data domains, we had one data domain for CRM, one data domain for finance, what we did, then else, of course, building up and inform a human community which should bring both parties together. And for this, we were even organizing a global workshop, which should actually be conducted or should should actually take place in unison. So everyone should be able to use them for a few days, just talking about the commonly shared data objects, and to come up with good solutions there. Because actually, the data ownership was with the CRM department, which was already defined as part of our project. However, we also could not change the data domain definitions anymore, right? Such such issues to identify after after six months and sounds like that, this is nothing people tell you at first they have to project you experienced that over time.So what we had to do is we still have to do data domains and build up an informed community. So people collaborate, wherever, if you experienced this in many projects, I can imagine that the next project you which approach approach differently in would rather try to write very bring the eight people and the finance people together. And then let's see whether they they want to become one data domain or whether they want to work separately. Like you have to, you have to talk to the people I would say, in order to get the best results for both options are possible for you.

Koen Hendriks
Is there a reason why one would work and the other would not?

Energy consultant 1
This brings us back to one of your first questions about the culture. From my point of view. It's a culture saying that our departments like to work together with other departments or not, whether there are any political restrictions within the company or not. Like this is really individual, usually, from my point of view, cultural and political reasons, I would say. What you see quite often is that people, the ground would really love to do that. But they don't have the backup of their management. Like we have seen this like like every day actually like talking to the people at the bottom. They really, really wanted to drag this this topic forward. But then there were some particular political reasons in the upper management, which were blocking the people, actually.So if you haven't had an open mind that management, especially the middle management, from my opinion, if you this could work quite well, if your middle management is blocking this won't work that well, from my point of view.

Koen Hendriks
Okay, that's interesting as well.

Koen Hendriks
Okay, are there any other important key aspects that you would like to touch upon?

Energy consultant 1
No, I believe to you talking about key aspects, its processes roles, and Data Domain definition from my point of view.

Koen Hendriks
Then we, I'm looking back at the questions,then we are going to talk about some problems. Because there are a lot of people running into problems while implementing this, or migrating to the data mesh, and there are like, some problems occur more than others. And I'm trying to find like, best practices to solve these. And also, I want to know, how important these problems are. So I have a few problems. Let me first start with the first one. Did you experience problems with defining the roles, responsibilities and accountabilities? So the RACI matrix of the data governance operating model?

Energy consultant 1
Well, that depends. What is more difficult is to really bring them to life because as soon as people recognize what an accountability actually means that we'll get back to you and have some some concerns or questions and stuff. So the definition of it is usually quite easy. But as soon as you really start start working in that setup, and someone is accountable and responsibility for the responsible for it. You might, you might see deformity other blocking the initiative. However, at least like with people and the way we approach that as well, like we had many discussions upfront of the people that will be involved in this setup later on and so on.We didn't really experience that.

Koen Hendriks
Okay, that's good to hear, but then the bring it to life, do you mean then when people are realizing that they are accountable, so they need to do more, or bringing to life as extending their capabilities to be able to do the responsibilities that they are accountable that they are assigned to?

Energy consultant 1
Um, let's let's provide you one example, we have really much focused on the Data Domain managers at the very beginning to identify the Data Domain managers for the data domains, because the time we can start working in their respective domain and to ramp this up. Therefore, we had like really intensive discussions upfront with the domain managers and everything was clear, ever clear about their responsibility and responsibility and also about accountability. To to all the discussions, what happened afterwards is when we when we onboard to data domain and we try to identify additional stakeholders and the people that might be involved in or should be involved

in such initiative. You had like meetings with plenty of people, and all wanted to do something with data and contributed but no one really wants to take over the role as a data steward and become accountable and responsible for additional topics and stuff like that. And this is what I meant with like bringing to life. At very first meeting, we were showcasing, Data Domain manager we already have, what we now need is for the sub structures, maybe even sub data domains focusing on different data objects and so on. We also need to take us towards and so on. Who of you feels responsible for that? Nope, no one feels responsible or accountable for that. But everyone has an opinion and wants his opinion to be considered and in a data data productive condition.

So I would classify that more as the motivation to take accountability for the roles and responsibilities of roles actually.

Koen Hendriks
And how did you solve this problem?

Energy consultant 1
We did a couple of workshops, like really a workshop series, where we started identifying and brainstorming with all the people that wanted to be involved about certain data products, which they think they are valuable, and they think they belong to the Data Domain. Then we add a set of data products. Afterwards, we have done a data product prioritization in order to identify which ones are the most valuable for us. But also, what's the effort or the complexity to build up these data products? And then actually, we

said, Okay, for the, for the first iterations, we just focused on quick wins. So easy to implement data products, which provide a lot of business value. And now we had concrete data products, where everyone who was in that meeting said, that's important, and that's low effort. And we need a data product manager audit and data steward for it. And then we asked the question, okay, you all all together, you all said, This is important for you, and you want to drive this forward. So who is now the person doing this, who's our key contact person, and then someone had to raise this is his or her hand. And this was actually working quite quite well, because, like, people recognized that considering the entire audience wasn't admitting that it's actually a good thing to take over the responsibility for something. And to show four weeks after that you already have to first results and stuff like that. Yeah.

Koen Hendriks
And if you look at the problems, so the problem of finding people accountable for their roles or responsibilities, because there was kind of the workshop for the accountability, assess, assessing accountability for someone, you said that this was kind of a problem. How would you rate this problem from one to five? The importance of this problem?

Energy consultant 1
Okay, the importance of this problem is the a 4. For because from my point of view, if you don't solve this, you have like a bigger problems after after running this project for several months, because, like if you do something, instead, it's not implemented in the company. So the importance, I would write quite, quite high. The way to solve this problem, I would write quite low with maybe a two because as you've seen, you can solve that and provide something to the people they aim for. And then of course, we had, we had the data product managers already defined because there was one person dedicated to the data product was supporting us in the definitions and stuff like that. So the data product manager was there. And then of course, the idea is, if you do this over months, there is one person which is maybe responsible for five different data products, or 10, and stuff like that, then we actually this was our idea, we could say, maybe we can group this. And so it's sub data domain, you become data's word for it, and can delegate delegate the data product manager role to someone you trust. Because as soon as a data product is implemented, actually, the maintenance is quite easy. Of course, you have to, you have to say submit, you have to missing click on a button if someone wants to have access and approval, excellence approvals and control data quality may be extended by certain fields, if required and stuff like that the maintenance is quite easy. And this was then our our approach, like first of all, now having the data domain managers identified, if a Data Domain manager becomes responsible for many data products, maybe make him a data steward and delegate a data product manager role to someone else.

Koen Hendriks
Okay, nice. So I say the importance you skill at the 4, and the complexity of solving the problem, you would skill as a 2.

Energy consultant 1
Yeah, exactly. out of out of my experience, I would say so. Nice. Okay, I'm looking at the time, we have 17 minutes left, we have a lot of questions. So we kind of need to, even though we could talk. I would like to speed up a bit. So let's go to this to the second problem.

The availability of the capabilities. If you look at the data governance capabilities that are needed to alter the data literacy, did you experience problems there during the design or implementation? Well, I mean, that there are people from the business domains, you should get data specific capabilities. So, whether they were those capabilities were available or not. And if they were not, how did you solve that problem? Why did you solve it that way?

Energy consultant 1
So what kind of capabilities were in the business? Very, very different from department to department, I would say, usually an engineering departments and reduction departments and so on. You have quite a lot of data people on when it comes to data. So because you have quite a lot of businesses and stuff, who work with data in their day to day business. We even have already the first data models for for the area like just in their pocket because they thought it's

important for them, and so on so forth. Whereas in other departments, for example, in sales departments, where people were either like focusing on the company connection to the customer stakeholder management, customer management and so on relationship management is not that well defined. So it's quite different from from from department to department, first of all.

And other than that, of course, besides the data literacy and so on. We also introduced new tools, of course, like a data catalog and so on, which should also then provide a marketplace capabilities and so on. And where we had to skill people differently. So meaning a data user uses the data catalog, maybe in a way that he searches for data he wants to consume is interested in what searches for data definitions and stuff. So these kinds of users just needed a general introduction to the data catalog they need to use, and the user end And primarily, that's it.

However, if you're really working in the data domain, if you're the Data Domain manager, the data steward and so on, you have certain tasks you have to work on in the data catalog, like implementing or modeling the logical data model, providing data definitions, and so on, so forth. And of course, this is something people had to be trained in

because this was was completely new, I mean, new to them. However, it's rather a tool training. What we recognized quite often is that like basic skills are also needed. Like for example, what are rotations then for logical data models, entity relationship model, different rotations and stuff like that? If people had experience with that, they had experience with a certain standard, but maybe this wasn't the standard we decided to go with as a central solution as part of for best practices. For for. However, many, many people haven't had experience at all in this area. Of course, we have to train them. Yeah, so so very different. But data basics, data modeling basics are definitely required to specific are definitely required. And then it depends from department to department, or I would say,

Koen Hendriks
would you organized some kind of training for them? He chose to do that instead of hiring new people from outside the company to fill the gaps and capabilities.

Energy consultant 1
Both. First of all, of course, what we did, we identified new roles, which didn't exist before, when it comes to working with data, so not just the governance won't like Data Domain manager and stuff like that, but really like role descriptions for the HR department. In addition to that, of course, all the career paths have been defined, which didn't exist before. So people also see what's my career path in if I take over this responsibility for something. And then of course, the we had the training codec or a curriculum defined meaning what kind of trainings do you need? What kind of skills do you need in order to fulfill a certain role. And in addition to that, for roles where we definitely knew okay, this is something completely new or they they're really demanding an additional employees. Of course, we also provided them with some job postings. And they hired people. In particular, for example, they hired a person for the data governance lead was leading this entire initiative, and to governance board and stuff like that. However, also, people on the operational level, were hired in the area of data science, for example, to force the use cases and stuff like that. So both approaches, new roads, career paths, then doing a fit gap analysis of curves, where do we have the biggest demand where do we need to hire and they can be trained on people.

Koen Hendriks
And then the question again to scored is ow would you define the complexity and the importance of this problem?

Energy consultant 1
Yeah, complexity probably a 3, importance a 4

Koen Hendriks
nice. Okay, then let's continue to something that is particularly interesting, the interaction model. So how the, for example, the domain teams interact with the governance leads and also the other roles that you defined. You sent the PowerPoint slide with the data governance and their you created like a pyramid? I don't know whether that's also just Energy 1.

Energy consultant 1
Yeah, it's also Energy 1.

Koen Hendriks
Okay. How would you? How did you design the interaction model? And did you run into problems here? And how did you solve that? Kind of take me through this design process?

Energy consultant 1
Yeah, yeah, of course, the interaction model was actually the model you can see on the slide is pretty straightforward, I would say, because likely, we had the chief data officer who was our executive sponsor, we were building up a CDO office. So with all the capabilities and Target Operating Model they needed for that CDO offers, and actually, the CDO office was aiming for introduced data mash. And that's the general setup. As already mentioned before, the CDO wasn't the data governance lead in that example, because like, he was really the business sponsor on C- level. And the data governance team was like one level below part of the CDO team, of course, or part of the CDO offers, and the data governance lead is responsible for managing all the data governance initiatives.

So managing and approving maybe also what kind of role definitions do we have here to sign off than ever, we have defined a process here to sign it off. And of course, he had to do all the stakeholder management to bring them to

life to enable the team's bringing, bringing these worlds and processes and stuff to life. Other than that, one of the first

topics we were working on this the definition of data domains, we've tried different approaches, we have discussed different approaches and the company. And then finally, we have considered actually the business.

The business capability model as our reference for the data, data governance models, meaning in a business capability model, they have defined capabilities like CRM, finance, production, and so on so forth, which we use as a reference and also define or use this existing structure. Because you can imagine in such a big company, it's really, really hard if we want to come up with something completely new. So we have used that as a reference and brought us to life.

And then of course, we started with a lot of discussions identifying top down, first of all the data domain manager. And then of course, in the next steps important to break this down into data steward identification data custodian identification on the operational level, however, the technical level had to come first. And there were quite a lot of interviews going on in the company, identifying the right data domain manager and so on. And finally, then , we came up with one slide, adding all the data domains or the Data Domain managers, and the data governance council has been introduced, bringing all the data domain main data, domain managers and data governance lead are on table to discuss about strategic topics to this cost about issues to have also one instance of escalation in place, if for example, many, many data domains be responsible for the same data, like all these things were discussed in the data governance concept. However, domains itself are running independently, and it was rather like a top downapproach. And if you if you take a look at the report, which showed you the operating model, Target Operating Model model of the CDO office, they had one unit for data governance, they had one unit for the infrastructure and platforms. They had a third unit, which I can't remember the name right now. And they had a business unit for cross delivery teams, which should then support business, in the implementation of data products in the implementation of use cases and so on. So they also provided centrally as a center of excellence, implementation teams, which can be flexibly scaled up and down in order to, like build up to business use cases with the data business is managing the data match.

And, of course, on the long run the ideas, of course, to provide more and more of these implemented implementation capabilities from the center of excellence in the CDO office to the business units. However, at the beginning, you definitely need that to for the development.

Koen Hendriks
Yes. And those are then also the people that sit together with the Data Domain managers about the about the inner governance council and nice. Was it difficult to come up with this structure?

Energy consultant 1
Like the structure you can see on the slide, probably not, because that's quite intuitive and top down approach. What what's what's definitely difficult is again, like, how do we shape the data domains, and then identifying the right people that take over the role as the Data Domain manager, because these are the first people you need to identify to establish a data governance cause? Because without these people, the data governance councils just with the data governance need, who was organizing it?

Koen Hendriks
And is it then also true that the data, product managers, they escalate towards the Data Domain manager, and the Data Domain manager then escalate towards the council?

Energy consultant 1
That's exactly had also slides explaining this, this is usually the path. What you can also you cannot see it here. But what we also try to establish is not just having the data domain, but also having like sub sub data domains. For example, let's say we have CRM as the data domain, and then sub data domains for customer for sales order for different data objects, basically, or, or business objects and all the data around these business objects. Meaning in the first instance, actually, the data product manager would escalate to the data or sub Data Domain manager, and then it will date the domain manager, and then we'll add it up. However, these sub data domains will just like introduced to provide a bit more structure to toward his people.

Koen Hendriks
How would you rate the importance of the problem of designing the interaction model? I would say,

Energy consultant 1
With 5, like, these interactions are essential. And you have to provide them with some structure, when you when you talk to such data domain manager or you want to talk to people about the data product and so on. It always helps to have something at hand, which explains why are we doing that? Where are we already in the process, what comes next? Who want to talk to next and so on support? From my point of view, therefore, this entire interaction model is like really, really crucial to have standardized processes and so on.

Koen Hendriks
Let's let's continue. How did you encounter the prioritization of the computational governance and dynamic topology?

Energy consultant 1
So from the governance aspects, and you're really focused on the data governance, computational governance, of course. The roles like IT Landscape owner and probably responsible, it's defined like, besides the data governance

need, I already said there was another subunit, focusing on the platform and infrastructure. And they were then primarily responsible for the computational governance. However, the you would not find any RACI matrix or stuff like that these were rather the technical guys. So computational governance, probably important, however, has not been a focus area for that project.

Koen Hendriks
Should I ask that to Energy consultant 2 Tomorrow?

Energy consultant 1
You know, he will be a good counterpart for that area.

Koen Hendriks
Yeah, that's, that's nice. And I am, will ask him the questions about like, the more technical computational parts.

Energy consultant 1
But prioritization on the content, lets get in detail, yeah, tell me tell you something about so prioritization of use cases, data products and platform capabilities. Maybe let's, let's start with the with the platform capabilities. Because this is, of course, crucial. At the very beginning, you need to define a target architecture. And for this, you need to identify and define first of all, what kind of capabilities do we actually need, then you define out of the capability overview, you define the target architecture, we'll proceed certain components, and then you say, for example, we need a data catalog, you need an ETL tool, and so on. However, in the next step, you check which of these capabilities that you require, do you already have available in your in your company, and then actually, you like you can color your target architecture or whatever, and identify those capabilities, platform capabilities that are still missing. And then use can start usually, or usually started in a selection process for those that you consider as the most important ones for you. And like, you can imagine what what is more important, what's more important, an ETL tool to do the pipelining and to data storage or visualization tool. So like usually, you cannot use a visualization tool if you don't have the data available. So prioritization is quite intuitive, I would say in this regard. When we talk about data catalog, data catalog is simply important, because it's definitely required. If you want to do data analysis, you need to have a central catalog, where you put all your data and sort of requirements can come from different areas, from the governance area, from the technologic, logic area, and so on, so forth. And you just have to prioritize them against each other, which you can usually do that as part of like a workshop or something like that. I would not say that that's too complex. Prioritization of use cases, like very important. We have a dedicated works team. That's the last verse team I was missing earlier, we had a dedicated project work stream, which was called Value Management, and Value Management was defining the Value Management Process future first of all, which which I mentioned already, and they were leading the value management process, meaning they also did update business partner board, also on the sea level or business partners, where they discussed and forever prioritized use cases. They were responsibility to estimate the business value and the effort to build up these these use cases. And they were basically like hot, shortlisting, or use cases which have been identified during workshops and interviews and stuff, and came up with a shortlist of use cases we have used for POC implementations, which were at Energy 1, like more than 100 of use cases out of which six have been prioritized for the MVP phase. As an example, and this is a continuous process, and really, really important. What kind of problems to be experienced there. Of course, it's not always easy to to estimate the complexity or the effort will have such a use case because sometimes a single case doesn't that data is actually quite easy to get, but there are then maybe in the next step certain GDPR requirements that you forgot. Or you think another use case is not as valuable however, it contributes to us. strategic initiative and you just have to do it because of political political reasons and stuff like that so so do you usually try to make the prioritization as transparent and have made the prioritization as transparent and objective as possible. There's tools, political reasons and so on to be considered as well, because you want to satisfy all your sponsors. And usually, then you have to provide one use case at least postpones ourselves like that. Then prioritization of data products we discussed earlier, right within our data domains, we have discussed the data products they need in order to bring our governance model to life. However, actually, actually, we didn't really care about the data products Data Domain needs. Actually, what was more important for us are the data products we need for the strategic use cases. Like like this was our hidden agenda, which you had to keep in mind, like this other exercise was just like to get to be people on board on into to satisfy them the same results. However, actually, the priority topic was to build up the data

products for the strategic use cases. However, in turn, the selection of the strategic use cases wasn't that bad. So that actually the same use cases or the same data products have been identified as important as the ones which contribute to the prioritized use cases. So HB, eight, this was not at all the same, at least with certain master data and so on and about transactional data, whether you do this the sales orders first or your returns, or whatever, actually, people really don't really care about. That's, that's, that's the thing. Yes. So actually, the prioritization of data products went quite well. However, on the strategic level, of course, the prioritization was quite easy, because we had to work on the data products first, that contribute to the use cases and not on the other bots. If in the data domain, you identify a very good use case and argue that you need to stay there, of course, you can, you can take this use case. Pitch it in the in the business partner board, so that it's addressed in follow up in the innovative management process. And then again, cycles close.

Koen Hendriks
Okay, wait, I'll come back to that. Okay, and so the, if we are going to, it's nice that there's like, there are different levels of prioritization. And also understand that the use case prioritization is kind of the same prioritization as the data products, because data products are made for use cases

Energy consultant 1

Yeah, yeah, exactly. This was the idea. Like we said, we do not implement a data product for which we don't have a use case, because what happens, the data product is available and registered in the data catalog. And like, that's it stop. Yeah. So, even if a data product is really easy to to get in the data product catalog, and there is like a use case for it. But the use case is not prioritized. So for example, the use case is at priority level 100. So very low priority, but it is fairly easy to make the data products available. Is it even then? Well, we're not going to do it because we still have 100 use cases first, or where they're like, what you could do is, for example, to still provide this data product as a as a pilot, or as something you want to use to demonstrate how it's working and that it's working. Of course, you can you can do that also the program lead on the client side was saying okay, now, we have defined a process a data product creation process for months, and he only trusts us that this process is working if we ran it through as defined more one time and for this exercise, of course, we said let's take just an easy data product, we do not need to do this with the most complex one, but we ran this through one time and then you could for example also take such a such a such a data product, however, that usually a data product should have use case and the use case should also have a certain priority, because otherwise we can spend this this this effort to to other data products.

Koen Hendriks

And then you said like pitching use cases. That's the you can pitch the use cases to the use case stream that defines the business value. Especially the use case, then do you also have use case owners?

Energy consultant 1

Yeah, indeed, you had like the use use case Omona in this case, so the person that wouldn't derive this use case that maybe also controls the budget, and thus the stakeholder management tries to identify a sponsor and stuff like that. But yes, the idea and also on the site, we mentioned them as use case owners and as data product owners. Yes. Which have to work together. Yes.

Koen Hendriks

Okay. Nice, then how would you rate the complexity? I think that it's easier to take ability, capability development and use case less data product prioritization, like, separate.

Energy consultant 1

So no, we can we can keep this together, I believe in general, the prioritization is not a big deal. You should make try to make it as transparent as possible, involve all the people that you need to there so that you do not prioritize something without an important stakeholder, for example. That's, that's of course important. Um, however, the complexity is, from my point of view is a 2 to because part of the prioritization is our daily business in any project, you will see that. Yeah, and of course, the importance is added say it's a 3/4. Of course, that's, that's important to have emson prioritization, because otherwise, you do not provide a business value where it's really needed. And especially in long running projects, usually need to show that you have a business where you enter return on investment, to get additional budget process as consultants, but also for the CEO, or wants to drive this initiative. Good. Nice.

Koen Hendriks

Okay, we talked already about the computational governance that is currently not applicable for this and then the dynamic topology, you already said that you use semantic modeling. Can you explain more about whether you initially run into problems with that and how you solved it? And why did you do it that way?

Energy consultant 1

Yeah, yeah, like, of course, the ontology editor was already available at energy 1, it's nothing we have introduced, they already use that for couple of years. However, I'm not sure if they were already pretty life. Not not sure about especially in the engineering department with all the visitors and building up digital twins of their of their engines and turbines that they build up. Of course, they had already quite quite a lot of information modeled about their their assets. Their goal with regards to the ontology was also different one because actually, they wanted to develop a billing information model, which they could use, or can use in order to connect with all the third parties. They're working with, like with suppliers, with customers with support partners, and so on. And you can imagine if you have some data, maybe the column is named ABC. And as your customer, the same column is named XYZ, that and suddenly you need to maintain this mapping and stuff like that. And this dynamic topology, right? You have multiple, multiple descriptions for the same object, so to say in different in different entities or domains. And this is what I wanted to model with the with the ontology editor, and also as part of the ontology editor. There, they've chosen a vendor, which provides low code capabilities, meaning as soon as you have defined all your like ontological knowledge, you can automatically create API's out of it. That's the upper layer API automatically generated from the ontology. Looking at the bottom layer of curious data. They've distinguished the ontology into different components that the domain ontology, which is about the domain knowledge and the entities that existence owns first, however, bonded below and connected to the domain ontology. They had the tool ontology, meaning they have modeled in particular how this onto Balaji is represented in SAP in Salesforce in whatever system. And by doing so they could actually then connect it to ontologies. And could could, you could easily see maybe the customer field for the domain ontology, how it's represented in the different systems and my dad and stuff like that. Meaning actually, it was not just an ontology with the option to store the data, for example, in snowflake, and snowflake was connected to the ontology editor, you actually had a knowledge graph, which is even more valuable for them. Yeah, that's, that's the goal setting like having actually the data, having all the semantics around the data and the ontology and being able to automatically generate an API out of it with you, which you can register in the API catalog. And that's a data product. In addition to that, we wanted to introduce a data catalog. And

now you can imagine, we say, okay, data catalog, meta data is loaded. And they say, Okay, we also have data data. However, actually, their meta data is actually knowledge. That's semantics data data. In a data catalog, it's quite simple.

It's a short description, and maybe a logical and physical data model and stuff like that. However, in particular, when we talk about the logical and physical data model, the logical data model is more or less the domain ontology, the physical data model is more or less the tool ontology, which represents the representation and the tools. Therefore, we have proposed, how can both systems actually interact? Because they should not be standing on? And how can they interact? But also, what's the separation of concerns, which tool is leading forward, where we said, okay, data catalog is actually the leading tool for metadata, and the marketplace information and so on. Whereas the ontology is the leading information about semantics and knowledge. And this is actually the separation post. And having an inter interface between both systems where they can exchange on the data models. So what columns are available, what's the feed legs and stuff like that they should interchange and keep keep each other in sync. So it's actually be directional. This approach has also been implemented as a POC by us and was working well, it also was also working out. And then actually, the next step, as part of this project would have been like to stabilize the solution that we rolled out at the end. I can also sent you the link to that paper to that as well, which we have proposed to them, because that's the smallest one available. Okay, if I can send you the link later. Our management summary about what's the separation of concerns and how can bolster towards intact. And of course, as you can see, we have actually really touched the topic of dynamic topology. And when we talk about droplets, of course, the problem is there is already an existing system, and we want to have the data catalog and people from the ontology choosing to take over responsibility, or we do six files and stuff like that. Of course, they they see a, they are afraid of this new system, of course, and we had to do quite a lot of workshops, again, understand what they are doing present what we're doing, and provide a best fitting solution actually today version really leverage the core capabilities of postal systems. Of course, that's the that's the problem there. And also, of course, if we would have only implemented a data catalog, yes, all the colibra and Informatica guys would now say dynamic topology is also possible, however, not possible to that extent as it is possible in the Indian polity editor, that that's definitely the case.

Koen Hendriks
Yes, so the dynamic typology between for example, if we take it back to the domains, because you're kind of flexible in your domains, there is a possibility that there is a dynamic typology between domains which made them sometimes not fully interoperable.

Energy consultant 1
Customers. The best example again, you have in sales force with the sales team, completely different definition of customer compared to ERP systems and finance teams. So it's in different domains different in different systems, it's different. And of course, there are many fields that do one or the other departments simply doesn't care about. That's that's

definitely the case. What comes in additional like the CRM team calls customer the finance team calls that account because of SAP Quite a count and a customer or business partner, you already have three names for the same entity.

Koen Hendriks
Yeah. And then you solve this problem by implementing, if I just sum it up right now, you solve this problem by actually using your ontology editor, which was already kind of implemented into clients, because of the sector in which the client works. And that enables people already to make the kinds of topology, this kind of ontology, because there it's important. And that's why this was this problem was kind of less complex to solve, because there were already ontology capabilities there.

Energy consultant 1
Yes, yes. Like if the, if the ontology editor would have been more mature with regards to domain governance and like, again, if you want to really live and ontology you need domain governance, you need also like governance boards about that. Then, if you, for example, also discuss topics, like in an ontology on the fifth of every philosophical layer, what exists, like it's really a philosophical discussion, actually, what exists and what should be modeled there, and how should it be model modeled, also, considering ethical aspects, and so on, so forth. However, all this wasn't really existing 90 ontology was something termed by technically guys, it guys mature, like it quite nerdy, you can imagine that this if we sing around that, that you actually need, in order to bring that to life was kind of missing. However, with our governance framework, we've incorporated that. And that's a huge benefits of benefit for us. And

Koen Hendriks
yeah, nice. Okay. And if we are going to scale it again, how would you rate the complexity, let's say, let's say the complexity in case you have like the ontology that you did, so the complexity in your position, and also the complexity, if you would not have used would not have that ontology. I'm like, kind of interested in both because you implemented it. So you know how complex it could be,

Energy consultant 1
Actually in both the cases, I would say it's a five, because it makes essays really, really difficult. Just talking about the customer, using the same language is really, really hard. And people from different areas have a different view also on same objects and have different names for the same object. This makes it really, really hard in the communication with the people in the workshops, and so on so forth. Therefore, complexity definitely five in order to solve these things out, and to create an understanding, perform personal portfolio department about the other department and vice versa and the importance, also a five. This is where data literacy starts, right? Of course,

they're easy objects, but central master data objects, usually people have different perception and views on material, that's the same product, that's the same then of course, you again have to question in some ERP, the products are modeled as finished materials. And then you already discussed this a group material a product is a product or material or are not like it at all these different perceptions, because like people are used to how its model in the transactional system of choice they're using using a data source. And then like, like people from the outside saying, Okay, why why should the product be material and stuff like that, but people use that and have dynamic views on it, different views on it, therefore, definitely complex and definitely important.

As well, those conversations and discussions I think, will be very interesting, because there is, of course, your thing, which is why we said, Okay, we cannot solve that just by working remotely and having two hour workshops or something like that. This is really why we proposed if we want to get over this, we need an onsite workshop for one week, all the important data domains together at one table. And then we have one day to discuss about customer one, day two, and discuss about production app, product material and stuff like that. And then everyday we can focus on on another data product. Nice to get over.

Koen Hendriks
Let's wrap it. Wrap it up. Do you have any other things important problems? Best practices? Key takeaways.

Energy consultant 1
Actually, not? Not really. I already talked a lot. Yeah. Emphasize on all the important topics during my explanations. I do not have nothing to add at the moment. Okay.

Koen Hendriks
Nice. Yeah, I think that I can, because you're just so mature with this whole implementation. I think that this could really be like a nice, really contribute to like a good understanding of best practices. So that's really nice, then I think we wrap it up thank you very much. I will stop the recording now.

TDIT2 Design and development interview Energy 1 Energy consultant 2

Koen Hendriks
The recording has started

Energy consultant 2
Alright, yeah, I started Yes, please continue. Okay, good. Yeah, I was I was about to start with in general my experience in in data mash that, I think we had quite a comprehensive program at Energy 1, which was one company and the other one, which we now recently kicked off at Energy 2, which has a completely different approach, simply because the company is different by nature, from the mindset from their history, from their size, and so on.

So Energy 1 has very big one very long history with very established organization structures and processes already. So I would rather say it's a company which is not, not not not willing to to change much. So changes really takes time. It was a top down approach to introduce a new CDO organization, as News Chief Data officer organization, which was supposed basically to push from top down the principles and paradigms of data mesh in all matters, covering new roles that get introduced new processes, different data mindset in terms of data sharing. And on the other hand, as well, in the technology side, to introduce a complete new platform, for enforced basically, to the organization to to use that one, where it's at Energy 2, it's rather a smaller company, which is quite new, very flat hierarchies, very democratic by nature, very open mindset. And there we can, rather, so there's a bear, they have a chief transformation officer introduced now in January, and the whole project is also assigned to that organization. So it's basically in the middle of it and business, which is, from my perspective, the perfect approach, because doesn't make sense to put it under business. And it doesn't make sense to put it under it completely. And there they are, it's rather a bottom up approach. So they define the data domains, they define also, we'll start already building up certain platform capabilities, and then approach basically the business units and convinced them from from the benefit of sharing data, basically, and leveraging data in analytical applications.

So It's all in the in the Transformation Office. So they have a department which is called digital projects. And in this department, they have a sub department which is called analytics, a date data integration and analytics. And in that organization structure, we are now building up a data and ailment platform which consists of Informatica IDM, CS or cloud based MDM system deployed on an Azure Stack. So they have Azure as a as a cloud hyper scalar selected and most probably we are currently helping them with finding the right tool for the data warehouse component of this target architecture they they pursue using snowflake. So these are the components we're currently starting with. So having a data lake as landing zone for any kind of data, having IDMC from Informatica as our MDM tool, which they want also primarily used for any kind of data integration data quality processing activities. And we have the most probably snowflake as a data warehouse component on the front end side, they primarily use Power BI them to to the Azure Data Lake as well. So, this is the stack they are using and where we helping them building this app. And besides that, we have a team introduced which is supporting them building a center of excellence. So this is an organization which is supposed to define standards define or make architectural decisions define blueprints probably also work on certain POCs which then will be passed into the platform team who is building up the infrastructure, which will be passed into the data Team which supposed to build up data products within the data domains, as well as which will be passed into the so called value generation and analytics team, which builds data assets out of the data products so basically consumes these data products in order to generate finally value for the business units.

And these data assets can be anything because there's always a huge discussion of the data match concept, but what

is now a data product actually. So, the way we define it their data product is is more or less the data according to all these definitions that you already know right in terms of data quality, reusability and so on. and security and whenever you start consuming these data products within the data domains in order to provide insights to certain end user, be it a report, be it an analytical application, or be it even a AI solution, so means even a machine learning model, which consumes the data, and which will then be used in various contexts and scenarios. This is all then called data data assets. This means in the end, we are talking about a center of excellence on the one hand side, a team, which is building the infrastructure of the platform. So setting up IDMC as a data lake and so on the team, which is focusing on data products, and a team which is focusing on data assets. And these data assets, data product teams, in fact, more in most cases are merged into catalyst teams. So it's actually in the end one and the same team. But we distinct should pick between both because you need different skills. So in the data product team, you rather have data engineers, so people can build the pipelines, security engineers, and so on and in the data assets part you have rather people, data scientists, ml engineers, bi developers, and so on.

Koen Hendriks
Okay, so, from that I have two questions. You talked about the difference between data product and data assets? Is it correct to say that data assets is then the analytical part when there's analysis done? And that's the output of the analysis and data product is before the analysis?

Energy consultant 2
Yeah, I would I would rather explain it around a marketplace. So data product is rather than data provisioning, sort of to the creation of data, which is as a product and data assets is rather the consumption part. So any kind of application analytical application or whatever, which consumes then these data products. So what would rather come from the provisioning and consumption.

Koen Hendriks
Okay, so then it is more the component that is consuming a data product is a data asset or multiple data products.

Energy consultant 2
Exactly. Yeah.

Koen Hendriks
Okay. And you talked about the Center of Excellence. That is like a group of people, I presume, are they're like representatives of the platform have the data analytics of the data team.

Energy consultant 2
Exactly. So, it is actually a group of people that belongs in that case to this CDO organization, something similar we had at Energy 1 under the CDO. And this is basically a set or a group of people that are rather Senior which are architects for instance, which are data governance experts and so on, where we define very specific roles, I think the most important one is there is a platform and service lead in that COOE. So, the person who is in the end accountable for any for the target architecture for any new component for for the evolution of the architecture into the future, in order to enable new use cases, there is a data governance lead included in that organization. So, a person who was in the end responsible for all the processes around the data mesh paradigms, right, there is a data and analytics lead. So a person was rather a Data Architect and responsible for how data will be consumed. In the organization, and these three roles are in the end, the people that will do make decisions at the very end, and also who are a very important role in the approval process. So ideally, out of this center of excellence, standards and blueprints will be defined according to the people from data and analytics can can can implement the use cases. But in fact, of course, there are also use cases popping up for which you don't have a standard or blueprint yet, which are basically new. And then the question is, do you really structured in a way so that you say, Well, we are federated a governance completely here so that team can do whatever they want to in a decentralized way, what do you to a certain extent centralized at least, the approval process means that every team needs to go into in that case as a COE board in which they are present basically, and the solution design have a use case for which there is no blueprint yet in order to get the approval from the platform lead from the data governance lead and the data and analytics lead. And this is actually an interesting part because you can have different opinions on that, I think advantage of having such as UI and the approval process as such is that you ensure a certain certain quality, you avoid that teams are working independently on new data products on new and data assets. And you in the end mess up actually the your new platform. On the other hand, it's of course, not really scalable, because if every team needs to talk to this coe you basically you block innovation always right, because everything gets delayed and therefore, we have on both sides. So, at energy 1 and also here we are pursuing an approach where we say this whole governance policy needs to be seen out of a dynamic perspective means that at the beginning, we might centralize it in order to have a certain mature set of of standards and blueprints and also to have a certain amount of or degree of maturity and experience in the federated teams. And once you have reached a certain maturity level, a certain level of experience in the organization, you can further decentralize it means that you can resolve these approval process by having rather an advisory process that you can of course, talk to these people that can give advice, but they will never ever be mandatory and required to to proceed in the process of development. And that's basically the approach that we have aligned in both the clients yet.

Koen Hendriks
nice and actually new in other cases it also like doing this kind of flexible governance of slowly in time seeing whether they can find good equilibrium between centralizing and decentralizing so as far as that there are other people. [. . .] So, if you How did you define the roles, responsibilities and accountabilities?

Energy consultant 2 13:37

Yeah. So, there, we started actually, first of all with with some kind of a business glossary. So, we came up with with a list of terms and and definitions and specifications of in that case roles that are required in the data mesh concept, right. So, that means we had for instance, in the data domains, we had the data domain owner, so, every data domain just had one single data domain owner under which you can have multiple data product owners. So, the data domain owners then the overarching accountable person for any

data product that gets created in his or her domain, the data product owner or rather those that are responsible for on an operational level to specify these data products and also support the entire implementation and also to beyond the implementation phase, to decide probably whether a data product needs to get retired or not. Under these data products, you normally build a team of a data steward and the data custodian for a particular data product. So this means if there is a demand for a new data product that pops up in the data domain, and the Data Domain manager says yes, I will be responsible for it and I see the business case behind this data product so I will build it up and and share it in the central marketplace. Then he assigns a data product owner for that, and the data product owner assigns, a data steward and a data custodian. So, the data steward is rather the functional operational person who will understand and also maintain the, the, let's say the the logical, the logical data model, the meta information about this data product and so on. So, everything which is related to functional requirements, the data custodian very easy as rather the technical person usually still part of the data domain. So, it's still part of the of the functional business side. But this data custodian is rather looking at the technical design at the test concept from a technical point of view, and so on and so forth.

And these roles we have defined on the domain side, and then on the on the, in the central unit, so be at the center of excellence and so on, we define these roles that are important for the approval advisory process. And this was read more or less similar like to what I explained with the platform lead, the data governance lead and data architecture lead, and the data analytics lead and so on.

Once we have defined this glossary, it was important that we get a common understanding across the business units of this. So therefore, we have built at energy 1 a single data strategy team per domain, who were then reaching out to the various data domain owners, in order to make them understand the whole concept and convinced them from the overall idea of data mash, even though it was pushed on top down, however, you always need to make some some kind of homework in regard to convincing them and making them having them all as a joint and common understanding of the different roles. And then it was about assigning people to the respective roles for certain data products. And this was then handled use case by use case. And so there was no full waterfall approach in terms of you define this once and then you assign for any use case, data stewards and data custodians and data product owners, it was rather a two fold approach that we said, we funnel now all the use cases that are out there.

So we had another team, which was basically looking at all the use cases, which were more than 100. Then selected 15 as the as those that we want to start with first sort of decision criteria were on the one hand side, what what are the use case was the highest business value? On the other hand, what are the use cases, which is also from a technical point of view feasible at the beginning already, so that it doesn't delays? In order to build up the technical capabilities first, before we can start with these use cases. And as a result, we had 15 use cases that we started with out of, I think, was five domains. So therefore, we could pick up these use cases in order to approach the domains and tell them look, we would like to stop novices use cases, we need to have a data we need one or two data product owners, we need some data stewards, stewards and custodians. As a result, of course, they have no not such skills. Right, especially data custodians, you haven't found that any of these data domains as they normally do not have the technical skills that are required for that.

So therefore, either we from Deloitte stepped in and took over this role, or someone from the central organization from the Center of Excellence or whoever could also step in in order to take over this role. And this was then basically the the approach that we said, if there is a skill not present yet in the department and the decentral federated teams, then the knowledge gets actually spilled over by putting in people who have the knowledge, right, and that they can either build up new people in that in that group or yeah, they can save let's let's proceed like that. Because I don't have money for for these people.

This is different from Energy 2 as we just kicked it off and they are by far less I mean, especially due to the size I doubt that they will have a single person per role, it will rather be a person probably that takes over multiple roles as they are by far smaller and we are at the very beginning. So we are currently actually rather coming from the platform side building up the first supplier domain. So we start with the supplier domain. So therefore we just have one domain counterpart currently. And even this counterpart Is, is partially covered by someone from the center of excellence where the chief transformation officer department. So therefore, it's a complete different different approach. And we are at the at the very beginning of the journey. So therefore, let's see whether this works out.

Koen Hendriks
Nice. Okay. So you've talked about the roles, accountabilities, responsibilities, you also talked about capabilities, like the capabilities of the person that need to address these roles. So that's nice. You also talked a bit about interaction model. [. . .] can you tell me more about how you came up with this interaction model?

Energy consultant 2
Um, yeah, for that we have defined the process. First, which we then aligned with the with the respective data domains that were chosen as one of the first use cases. And in that process, you basically, we defined these feedback loops. And so we started from a certain demand for it for a data product, which then the federated team was supposed to start with. And then they start with the specification phase, which they can rather do independently. Because they specify then from a functional point of view, these data products, and from a technical solution, design

and solution architecture point of view, there, they already needed help, because there, they had more technical experts in that team. And then the feedback loops were as follows: you go first into the data governance board, in the data governance board, you you could clarify whether the data product that you actually want to specify, is already probably in the marketplace, present or not. Because the idea was also to resolve these data silos in the organization. And these silos, of

course, if people do not know from each other, that data is actually already available somewhere. So therefore, the domains themselves can't know that. That's why they need to go first to the data governance lead in the central organization, let's call it COE for the time being. Because he has the overarching view on what's going on in the entire enterprise. And they also have to tools in order to check is are these data products already present in the marketplace? Or is partially already available, but just some attributes are missing? Or is it completely new, these three scenarios are conceivable.

And once this has been unqualified, so we also call it the qualification phase. And then you take what guys, we give the feedback back to the data product owner in order to advise him on how to proceed. So let's assume now, it is a complete new data product that needs to be implemented, that the specification phase kicked in. And in the specification phase, they have they came up then with the solution design and with the with the specification of functional requirements. And this has been presented to the platform board. And in that platform, the board the platform lead that was present and also the data analytics need was present in order to check whether the way they want to implement these data products is actually aligned and ensuing with the guiding principles was the design principles that they have designed, probably also with some blueprints that they have already present and so on and so forth. And then there are different results conceivable. So either they say everything is fine, you can go ahead, then they go ahead and do the implementation and publish it on the marketplace and that's it. Or they say, Look, guys, you need to change your solution design. You need to use another tool that you actually wanted to use because our standard tool before ETL is I don't know, is Alteryx or whatever. Right? So these kinds of advisors they gave back so that they can be find a solution design go back to the board in order to get the final approval. The third possible result is that the central team figures out well, actually your use case is something we haven't considered yet at all in our environment. We don't have the technical capabilities, or we don't have yet a blueprint for that because we haven't thought of it Let's take an example. They come up with a Iot streaming use case or something. We don't have streaming capabilities yet in our platform, you need to build this up first. So then it actually it's a task that results out of it within the central Center of Excellence, to make a tool decision to I don't know, make a certain POC to create actually blueprint for future use cases. Why is it important, because if you do this once, any other team that will come later with a similar use case can make use of these blueprints and the experience that were gained in the cu e. And that's how you basically avoid that each and every team is actually going through the same process hundreds of times, figuring out how I can realize an IoT streaming use case, if you just always do this once out of a team, which is very senior by nature, very experienced, who comes up with a blueprint. I mean, the drawback is that the team that comes up first misses requirements needs to wait, right? That's a bit of drawback. But on the long term in terms of scalability, of course, by far more efficient. Yeah, and these are basically these two feedback loops. So first was the governance lead. And it's always important that you have this loop as the governance lead first, because I also talked, for instance, to another client who's also intending to implement data mesh, they do not have this feedback loop with the data governance yet at all. So they have it quite late. So they start implementing with something and then later on, once it is implemented the acid data governance, so it's this data product that we have created. Now some are fitting in one of the data domains, and then it turns out no. Well, it turns out, well, actually, you have not spent a lot of money into a data product, which we already have made, provided, published by another team. And therefore, it's very important that data governance, feedback comes first before you talk to the platform needs, because then you can avoid sunk costs by creating redundant data products that are already out there. And also that you have right from the beginning, before you start with implementation, someone who feels accountable and responsible for the result of this data product. So you talk about the business case behind the data product, by far earlier. And that's that that's how you change the mindset of an organization in terms of every data handling processing, every data asset that we want to build must somehow be cost efficient as well. And this kind of discussion you only have if you talk to the data governance lead first.

Koen Hendriks
And do you perceive that as scalable, the connection between the data product owner or the perceived use case owner, maybe he wants to check whether a specific product is already on the market? So he then he goes to the data governance lead? But then still, you have like one person everybody has to go to me that can cause problems? Yeah. And how do you? How would you solve that?

Energy consultant 2
It's exactly the same same approach as the one that I explained earlier at the beginning. So Energy 1, for instance, they did not want that at all, as a as an approval, they said the data governance leads should always be a facilitator. So he should just check and qualify the demand for the data product, and then orchestrate basically meeting a man potential data domain owners, but you should not be the person saying You are not responsible for it. And you should also not be the person who gives an approval. So they said, We don't want to have approval processes, or we just want to have the central organizations as an advisory unit. However, in that case, you can you could pursued a similar dynamic approach that you say you centralize it for a certain timeframe, depending on the on the company itself, how this works out for how long you will centralize it. But in best case, the data steward and the data product owners in the various business units are actually capable and enabled to do this research themselves, so that they know how to use the tool of the marketplace in that case, in order to figure out is there already a data product, and also, we have introduced an ontology tool, which they use as well for the for the entire data modelling and also the to train the people the data stewards and data product owners in using that in order to figure out is it maybe

just a synonym? So is the data product that we are searching for? Is it probably also under a different name, but present already in the marketplace? And these kinds of correlations you could capture with an ontology

tool. So it's on best case if they are enabled to use these tools. And if the process is probably also automated already. So they use Alfresco to build an automated workflow in the end. So if all of this is in place, then of course, it's scalable, right, because then they can work independently on that, which which should be the vision target picture.

[..]

Koen Hendriks

If you if you like, look at the process of what is the first thing you would do, while designing this whole structure?

Energy consultant 2
Yeah, so the very first step should be analysis of the use cases. Because if you if you don't know which use cases you start with, then then it's quite difficult to to to kick such a journey off. Because you can't say you define for us all these roles and then you let everyone start with any use case, that's that's impossible. Because then also from a platform perspective, you get overwhelmed and you can't handle all of that at once. So that's important is this use case funnel is I think a very crucial part for the success of the of such a journey of such a transformation because it's, it helps you to create momentums and success stories that you can convey on both sides on business and on it side. If you if you if you select them in a smart way. Because the worst thing is, you start with a use case which is super complex, it's hard to define or find your domain owner it's also difficult maybe to implement it and then it takes ages and people lose basically the confidence and and and basically do not believe that your intention is working out that's why I would always start with the with the use case analysis. So that you select Use Cases where you say this is a quick win on business side. For that we already have the capabilities and on the other hand, you select a few use cases where you say probably a bit more complex takes a bit longer but already requires technical platform capabilities that we can easily set up right because then you have a success on on platform side and on business side. Think that's the that's the most important piece that I would would start I'm working on

So this whole use case funneling, I think it's very important. The other one where we make good experience is starting with the Business Glossary early, because we also started at energy 1 probably rather from an architecture point of view, what kind of technologies and tools do we need. And then once we have set up the data catalog, we, we actually started reaching out to the data domains, which was by far too late. And therefore, having this this business understanding a common understanding of the new roles of the of the value add of data mesh, and so on. So defining all of that should be one of the first steps in order to convince the most important stakeholders, probably also figuring out in the organization in the target organization, who are the decision makers, who are the important persons who are probably also persons, and where, you know, if you convince them already quite a network of people are convinced because they are also influences, right. And so I think this this, this, this would probably be very important in such a transformation journey. Yeah, and then on the other hand, building authentic capabilities on technology side that are required for the very first use cases. And there, it also depends on the company, what what it actually means. Because some companies have already some some components, some capabilities, others don't. And therefore, yeah, these are the two areas I would say. So use case funding on the one hand side, and the business glory glossary, and the most crucial technical capabilities on the other hand side.

Koen Hendriks
And these processes seem like they are more part, the first steps of the implementation to get like, get a living in organization. And if you take like a few steps back into the design of the structure of the how teams interact, whether you need the Center of Excellence, who will be in its center of excellence, if there's a data governance board, decisions like that structure, that organizational structure. Wereyou also part of that specific process like designing that structure?

Energy consultant 2
Indeed, this was something so the Target Operating Model, it was called, which was in the end, I think, a deck of 150 slides. So this was one of the very initial work streams. However, this Target Operating Model back then did not come include this operational collaboration model. Let's get caught up like that. So these processes that I explained to you now, it's a different feedback loops and so on, and the way how we continuously build up capabilities in the studio II and on the other hand, enable the use case teams, this was not part of the Target Operating Model. So therefore, you could say theoretically, the Target Operating Model to have two roles defined and then break this down to an operational and operationalize the operating model, right, considering the circumstances of the of the individual client. Now, so very important point, the Target Operating Model must be in place at the beginning already.

Koen Hendriks 16:21

Okay. Nice, have you any of you have, do have any other additions in types of best practices that I should definitely take into account while creating the design guidance for this operating model?

Energy consultant 2 17:02
I think the way to specify the domains, I'm not sure to which extent you have talked to Energy consultant 1 in regard to that, because I think Energy consultant 1 also contributed to a white paper, which is capturing that topic. Because I remember, this was also something where you don't find many references or experiences, if you just Google

for it. So how do you structure these domains, because there are different ways either you structured according to the processes, the global processes and organization or you take the organizational structure. So the corporate functions, or you, you're gonna use systems, right, and all of them turned out not to be the right approach. Because in the end, you want to break silos. So therefore, you rather, I think they call it context based domains, which you build, which, which goes then across functions and across processes, on purpose in order to break the silos to break to bring different people together that haven't been talking to each other yet. So that's the basic idea, nothing there, it makes sense to spend some thoughts into it before just running around and creating the first domains. Now, have you talked to Energy consultant 1 in regard to that, because I think he would be even the better person.

Koen Hendriks
I've talked a bit about in the domain specification, how they did that, and why they did it that way. And also my first interview with Energy consultant 1, I also talked about the context based domain specification. So based on the use data, and not specifically based on the special business divisions, but the context and use of the data. So I talked a bit about that. So that's, that's nice. Any other best practices I should take into account? Like maybe on the organizational side?

Energy consultant 2
Yeah, I mean, the other one is this hybrid federated governance, I think there's this dynamic approach that you basically do not that you centralize first the knowledge and then you continuously release it and federated into the into the legal entities and there, I think it's the it's the it's the art and the challenge to find the sweet spot on when do you stop the centralization? When do you start releasing responsibilities and capabilities into the business units? The other one is the general twofold approach means that you have a use case driven approach on how to build up data mesh capabilities, that you do not have a waterfall approach, but rather, create a use case funnel pick the right use cases, which provides business value early In which uses already existing capabilities, but also builds up new capabilities that are required for the more sophisticated use cases. So these two fold approach, I think it's a one that was very valuable in both of the clients now. The other Yeah, I think that's, that's important.

Just thinking regarding the technical capabilities, because they're also turned out to be a good approach to start with the business glossary and data catalog early to to not make the whole data mesh transformation, a technical it topic, but rather a business topic. Because then you need to involve business right from day one. So involving them from day one is very important. So therefore, business glossary and, and data catalog as something to start with. And then probably as practice as well to make the stakeholder analysis because I think that's in general, our recommendation for any transformation, that if you want to change mindsets, or change ways of working, that you understand who is who's an influencer, who's the decision maker here, and that, that in these domains, and whom should I convince first in order to have network effects in terms of reaching out to many people by just convincing a single person which is imposed important? Yeah, that's important point, another one probably in regard to the delivery.

So organizational wise, we also I would say, learned a lot when it came to not having a transformation office or nerve center, or corded program. Overarching instance, which, which basically brings the different pieces together, because we had at Energy 1, we had work streams for for governance, we had a work stream for platform and services. And we had a work stream for value generation who builds the first data assets. And there was at least not enough, I would say, not enough resources, planned on an overarching level, who brings all the different pieces together. I think if you Google it in or if you look in our Deloitte network, you'll find the term nerve center for for transformations, or you find a transformation offers when it comes to cloud transformations. And the idea is always the same that you have a central overarching project organization unit, which brings in the end all these things together, which orchestrates them and tracks them and so on. And so something like that we haven't had at Energy 1 or not, to the extent needed, because we have so many people running around. And now at at these other clients, we also don't have budgeted for some like that. But we we are small enough to to handle it out of the PMO currently, but it's on the long run, also not handable. So I think that's also another best practice advice, I would say to consider something in the delivery of such such a program.

Koen Hendriks
Nice. Okay. Yeah, I can do a lot with it. I think. Some of them are new to me. And some of them I'm sure you score. So although that is interesting. I think I have enough that I can analyze this together with all the other interview. I'm going to reach out to Tom is his name, right. You sent his? Yeah, Tom. Okay, I'll send out an email to him. Thank you. So for the context, I think it might be really helpful. And also, thank you for your time. And do you have any questions for me for the last few minutes?

Energy consultant 2
I would be interested of course in results. So whenever you use summarize and make your conclusions would be very, very interesting to see and know that. And also, probably we can stay in touch because I think the more we proceed also with the current client now into the data mesh area, and probably new insights will be come out of it that we should exchange. [...]

TDIDG1 Design and development interview Data governance expert 1

Koen Hendriks
Could you please introduce yourself and give a brief introduction on your data governance and data magic virions?

Data governance expert 1

For sure. So my name isData governance expert 1, I'm a manager within Deloitte Consulting, I have a little bit more than six and a half years of experience with the data governance and data strategy. We'll touch upon my experience a little bit later, maybe first, starting with the industries that I've worked in started off in the financial services industry, so did a lot of projects for banks and insurance companies. And in the last three years, I moved to the consumer business. Overall, what I noticed is that within within financial services, there's much more regulation, of course, so the maturity when it comes to data governance, and data mesh is a little bit higher than than consumer. But within consumer, you see that it's becoming a much more important urgent topic as well, which I really enjoy.

So the projects that I'm doing is, I always use the name data strategy and data governance. It's quite broad. Actually, it's all data management, I would say. So starting with, with data governance and data management, I've been working on a lot of data governance projects, where we really built the data governance foundation from the start. And with that, I mean, there was not a lot in place. So I really supported the client in creating the data governance framework, creating role descriptions of a data owner and of a data steward, helping them with the data governance operating model, which we will most likely touch upon later as well. But also, we're executing data governance, so that is already assigning these data owners and data stewards working with them, onboarding them, showing them how they should be successful in their role. I, in these projects, well, so set up data governance councils, where a lot of senior people meet regularly to discuss strategic data governance topics. I've been doing data strategy projects as well. So that is more a little bit more on the strategic side, obviously, which we sort of dive in what the client actually wants to achieve with data. So a lot of clients do see that the Googles and the Amazons of this world, and want to create value from data as well, but it's difficult for them. And they also don't always exactly know what value they can get out of data, and what they should focus on. So a lot of the clients have low maturity when it comes to data analytics. And then they ask us, like, can you help us with setting up a data strategy so that we have a better direction in the coming five years? So help us to determine what our current maturity is what we should focus on what data topics are important, like, is it data quality? Or is it data governance, or is it data architecture, but also really showing what the connection with the business strategy and the data strategy is, so this is more, I would say, strategic project where you do a lot of interviews and a lot of workshops and a lot of sessions with senior leadership. And eventually it results in a deliverable or roadmap or strategy. And I've done some projects on data quality as well. So often, when you touch upon data governance, you sort of connected to data quality as well. But I, for one of the insurance companies, I created a data quality policy and for the food company that I've worked for in the last three years, but also with a lot of data quality projects where we work with the data owners and data stewards that we assigned to help them and sort of build data quality roles and set up processes so that they could actually solve their data quality problem. So overall, a lot of data governance, data management, data strategy, and I see data operating model as part of data strategy. And then I also see a difference with data governance operating model, but let's leave that for later in the in the interview.

Koen Hendriks
Yes, yes. As you mentioned, there is also recognize the difference between an operating model data strategy and data governance operating model. And this interview is mostly focused on the process where there's already a data strategy and data operating model in place and then translating that towards a data governance target operating model. And with that designing that operating model, designing the processes, the roles, the interaction between the roles and teams. That is the main focus of this interview. So it's nice to hear that you have experience in that. Okay, thank you for the introduction, then, let's dive deeper into the creation and development of those data governance over any models.

Can you explain what is like the normal process of designing an operating model like that? So

Data governance expert 1
the normal process, I wouldn't say that I always follow the same process, because every client that I serve as sort of a different starting point, like you said, some of them already have a data strategy in place and are a little bit more mature and have a lot of data talent, and they have teams already decentralized everywhere, and others have nothing. So

what I would do normally to when when client asked me to support them with setting up a data governance operating model, I really need to know what the current status is. So what type of data people do we already have? What type of are there already data owners and data stewards in place? If so, where do they actually reside? Is it in the functional teams, are they in this more central data team, so I'm collecting a lot of information to understand how their data landscape looks like, but also talk to people to understand the maturity of the organization. And this is something that I need to know. Because you can imagine that when an organization is not that mature, it might be helpful to, for example, have more centralized operating model model where we have more people centrally in one place, because it's from there easier to sort of distribute the capability to the organization and to the domains, you know, so in the beginning, I tried to collect all this information.

What we normally do is also show some reference models. And to be honest, when it comes to data governance, there is not really one reference model that I always use, what I tried to do is to show the client different type of models that we apply in other types of organizations that are sometimes similar to this organization. So when it comes to reference models to apply data governance, those we have, so for example, there are multiple data governance frameworks that we use within within Deloitte, there is the Deloitte data governance framework, I my experience is that in the last six or seven years, I've seen many and if I start these projects, most of the time, I work together with the clients to fine tune the model a little bit. So that works for them. So for example, some of the models are a little bit too much focused on policies and processes. And there's not a lot on roles and responsibilities

and data quality in there, and then I sort of fine tune the model a little bit, so that we have a good framework or a good model to apply. But going back to the operating model, so collecting a lot of information, understanding what people there are already there, like I said in the in the organization, also where they are. So for example, if I project that I did in the in the food industry, we noticed in the beginning that some of the domains so for example, supply chain team already had some data people, and already some data stewards in place, but the commercial domain and the finance domain and the other domain did not have anything in place. So there it wouldn't make sense, for example, to go full data mesh and say, Okay, our goal is to have data or create a data governance operating model where all the capability of data analytics, but also all the data owners and stewards are in the domains. Because we said that that's a little bit too soon, because there is there is no one yet. So then, based on the what we often do is we show the client, what the options are, right? So most likely you've you've seen already the slides where it goes from centralized to decentralized and These, we used to have a discussion with them, you know, we showed the pros and cons and we help them a little bit to make that decision what would fit their organization best, then at some point, you create the first draft of the model. And then it's a matter of validating, of course, with the right people making sure that senior leadership buys into the operating model, and everyone is ok with it, and then you start to execute it.

When it comes to design principles, I think that's a good good bridge, your next question? So, I already said that, I look at the maturity of the organization. Where it would make sense if the organization is not that mature, to do to not have a full decentralized data governance model in the beginning, and try to, for example, set up some data people and maybe Data Officers, for example, in a central team that support these more decentralized domains, right, but when, when already, you come in a see oh, I have seen domains like supply chain and, and finance and commercial, who are already doing data governance for a long time. And they are they have data owners in place the data stewards in place and doing a really good job, then maybe it's better to have a smaller central team and push more decentralized. So that's something that I also look at, then what is really important is, are we talking about the global organization or the local organization. So this is information that we really need to decide how the operating model will look like.

Because it's often quite challenging to decide where you're going to, for example, assign these data owners and data stewards. I look again, at one of the companies that I worked for, it's a really, really large company. So it's a global company, which has four enterprises. And then within these enterprises, you have businesses and within each business, you have regions. So then it's really difficult to decide, decide like, are you going to decide to assign data owners on a regional level? Or are you going to put them on a more enterprise level? Are you know, so there are a lot of decisions that you need to make. And I will say this is information that we try to try to collect upfront to, to this eventually decide on where it's best to, or how this operating model should be designed.

Koen Hendriks
If you look at the reference models that you that you have, and also like these different design principles, do they sufficiently support the decisions that you need to make in a data mesh context about the maturity, the scope and decentralization?

Data governance expert 1
No, definitely when it comes when it comes to the data mesh, so I'm, I'm referring now to the project that I did where we sort of implemented data mesh, but without calling it data mesh, and it's, it's only a fraction of data mesh, because just maybe a quick explanation, what we did there is, we said, our goal is eventually to have sort of self sustaining business domains when it comes to data and analytics. So we created a plan where we said, these are the type of roles that you need to have eventually, in all of the business domains. And this is us centrally as a data analytics team. These are the people that are in the team and this is how we will support all of you. So that's the only project that I did I did there. And our reference models that really helped you with that decision making on what a little bit. So, for example, within Deloitte, there are points of views on on data mesh, you might seen it, where they, for example, give you some guidance, like this is how self sustaining business domain could look like you have three to eight people, for example. So you do use those models, and you try to get some inspiration of other projects. But my feeling is, is that because the topic is quite new, there are not a lot of models,

or use cases out there that really help you with setting it up. You know, it's quite new, it's quite innovative. So I look at it. But I noticed that it's not such a model that easily will help you through the entire process.

Koen Hendriks
And where is it exactly lacking them?

Data governance expert 1
For example, one of the important thing is the type of roles that you need in each of the business domains and also how many people you need, you know, and this is I think this is really really dependent on the type of organization that you have and the goals that you want to achieve. And I've seen models where they say, three to eight people. But if you have no one in the domains, yet, it's quite difficult to already work with these numbers, you know. So the models that I saw, were showing how the ideal world looked like, and not really taking into account that a lot of these organization wants to start small. So, you know, how do you actually start with all of this? How do you start with data mesh, for example, in one of the domains, one other thing that I think is really showing that this whole data mesh concept, and the model that they have is really, the ideal world is that these companies have a really hard time hiring people when it comes to data analytics, mostly. And when you talk about a data mesh operating model, you need a lot of people, you need a lot of data analytics, talent, you need to have domain owners, you need to have data engineers, you need need all to have all these roles, but they're not in one place. But for each of the domains. And that is sometimes something that I think is a little bit overlooked, that it's difficult for these

companies to hire all these people, you know, so you need to work with the limited talent that is in the industry or in the market. And it's not easy to sort of set that up and eventually work towards an organization where you have perfect teams and all of the and all of your domains.

Koen Hendriks

That's actually interesting,because that's part of question seven, let's quickly go to that one. It's about a availability of the capabilities. So, the data capabilities, as you said, it is it is a problem, it's also a problem recognized often in literature and in practice. Can you elaborate on this problem?

Data governance expert 1

So if you, for example, are facing difficulties with data analytics capability, so you don't have it internally, but you also have difficulties to hire these people for your teams, then it would make more sense to start with a smaller central team that can serve more of the decentralized domains, right. So obviously, you cannot help all of the domains, you cannot build 20 dashboards and solutions for each of the domains. But at least everyone can see what the value of these capabilities are. And then from there, one of the important jobs of the central team is to distribute the capability. So most of these teams don't know what they are looking for. And that's that's exactly the support that they need. So you should be really specific in these are the type of people that you need. And these are the ways that we can support you to hire these people. Training, for example, could be one of them, you know, so there could be already a business analyst somewhere out there that could take the new data mash branded role.

Koen Hendriks

So yeah, so you would suggest to other companies, for example, choose to take the consulting model, but then on the capabilities itself, so they lend out their own capabilities to make sure that the domain has the capabilities. But as you would solve this problem is more, making sure that the team knows which capabilities they are missing, and then helping them getting those capabilities instead of lending your own capabilities towards that team.

Data governance expert 1

I think I think both both is is an option, right? So let's say that you have four roles centrally that are focusing on ICT and developing policies. Setting up the strategy And for other people that you could sort of lend out to these domains, that they can help them setting up that capability, but also executing, for example, something that they want. So let's say that they want to have a dashboard, you could say that, okay, I have this visualization expert centrally in my team. And we decide now that this visualization expert in the coming six months will sit in the supply chain domain and in the finance domain and help them with the dashboards that they are creating now, then it's a perfect, it's an ideal solution to have someone next to this person so that eventually they start to learn how this person is building these dashboards so that eventually they can build the capability themselves. So in the beginning, we, we don't expect everyone to have that capability. So it would definitely be also an option. To have it centrally. To give one example, what we now have, in one of our projects, we said in the beginning, at this moment, the analytics capability is limited in organization. So we have a team of six people who can do more good analytics, they build dashboards, they can they can, we're not talking about AI at mostly that's called a dashboard, they can work within digital products, they can do really good analysis, for example. And in the next two years, we'll make sure that these people sit, for example, 50% in a central team and 50% in these domains to

help these domains deliver what they want. But then at some point, we do expect these domains to hire the capability or to build the capability themselves. Right. So then, after two years, what we expect or let's say two to three years, we expect, again, that this central team doesn't really lend visualization people or dashboard people, but then goes into data science, and then starts over again, okay, new type of capability. So we now want to have do more with AI to do more with machine learning. So we again, hire, for example, three data scientists centrally, they start to experiment, they start to help these data domains. And eventually, we hope that this capability is again pushed to the Domains by training the people or by hiring new people.

Koen Hendriks

Nice, nice. So it's more like a gradually going from centralized to decentralized. And then we're capability, making sure that it's all available in teams.

Data governance expert 1

Exactly, exactly. And I don't know exactly what data mesh says, right. But I, I don't see a way to, to go full decentralized right away. Maybe, if you are really, really far with your maturity, and you're a Google, for example, I can imagine, imagine that maybe Google because they are so mature. And they do everything with data that they don't have a central data team, because every team is data on literally every single team should work with data analytics, you know, they're there, this could work. But then most of the organizations that we work for, that are really large enterprises that have been there for a long time that have already have legacy systems, but also do not have a workforce that is so super data savvy, it would be good to start centralizing and eventually go more and more decentralized.

Now maybe one thing maybe that I would like to mention, as well as we, this difference between data operating model and data governance, operating model is sometimes a little bit difficult. And we are now using sort of both in the last few minutes mostly talked about the data operating model, right? I don't know exactly how you refer to the difference between these two. But when we talk about the data governance operating model, I don't really look at setting up this organization. To deliver analytics, I mostly look at setting up an organization with the right ownership in place who can ensure the quality of the data? And then for example, publish the data wherever so that others can make use of it. Only that part. Right? I don't know. Is both of them interesting for you, or is it mostly the data governance part?

Koen Hendriks

Well, I think I think both those topics are interesting. And also I think that there are also data governance capabilities, which are important because you federated data governance, which means that the teams also need to do their governance. So that makes almost every for example, if you brand new teams to have data analytics capabilities, you also need to train those deemed to have data governance capabilities. Yeah, because they are flexible. So distributed capabilities is interesting because one part of that is also data governance capabilities

Data governance expert 1

for sure, for sure, definitely. And maybe if I look at one other, let's call it a problem with this activity with pushing the capability is when it comes to data governance. Again, in an ideal scenario, it would be really, really nice to push the capability and to say, their data domain, you are now responsible to set up or to to own the data to ensure the quality of the data to make sure the availability is good to do some metadata management as in define your data. And but the problem is, people just don't do it. Because they don't have the capacity for it. Sometimes. They don't have the skills for it, but also they don't often see the urgency, it's still the concept of the business owns the data is difficult for them to understand to grasp, because even today, 2023 people still think like data that is like it thing, right? So why are you why are you reaching out to the commercial late to talk about getting the ownership of data and talk about ensuring the quality, you know, they sometimes still look at others, whereas they are they have most interest in the data? So

Koen Hendriks

What you're saying now, it's more related to question six. So it's about defining roles , responsibilities and accountabilities. So let's pick that now as a focus, yeah. Can you tell me more about this process?

Data governance expert 1

Yeah, definitely. So again, here, there, there is a lot of theory out there. So DAMA DMBOK talks a lot about roles and responsibilities. Within Deloitte, we have multiple frameworks and multiple point of views, which lists a lot of data governance roles, you know, data owner, data steward, but also data custodian data governance, manager, data governance, lead is a lot of roles out there. And when it comes to the process, I often just I, I really tend to stay with the less is more mindset. So let's see what is really the bare minimum that is needed. And let's start with that. And eventually we can we can expand. Now, there are a few things that are non negotiable for me when I implement data governance and define roles and responsibilities. And that is, when it comes to data ownership, it has to be in the business. So we discussed that already a little bit in the last

question. The only way that data ownership to actually work if someone from the business has really interest in the quality of the data. And in that day, there has motivation to act when the data is wrong, for example, but also has, for example, budget and mandate and authority to act when there are problems with the data. And this is the one thing that is for me non negotiable. So the ownership of the data has to be in the in the business. Now. This is a concept that I would say maybe it's it's it's a small problem. It takes time to explain this. So you're now talking to, for example, senior business leads, where you need to explain like, hey, we defined roles, we have data owner, we have data steward, and we want you to assign someone within your team with these criteria as the data owner. So I wouldn't say it's a big problem, but it takes time for them to understand this part. You know, sometimes they say, okay, you know, assign me as a data owner. And then if you for example, one month later, check in with them again, they already forgot that they are the data owner and that they have certain responsibilities. So that's one of the one of the biggest problems I would say is keeping them engaged, making sure that they are not only assigned in the role, but they're also operationalized.

Koen Hendriks

Someone else mentioned that as defining roles and responsibility isn't really a problem, but bringing them to life in the team and making sure that they are aware and act acoording to their responsibilities is a real problem.

Data governance expert 1

I fully fully agree with that statement.

Koen Hendriks

And if you look at the new roles within a data mesh like the data product owner, datta product developer and the federated governacne team, do you think that DAMA DMBOK is lacking there? Or could it be interpreted in a way of data mesh principles?

Data governance expert 1

This is, this is for me a little bit a little bit difficult. And it touches upon this, this this difference between data operating model and data governance operating model. So when the DAMA DMBOK, was launched, there was no data mesh, right, no one talked about data mesh, we did talk about decentralized models when it comes to data analytics, and when it comes to data governance. So I feel a little bit the DAMA DMBOK, at this moment is your Wikipedia, it's a lot of theory. And a lot of people spend a lot of time to write the book, and it was really helpful. But slowly, we are moving away from from DAMA DMBOK, because I think it's still really relevant. But it's a lot of theory, which explains the what, but will not explain you the how, if I look at the roles and how they work, it's a lot. It's a lot, and they don't really explain also like, hey, these, for example, data owner and data steward roles are the most important one. So start with those. And then maybe later in later states, you can add data custodian, you don't have to, you know, these type of this type of guidance, it is not there. And then if I look at the roles, it's a lot and sometimes I'm a little bit overwhelmed. The core is in there still. Now, if we compare that with data mesh, sometimes I'm having a little bit difficulties. But again, we have the data owner, but then we also now have the domain owner and data mesh, is this the same role or not? So if I look at the domain owner, and correct me if I'm wrong, this is not necessarily the same as a data owner, right? Because you could be, for example,

the data analytics domain owner of a data mesh, where you make sure that you're the one responsible for all the data analytics demand, for example, and you're also the one responsible for publishing the right data in the Data mesh. But it's not necessarily always the case that you also own the data. And, and there, I'm having a little bit difficulties, because maybe I need to read more, right. But I would hope that are all the organizations that want to do this get overwhelmed really quickly. So that's why I tend to always stay with these most important roles data owner data steward now if we introduce new roles, then easy like you said, we can define them, we can introduce them, but it's really difficult to get them operationalized. So that could be something that could match a little bit better, you know, when who is actually the person owning the data within within data mesh? And how are these roles like the data domain owner and the data product owner combined? I lack a little bit of guidance there.

Koen Hendriks
Yeah, I can imagine because the I also feel like the data mesh introduces the roles pretty, pretty good. Like, it kind of explains it in one specific, purely theoretical context. But there is no connection with the current roles, which means that also, transforming from the current data governance roles to database roles, can be very difficult.

Data governance expert 1
Yeah. And again, I think it's it's a matter of, you know, eventually the industry is developing themselves and we will get more and more information on this part at this moment is still a little bit difficult. But I do see a lot of overlap, right. So within data mesh, this whole concept of decentralized your data analytics capability, we always said that with data governance, we always said that we as a central data governance team are not able to ensure the quality of your data. Because you need domain expertise. You know, you

need you need to be really need to have, for example, supply chain expertise to be able to define what good supply chain data quality looks like. We can give you the framework we can give you the tools that you need. We can help you to show how you implement the data quality rule, but eventually it's your responsible at the end. And this part I do like when it comes to data mesh, it's full, it's all it goes full, decentral, right. So we say, given that there will be more and more data analytics demand in the future, it's a lot better to not be dependent on this central team, but to be self sustaining and have decentral teams there. So I do, I do see that it is interesting. That's why I'm really interested in it. And I feel like, you know, sometimes, I feel like some of the organizations have been doing data mesh without really knowing that it is data mesh, you know, they have been trying to decentralize, and we put this capability in all of the domains, without calling it really a data mesh. Now, the whole technical component we're not talking about yet, because I think it's not also not really the focus of your, of your research, right?

Koen Hendriks
Yeah. Yeah, that's, as you mentioned, the lot of companies are already in some principles of a data mesh. So for example, we only look at one principle, for example, the federated governance structure. There are a lot of companies that are already have like a federated governance structure. But then you see that data mesh combines like domain ownership, fully thinking and federated governance into concepts. And that means that a lot of companies already have, like, 50. So that combining that is

Data governance expert 1
Definitely, exactly, it would be sort of sort of, again, maybe it's out there already, but for me, it would be really nice to have this more sort of a unified model where where we say, Okay, if you want to do if you want to implement a data mesh operating model, this is the data governance that goes with that. So we are talking about a domain owner, which is not a data owner, because this isn't isn't reason this is this is how you should set it up. And this is how data governance and data management go hand in hand. Yeah.

Koen Hendriks
That is what I'm trying to exactly

Data governance expert 1
I understand. Yeah, so I'm also I'm posing a lot of questions for you, but I hope that I can give you some some answers as well.

Koen Hendriks
Yeah, yeah, you definitely can. It's also nice just to have a lot of input of all different experts to know which things are recurring, and which are like doing specific problems, or for a specific company. So that's also interesting yet, then the next question, let's go with further given the time also, if you look at the interaction model, and with interaction model. Do you see problems there in a federated governance structure?

Data governance expert 1
Yes, so when we say federated, do you then I think we always will have some central team, right? Or if you say federated, there is no central team at all.

Koen Hendriks
A federated model is a model where there's a central team that supports all the the decentralized teams.

Data governance expert 1
Okay, so there are two challenges in the federated model. [..] Maybe one challenge to start with, is what I faced in my own projects. And one challenge that I foresee, I didn't really our experiences in my projects, but I think that that could be a problem. So first thing in this federated model, is that on paper, we are sort of able to define exactly what the standards are centrally. And then we, for example, say we will define the standards, but then you all should execute it according to our standards. The problem that I have, and there are reasons why we do

that, right, because are centrally not able to implement data quality roles like you can we'll give you the templates, we'll give you the framework, we'll make sure that you are able to do it, but eventually it's your responsibility. The biggest issue that I face in my projects is that is their domains, sound them just don't do it. So you, you help them you do everything from a central point of view. But still, there is something that it's, it's not on top of their mind, it's not sort of the first priority within their team. Right. So to give an example, when it comes to data quality issues, you can be really clear and okay, I want to, for example, to start with, I want to make sure that everyone collects their data quality issues, as a first step. And then I want, we as a central team will monitor if you guys will resolve the data quality issues. And

we will be there to plan when you are finished, for example, and we will will chase you. So that everything, all these data quality issues are resolved at some point. So theoretically, this works really well. And it makes sense. But then when you go into practice, you see that they are too busy. [..]
So that's the first thing second thing is and this is what I would foresee as a problem is I can imagine if this if these teams are really growing their capability and at some points are self sustaining. If you don't have proper governance in place, you could say, what what do we actually? Do we actually need this central team. So if you want to, for example, create a standard that deviates a little bit from the standard from the central team, go for it, we have our team, we have our data scientists, we have our data governance people, we have everything, you know. So I can imagine that, that sometimes, if they don't have a really strong leader, and the central team with the right mandate, and everything to you know, and, and really push the standards in their own way, I can imagine that people steal, if they have their own central self standing capabilities, that they start doing their own thing. And then you have sort of the risk, again, that you don't have the standardization that you wish for. Right, so they start doing their own thing. Because I haven't seen a lot of companies that have full data mesh in place with really nice, decentralized data domains that own all of the capabilities and can do everything on their own. I can imagine that sometimes people might think like, you know, this central team helped us in the beginning, but at this moment, we can do it on our own. And we don't want you to tell us exactly what the how the standards should look like, or which fields we should use or should not use.

Koen Hendriks
I can understand what it's also like the flexibility of the domains to adhere to standards or not that is it, they are very flexible, but the moment they need to collaborate with other domains. This flexibility should be streamlined through the Global standards, of course, and that's then where the central team comes in.

Data governance expert 1
Definitely, when it comes to maybe one other small example, that's more data quality, it's not really data match. But what we noticed at some point in the insurance company is that there are, it could be that, for example, there are a lot of customer data quality issues. But the customer domain is not really feeling the pain of those issues. But the finance domain is feeling the pain of those issues. Now, it's difficult to say to the customer domain, you need to fix this as soon as possible, because it's not in their best interest to fix it, it's in the best interest of the finance domain in this case. So how so that there is definitely see a role for a central team to make sure that you have some prioritization mechanism or something in place, so that you can see, okay, what are the most important issues and even if it's not your own domain, that is feeling the pain, you should be able to, for example, prioritize days or prioritize that.

Koen Hendriks
Nice, that's also part of question 10 prioritization. Given the time, we still have five minutes, if we look at the problems that we did not address yet, the dynamic topology, the computational governance, prioritization, actually, those three are not really touched upon yet. Which one of three or maybe another problem, do you think is very important

Data governance expert 1
Let me think about that. We already discussed a few things, right. So the one of the biggest problems that I have with data mesh is that it will work if you have sufficient talent also in the market, or you have people that you can train properly to have all these teams right and there I, I foresee a problem, it's, it's too easy to say, you should hire six people per domain. Every domain needs to have a data analytics product owner and data engineer and this and that, given that it's really scars, and that's why a lot of companies have these teams together still, because it's too difficult to have them decentralized.

Koen Hendriks
Would you also incluce the gradual shift to the equillibrium of federation in the data governance model?

Data governance expert 1
Yes, yes. So, to give to give an example when we Say, Let's, let's say more more federated, let's say data mesh type of operating model for data governance. Again, in an ideal situation, we right for example, data owner is accountable and even responsible to ensure the quality of the data and to define the data quality roles, and to fix all of the data quality issues. So that's because in theory, that's the we all know that that's eventually how it should work, the data owner should be responsible for these things. But in practice, when the maturity is low, in the beginning, we see that the data owner gets nervous when you start asking them these things. They say, when I look at your criteria for a data owner, I think I'm a match because I, for me, this is really important data, I am senior, I have authority, but I never worked with data quality roles before I have no idea what you exactly are talking about here. My team doesn't have this capability at this moment. So then I can imagine that you take a little bit more responsibility centrally in the beginning, and then start to slowly grow that capability decentrally. So to give one example, and that is, I don't know if it's really data mesh, but it's more on Access Management, right? So

we said eventually, data owners, you should have a say in who has access to your data. Because again, centrally for us, we can give you controls, and we can help you with friends. But for for us, it's difficult to say if this is really classified data or confidential data or whatever. But in one of the other projects, we noticed that they said, We, I don't feel comfortable doing this at this moment. Because this is a new topic. For me. Also, data privacy and data security are new topics for me. And we offer them as a solution. So okay, what you could do is that you, for example, delegate your responsibility to the central team. So you're still the owner, you are accountable for access management, but you can give the responsibility to the central team. And then we'll, when these workflows come in, we'll make sure to provide access to the right people, according to the information that you gave to us, and we will sometimes will align with you. Right? So yes, eventually, you want to really push it, you want to have a decentralized, but you notice that a lot of these teams are not ready for it. And you need to have in between solutions to get everything up and running.

Koen Hendriks
Yes, I can imagine. Is there anything else you would like to add or something that I could take with me in the research?

Data governance expert 1
I think the most important thing is I think it's the purpose of your of your research, right? What I said about I sometimes have difficulties. I really work with these data owner and data steward and now data mesh comes in, and I think it works really well. But give me some guidance, you know, are these the same roles? Are these different roles? You know, should I start with maybe these two or three roles only? It will be really great to have some some guidance there. I think maybe it's somewhere in the book. But so far, everything that I have read, I feel it's a little bit difficult. Whereas there is a lot of overlap. I just need someone bright like you to dive into the topic and to do proper research and to see okay, what are we talking about with these with these roles, and the same is different. And so now that's that's it. And I already told you this is super, super relevant.

Not just saying that to you, because you're our thesis intern. But this is really, really relevant. And this is something that I can immediately apply in my projects. People are already everyone's heard, here's data mesh, everyone's looking to it, but they are excited and nervous at the same time. Like I think it's really important to maybe last thing is a lot of companies don't even have proper data governance yet. Really think about that. So this data mesh is now a concept that builds upon data governance builds upon data architecture, and sometimes assumes that everything is in place. But a lot of the companies that I come and that I serve, and these are really big enterprises within the consumer industry. They don't have proper data governance and with that, I mean they don't have that the ownership in place for all of their domains. They don't have data stewards in place when there's a data quality issue. Ad Hoc people will fix it, but not the right people. Sometimes it you know, so we're really at the start, and you need to go really gentle with this with this whole transition. Even the banks at this moment when it I think we were together in that session with with financial organization 1, yes. You see that when what they showed looks really good. We have 50 data owners in place, we have 40, I don't know, 90 data stewards in place. But if you start to talk and to ask them more questions, like, okay, but is this, you know, are they all operational? Are they you see that they also have their problems, how to really set it up? So, I think when it comes to data governance, yes, it's out there for 10 years. But I have not seen a lot of companies that have this perfectly implemented. And now we jump out to the new concept. So this is really something to take into account, the maturity is low. And yes, we all love innovation, but if it goes too fast, you'll lose everyone along the way.

TDIDM1 Design and development interview Data Mesh expert 1

Koen Hendriks
Thank you for being here. And let's just start. So can you maybe briefly introduce yourself and what you did in the context of data?

Data mesh expert 1
Yes, so yeah, my name is Data mesh expert 1. And now I been with Deloitte for around six years for that, you actually might have a background in computer science and information management, most quite common data engineering project. But there was a point, I really started working on the data mesh topic. What I liked about that, it gives me more high level point of view, it gets me out of my data engineering cape. You know, I have, you know, think about the bigger picture how things are working on enterprise level, how things aren't being connected in us. And that's something I really like, and I saw myself growing in that direction. So that's why I made the move towards a Data Architect role that specializes in Data Management. So far, I'm not tied to a certain industry, but more interested in consumer and telecom. But as a specialist, I'm not focused really on one industry, by itself. And, but most of the projects that I've done, were, usually with consumer organization 2 and consumer organization 1. And currently, I'm doing a data mesh implementation at Telecom 1. This is my only data mesh experience, I had a little bit of involvement with consumer organization 1 as well, they may stop it, but we weren't doing that for them. So we were just kind of helping them more with with some workshops and thinking along with them, but they were implementing it with another partner.

Koen Hendriks
I talked a lot with people about Telecom 1. So i would like to hear more about Consumer Organization 1 and 2.

Data mesh expert 1
Yeah, so over the consumer organization 2, it was data engineering, mainly. So no data mesh involvement there. But with Consumer organization 1, I was a senior data engineer, was there for around a year, and we were having one of the main data products that they use for reporting. And what happened is that they created a new platform

to analyze data, and they set it up in a data mesh way of working. So they wanted to set up domains and, you know, start getting data into this self service data platform. And they thought that, okay, we can be the first team actually to migrate our data products. And so we had kind of a say, of what a data product look like, what needs to be done. So we were just, you know, we took the advantage of the first mover advantage, you can say it this way, you know, and we started going into conversation with the data platform team, on certain things. And that workshop I had, we actually targeted not only the technical architecture report, but also we touch upon organizational operational part, because we saw, we saw a gap there, there was an issue that they were, you know, within data mesh, you have multiple layers of what some might call it paradigm shifts. And what they were doing, they were targeting technical architectural part and lagging behind with an operation organization of bots. So ask them, Do you have a central data office in place or do you have federated Governance Forum? No, they have nothing. But at the beginning, because the maturity is too low, it's good to have a centralized governance office and the CDO to guide the process. And slowly bit by bit hand over the responsibility to the domain teams when they are pretty mature. So it can start centrally, but with the aim that in, in a certain point of time, this responsible need to be distributed to the different domains, maybe according to their maturity.

Koen Hendriks
Yeah, so there, and then you mean the maturity of the organizational side of data mesh?

Data mesh expert 1
Yeah, exactly. I mean, like, the NOC, for example, domain things, they really, they start first, they don't have also resources, or they also don't have enough policies and roles of they don't have enough definitions in place, what is the data product, what is the quality roles that can apply on a data product, and so on. So that's why having a central organization, or governance at the beginning is very crucial. Because those people will be a bunch of very experienced and subject matter matter experts that can set up these policies and the path or the domains to get up to speed. And then bit by bit, you will find your organization maturing, and bit by bit, you don't need to do it centrally anymore. Because the knowledge has been distributed across different domains. And then, you know, it is it is how it should be federated and every domain is enabled to perform and work in an independent way. So there's a transition periods from the old organization to the theoretical data mesh.

Something also to mention, is that that every company struggle when moving data mesh, but the type of struggle of challenge differs. So give you an example Telecom 1 is very federated by nature, everyone getting things done, very smart, intelligent company, but no collaboration. So they are really lagging behind. Anything Central. Yeah. And now this is something we are facing now. It's a very big challenge, because, okay, how how can you guys get started? Because you see, when we talk to a domain, the first domain now implementing the first data product, they always have, we don't know what to do. We have capabilities to do it, but we don't know what to do. And we still have the central data office still absent. And that's not helping. No, you know, if you go more more traditional centralized kind of company, that concept, just that you are having different domains, and things are out going out of control. It's a cultural shock for them.

Koen Hendriks
Yeah, that's, that's that is that is really difficult. You also see that the that comes back to the four principles of data mesh, you, you see that a lot of companies that want to migrate towards data mesh, they struggle with different principles of the data mesh that's yeah, that is that is that is interesting. And if you well, if something I found also with my thesis is that some components of the organizational structure they everybody does it the same way and some parts they depends on how the organization is currently organized. So, for example, a data product owner is responsible for implementing the data standards. That is, well responsibility that almost all companies have the same. But then, as you mentioned, the responsibilities of the central governance team. Those are dependent on how your organization is currently structured. Do you also recognize this?

Data mesh expert 1
Yeah, indeed, I think the some of the roles are obvious like a data engineer data steward, data product owner, these are pretty clear, but what I noticed is that like the things that the more they go to the business and governance, the more difficult to get, because those those kinds of roles may change, change from company to company. And, and it also it also it gets affected by the cultural aspect of a company. As I discussed that, if a company that is top down centralized, move them towards the data mesh, they will this will be

so challenging from governance perspective, and why it's federated by nature, that will be very challenging from also as well. So, indeed the I think the the federated governance and the strategy, how to start and where to end and how to grow the data mesh, this might that can be really impacted by the culture and nature of the company.

Koen Hendriks
Yes, and for my for my the aim of my thesis is actually to do define the design choices that you have and the resulting structure. So, as you mentioned, some roles and responsibilities are set in stone, and some roles and responsibilities depend on the culture. With my thesis, I would like to define which roles are set in stone and which role or team structures are not. So that is, that is like something I'm really interested in, because I found that some people struggle with that, while designing the organization structure, like what parts do I have to get from the book and watch which parts can I design like my organization? Can you elaborate more on that with your experience?

Data mesh expert 1
Sure. I also just a disclaimer, at the beginning, I think I'm not the most experienced person in this organization matters. So but I think you talk to Telecom consultant 2? Yes, I see. You can provide that. But I will just share what I know for now. I hope it will help. So when we look at, at the working silhouette and data, so it's a domain,

right? Yeah, and if we look at and the whole idea for domain is to bring the data people, those are the business people and here, the business people, we need two kinds of business people.

We need people who know that data, and people who know how this data can be used. And this, these two can be in one person and can be in different people. So the I'm starting, by the way, with more, not the written in stone roles. So if you look at it from that perspective, there, I think the domain owner is, is a role that can be in this case, the person who's knowledgeable about the data, but also very knowledgeable about the business needs. But it really depends on the complexity of the data products or the domain.

Because imagine that in situations like if you if you have a sales department, finance domain, they have a lot of data products. I don't expect the domain owner to be knowledgeable about every data source they have, or every data product. So in that case, the domain owner, depending on the complexity of the domain, or the data products might shift more towards the strategy and the business and application of these data products for a business and let the data specific knowledge go to the data product owner. You see what I mean?

Koen Hendriks
Yeah, so then you mean that the division of capabilities, like the hardcore data capabilities about data products can be shifted either towards the upper domain owner, or more decentralized and between product owners, which means that the domain owners only responsible for the business parts?

Data mesh expert 1
Yeah, so imagine that the domain just started, they have one data product is fine, let's have this one domain owner, but then in the, in the coming year, they have three or four more data products, then I expect, yeah, this will be too tough for one person to manage. So then the domain owner will focus okay. Are we still delivering value from all the data products we are sharing? We're publishing? Well, what is the roadmap for us what other data products we are planning to do? He is always in touch with the business. So you see, he this person is the link between all the data product owner and as the with the business. And then the data product owners, they everyday the product owner can be responsible for one or two. It really depends on how complex is the the product? So I do see within the domain, those are the most let's say changing roles or can be Yeah, they can be different from even domain to domain, say not only from organization to organization, but domain to domain. And then you have

Koen Hendriks
one more question about this just you said that it depends on the complexity of the data products. I assume that it also depends on the size of the domains.

Data mesh expert 1
Yeah, absolutely. Yeah. Well, when I, I made an assumption that a number of data products reflect the size of the domain, but it doesn't always happen. So since we are in the domain, I would say the other, let's say, clear roles, would have a data engineer, a data steward. The steward would be responsible of the quality and applying the governance policy on a data product. So the basically the link between the governance and the data products, you see the domain owner, is the link between business and data product. This data steward tere is a link between the governance and the quality policies and the data product. The stewards is responsible for the data product. And the data product owner or data domain owner, is then accountable for that.I don't remember exactly, but I think I read it once that they call it the data product developer. It's a generic role. And under that, it can be a data engineer can be a data scientist, it can be a data source. So it is any person that contributes directly to the creation and transfer of a data product. Yes, because look at this, if it's a data product can be a machine learning model, it can be also a data set that that is generated based on some analytical model, you know, so there is a data scientist or machine learning engineer involved. So to simplify things, we call it data product developer. And this person can be any person who just contributes to the creation of this data.

Koen Hendriks
And just to make it clear, the data product developer is like a set in stone role with its responsibility?.

Data mesh expert 1
Yeah, I mean, you need someone to create a data product, right? You need a data model, or maybe you need a Data Architect, maybe you need the data engineer. But what's nice when you identify this, and it really depends on the case, on some data products, you don't need a data scientist for even some data product, you don't need even a Data Architect because they are maybe very straightforward, or they are mature enough, for example, but in most cases, when you're starting a new data product, you need maybe all four data modeller, data engineer , data steward and maybe a Data Architect as well. Of course, of course, they vary in contribution, maybe a data architect would work two hours, three hours a week on this. But maybe a data engineer would work there for full time, or four days a week. So that's about it. Yeah,

Koen Hendriks
And than that data product developer role, can that also be assigned to multiple people? So for example, that that one, that's a machine learning experts, and the data engineering experts, together are the data developer of a specific data products?

Data mesh expert 1
Ah, maybe I don't have enough knowledge in that moment, but maybe, maybe, yeah, okay. Okay. Because if you look at another example, data engineer can do data science, and data science can do data engineering, can do data modeling, as well. So it can be Yeah. Okay. Yeah, but a data steward. I'm not sure if a data engineer would be able to do that the store does. There are also ways to, I would say, to organize this data product development,

because from what you see, this is the role that is most in demand in the domain. And this is the role that lacks the most within the organization. If you look at any organization, they let the most data engineers and data scientists available much in the market and every company, yes, needs a lot of them. There is a slide that explains this all.

Koen Hendriks
I have it in front of me now. The point of view, right?

Data mesh expert 1
Yes, exactly. So one option is to have an so every data product developer is assigned to one domain at the time and stays there. But this is really hard to achieve. Because I mean, especially at the beginning, you cannot just hire people easily, you know, the other option is to have a pool of data product developers can contribute to met several domains at the same time. Or you can actually just bring a developer and assign this developer for a certain domain for like three months or four months until the project is over.

So one domain at a time, but they are not tied to to one domain. think that's the second. And that's the most logical option. Oh, yeah, I think the third option wasn't feasible. So
basically having one data product developer assigned to multiple domains, because you are actually creating a situation just like a data lake. So those people they become a
bottleneck. They are lost with priorities from different domains. And they ended up like, not able to really prioritize things pretty well, because then there will be clash of
interests between different domains and priority. So that's why I think the second option would be the best. Of course, the first option would be the happy path, but I don't think
it's feasible.

Koen Hendriks
Nice. Okay, so just to, to round this a bit up to check whether is correctly, the roles and the roles, that responsibility of the data product developer stays the same. However, the team structure in how it contributes to the domains, that can be different?

Data mesh expert 1
Yeah

Koen Hendriks

Yeah. Okay. Nice. Are there more additions to the structures wihtin the domain?

Data mesh expert 1
The domain, I think that's it for now. That's what I can recall. Of course, there might be more stuff, but maybe it's good to discuss with other people as well.

Data Mesh Expert 26:11
Nice. Okay. And yeah, this is really clear. And it's also nice that your explanation together with the slide deck, that's super nice. Okay, if we then go from the Data Domain team, to the other teams, and then most importantly, the federated governance team. Can you maybe explain a bit more on what is the set in stone roles and responsibilities there, and how it is flexible?

Data mesh expert 1
I think there is nothing set in stone and CDO or central federated governance model. Because it highly depends on the culture of the company, the way they are organized. Yeah, the nature of just like we discussed. But if you think about certain roles that are expected, you of course, the the you need the CDO. And with that I do mean someone that is on top of this governance, but also has those access to the executive board?I'm saying this because there is no way I'm data mesh succeed, without setting KPI from top in the organization, and track those KPIs. For example, within the executive board, there will be a CFO is on top of financial matter, he or she, or top on financial matter, and they report as well. And they are held responsible for all financial matters. And same for the executive same floor, and so on. And for, for when when you're transforming your whole organization to a database, you need a chief data officer, you know, to report on all the KPIs. So for example, a KPI might be how many data products we're gonna publish, what kind of tangible value, I don't know how they would calculate it, but what kind of tangible value we got from from this? Did we like how many data products that we monetize this year, and so on, so this CDO with the, with the federated Governance Forum, they need to define these KPIs and track them on a timely basis. This is very important.

Koen Hendriks
Do you also like that is one role in the central governance or the incentive governance team? It's called like that, that is kind of set in stone you need to have that for almost obvious reasons. Is it also when you look at the other roles, presented by theory, so for example, the subject matter experts and domain and platform representatives, do you think that they are necessary to have there and how flexible can you be with those roles?

Data mesh expert 1
Yeah, so So one thing is, it's very important to have a data platform representative to enable domain teams to use it, and also to have self service tools. So if they don't have enough data engineers or capabilities, they can actually use some of the tools very easily to publish or use or consume the data product. Having a data platform is a must. But then either the CDO sponsor, a new data platform, if they don't have a data platform, that's what happened with consumer 1, they didn't have a data platform for this matter. So they established a new one for you actually

use an existing one, like in Telecom 1. It's hybrid. So they already have a team called CCOE. They have like, the central center of excellence or something like that. And they have an AWS platform that anyone can use. So what happened is that they are just building on it. And then they are using some other tools, like Informatica and so on to connect to it. In either ways, you need over responsible for this data platform to be part of this for very important, yes. Okay. Because because they need to be aligned on the policies of the governance on the usage. And I'm not sure about financial them, I'm really not literate about it, but there needs some to be some sponsorship. With pay for all this resources. If the new platform has been established, then I understand that okay, if CDO is sponsoring, or if its platform, some other team or some other organization who would be sponsoring, so these things need to be resolved within the the federated Governance Forum. That's a financial aspect is likely that's not enlightened in the theory book. And, and then the SMEs, it's it really you need to have SMEs from different backgrounds, it to have SMEs on privacy on risk security, data architecture. And you need to have eventually representatives from every domain as well. At the start, you don't expect the domain to contribute that much. And this is the federal government or the central data office will be so much involved in guiding and helping, but eventually, the domain should be working by themselves, will be able to do stuff by himself, but then you need to have the connection still in place, okay, the CDO or central governance, they're not helping them anymore. Because they're mature enough that how to keep the line to keep the line has to have a representative from there, it could be a data store could be the domain owner. It really depends. But this person needs to have the line established with the governance just to make sure that they are up to date. So for example, imagine they say, okay, there is a new technology for sharing data progress. We want all domain teams to get up to speed on that. Maybe some domain, whatever, would reject this because you just it's not feasible for them all. It costs them a lot. So you need a representative to be there to convey the mesh to their team and make sure that okay, whatever is defined centrally, is feasible for the different domain.

Koen Hendriks
And when I look at the few presented in the point of view slides, you also have an enabling theme there. Can you maybe describe a bit what the enabling team does?

Data mesh expert 1
Oh, yeah. Yeah, that that was that came, I think, during one of the conversations that an enabling team would be if you remember, we talked about the pool of resources. Data product developers. SME is have developers of people who can intervene very quickly to support a domain getting up to speed and, you know, helping them out. So for example, a domain that is starting new, and they want to get into data ops, DevOps way of working. So you have in the pool, a DevOps expert, again, help them out. If that's not a must to do on every domain first. And then also, this pool is not supposed to fix all the domains issue, because they also need to take responsibility of their stuff. That's, that's interesting, because that is, I can, that is like a typical thing that you need when you implemented in practice. But it's not taken into account in the theoretical operating model, because there they give like the end end state without the statements in the states in between. So that's also really important addition.

Koen Hendriks
Are there any other decisions that are really important? While structuring the organization?

Data mesh expert 1
That's a good question, a point that I mentioned was that a platform needs to have enough. So it's one of the principles of the data measure self service tool, because you don't expect the teams or the domain teams to be mature enough, they might be also domain teams that lasts for maybe a year or two without a data engineer. So you need to support them with tools and people know, with enabling teams as well. So that's a very important one. And the other important thing, it's a rule of thumb, I can say is that a good roadmap would cover all the different levels, of course, you are interested in your thesis more the organization operational part. But it's really important when a company is addressing this to start with an assessment and see okay, where they are lagging behind because what you see if a company focused a lot on the technical part, without focusing on the organization, operational part, then there will be a lot of a lot of chaos, the domains will not be able to understand what is needed to happen, they will not there will be a lot of frustration. And if you focus out organization operation without technology, then also it's difficult for them to see anything tangible. And also they don't have the tools that can support them. So there's like Okay, where can we publish a data product? Where can we do this? So there is no medium to implement those ideas. So it's really important to assess at the beginning understand what is the culture what is the what is the starting point, what is the as a situation and then see okay, we can move in this direction, have we in the right direction have we put that 50% of effort in technology for lagging behind technology and then 50% on organization operational aspect, but this can be answered only after an assessment.

Koen Hendriks
And what exactly is included in that assessment? let's talk about the influences, the parts that influences that assessment. So you talked about culture, having a top down approach So what is it that drives it? And the current organizational structure, but also the technical capabilities? Are there other things that are really important for how it essentially will look like?

Data mesh expert 1
Yeah, I think you can start from the five levels. So infrastructure, technology, architecture, organization and operation, you can also address it from a wider perspective, like the ideal kind of framework.

Koen Hendriks
Anythin else you would like to add?

What you see in this companie, there is a lot of allergy to centralization. Okay, people are allergic to governance and

Central. So they like democratization more. And that was a challenge, because that's why they went so much into technical architectural, and then they lagged behind governance, because it's just against the culture. Although it's a, it's a consensus culture. So you see, making decisions there is very slow, because they always want to have as many people as much as they can, and then make a decision. So there is a lot of delay in making decisions. Telecom 1? It's kind of the same, but in different tests. They like to it's not consensus culture, but they're still also allergic to governance. Even though it's the same effect, the reason that effect is different.

Koen Hendriks

Yeah, the last the last question, which is actually quite difficult well. How do you have any best practices or tips on how to find the equilibrium between the centralization and decentralization of the governance?

Data mesh expert 1
One is, as I mentioned, start with an assessment with a very good understanding of where your organization stand on multiple levels and cultural level, organizational, operational, technical, and so on. And then have a clear idea. What you want to do with the data mesh is beneficial. Some organization, it's better not to do a data mesh, stay centralized. If it's not Like a small company or whatever. So this is very important one, and the second one was starting small, andthere will be some failures, there will be challenges. So it will be mistakes. But you cannot get it right from one time, one go impossible. So this balance between centralized and decentralized, I mean, it can be the case that they can experiment on smaller scale and see how things can go. And then if it doesn't go, well, they can maybe do another iteration with different kinds of setup. What organizations need to understand is, this is a very long term change. And it's important to break this change into smaller steps. And experiment. Let's see how, and keep a feedback loop. This is very important.

Koen Hendriks
you come to the point where you have to learn from your mistakes, but when are you mistakes or failures, enough to say, let's cancel the idea?

Data Mesh Expert 23:31
What is that, that brings us to the role of this CDO, when they define KPIs, and those KPIs are not met? Then you dig deeper into why these KPIs didn't met? Or weren't met? And then it's clear what went wrong. And then can you react? You can react on this? It's not like we're pinpointing on people or they know it's the whole ultimate goal goal is to make this a success and maybe deliver value for the whole organization. Yeah, to reach that point. What do we need to do? Yeah, I think it's impossible to tackle it just on one go. And that's something I faced now I'm talking to the domain team, the first domain team and the next week, they want the data platform team to tell them every single thing like what data models to use, or what data quality roles apply and and everything. The thing is that, this cannot work they need also to take initiative, be more proactive and try and trust that if something goes wrong, it's not the end of the world. You can really do something nice and the goal is so for example, For us now, what we are doing is that we are we want to connect the dots, meaning now we have a marketplace a data catalog. We have quality policies to be implemented centrally. So we need to have a scenario that works end to end. next iteration, we can refine it further. Yes iteration, we can maybe scale this into two or three more domains. next iteration, you see it grows bigger, and every time it matures, more and more, but you cannot just say okay, it's we were aiming for one perfect shot that we not gonna It's impossible. Yeah, that's my I think my, my, my second tip or let's say, I think There is no way to find the balance between centralization and decentralization without assessing first and then trying to manage it.

Koen Hendriks
I can understand. It is five o'clock. So officially, we are out of the Meet the meeting time is over. Thank you very much for your time and see you later.

Appendix G

Treatment validation interviews

G.1 Correctness validation interviews

G.1.1 Correctness validation interview protocol

Validation and Evaluation document	
Goal	Validation and evaluation interview for Thesis on organizational structures in data mesh
Research problem	Enterprise architects design the organizational structure for an organization that want to adopt the data mesh principles. The data mesh theory differentiates significantly from the practice and architects find it difficult to design an organizational structure that fits the data mesh paradigm and the current organization.
Research solution	This thesis tries to give these architects guidance by presenting important design decisions and organization structure. Six different design choies will be presented with their resulting structure (configuration) their (dis)advantages and the factors that influence these decisions. For my research I would like to know the validity of these components.
Link to document:	Guidance tool for design decisions

Validation questions

This section presents each design decision and

Slide 9 : 1.1 Steering committee
1. Do you recognize this design decision?
    a. If not, could you imagine that this is an design decisions that has to be made?
        i. If not, why not?
2. Do you recognize the configurations?
    a. If not, could you imagine that these configurations occur in practice?
        i. If not, why not?
3. Are there **configurations** incorrect or missing?
4. Are there **influential factors** incorrect or missing?
5. Are there **advantages** incorrect or missing?
6. Are there **disadvantages** incorrect or missing?

Slide 10: 1.2 Federated governance council(s)
1. Do you recognize this design decision?
    a. If not, could you imagine that this is an design decisions that has to be made?
        i. If not, why not?
2. Do you recognize the configurations?
    a. If not, could you imagine that these configurations occur in practice?
        i. If not, why not?
3. Are there **configurations** incorrect or missing?
4. Are there **influential factors** incorrect or missing?
5. Are there **advantages** incorrect or missing?
6. Are there **disadvantages** incorrect or missing?

Slide119: 2.1 Data domain ownership

1. Do you recognize this design decision?
    a. If not, could you imagine that this is an design decisions that has to be made?
        i. If not, why not?

2. Do you recognize the configurations?
    a. If not, could you imagine that these configurations occur in practice?
        i. If not, why not?
3. Are there **configurations** incorrect or missing?
4. Are there **influential factors** incorrect or missing?
5. Are there **advantages** incorrect or missing?
6. Are there **disadvantages** incorrect or missing?

Slide 12: 2.1.2 Domain representative structure
1. Do you recognize this design decision?

    a. If not, could you imagine that this is an design decisions that has to be made?

        i. If not, why not?

2. Do you recognize the configurations?

    a. If not, could you imagine that these configurations occur in practice?

        i. If not, why not?
3. Are there **configurations** incorrect or missing?
4. Are there **influential factors** incorrect or missing?
5. Are there **advantages** incorrect or missing?
6. Are there **disadvantages** incorrect or missing?

Slide 13: 2.2 Data product developer capability allocation

1. Do you recognize this design decision?
    a. If not, could you imagine that this is an design decisions that has to be made?
        i. If not, why not?
2. Do you recognize the configurations?
    a. If not, could you imagine that these configurations occur in practice?
        i. If not, why not?
3. Are there **configurations** incorrect or missing?
4. Are there **influential factors** incorrect or missing?
5. Are there **advantages** incorrect or missing?
6. Are there **disadvantages** incorrect or missing?

Slide 14: Enabling team
1. Do you recognize this design decision?
    a. If not, could you imagine that this is an design decisions that has to be made?
        i. If not, why not?
2. Do you recognize the configurations?
    a. If not, could you imagine that these configurations occur in practice?
        i. If not, why not?
3. Are there **configurations** incorrect or missing?
4. Are there **influential factors** incorrect or missing?
5. Are there **advantages** incorrect or missing?
6. Are there **disadvantages** incorrect or missing?

G.1.2 Correctness validation interview responses

TABLE G.1: Interview responses for validation study

1.1: Steering board	
1. Do you recognize this design decision from your experience?	Evaluation 1: yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: Yes Evaluation 5: yes Evaluation 6: yes
Additional comments:	
2.1 Do you recognize Configuration 1.1.1 from experience or could you imagine that this configuration occurs in practice?	Evaluation 1: No Evaluation 2: yes Evaluation 3: Yes Evaluation 4: Yes Evaluation 5: yes Evaluation 6: yes
Additional comments:	Evaluation 1: From my point of view, this configuration cannot have the power to drive the data governance initiatives to the rest of the organization
2.2 Do you recognize Configuration 1.1.2 from experience or could you imagine that this configuration occurs in practice?	Evaluation 1: Yes Evaluation 2: yes Evaluation 3: Yes Evaluation 4: Yes Evaluation 5: yes Evaluation 6: yes
Additional comments:	Evaluation 5: You can show that the decision making steering committee is interdiciplinary through the vizualisation
3. Are there configurations missing?	Evaluation 1: No Evaluation 2: no Evaluation 3: yes Evaluation 4: no Evaluation 5: no Evaluation 6: yes
Additional comments:	Evaluation 3: one other option is a steering board within the domain Evaluation 6: But it is not possible to be exhaustive
4. Do you agree with the influential factors?	
Organizational culture	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: Yes Evaluation 5: Yes Evaluation 6: Yes
Organizaitonal Strategy	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: Yes Evaluation 5: Yes Evaluation 6: Yes

Organization size	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: Yes Evaluation 5: Yes Evaluation 6: Yes
Organization complexity	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: Yes Evaluation 5: Yes Evaluation 6: Yes
C-level management	Evaluation 1: Yes Evaluation 2: other Evaluation 3: Yes Evaluation 4: Yes Evaluation 5: Other Evaluation 6: Yes
Additional comments:	Evaluation 2: Organization agility Evaluation 3: Data governance maturity Evaluation 4: Culture is most important, data mesh maturity Evaluation 5: organizational and leadership culture is most important Evaluation 6: regulatory requirements can play a role

5. Do you agree with the advantages

Configuration 1.1.1:

Fast decision making	Evaluation 1: Yes Evaluation 2: other Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes Evaluation 6: yes
Low human resource costs	Evaluation 1: No Evaluation 2: no Evaluation 3: Yes Evaluation 4: yes Evaluation 5: Other Evaluation 6: Other
Additional comments:	Evaluation 1: A steering comittee does not meet frequently, so the lower human resource costs is not significant Evaluation 2: Fast decision making is not always true because more effort could be needed for allignment which reduces the decision making speed. A benefitcould be shorter comunication to make decisions Evaluation 5: The steering committee does only meet once very few monts, so the lower human resource costs is not necessary true. It could take more human resources because it needs to be alligned with the strategy

Configuration 1.1.2:

Clear accountabilities	Evaluation 1: Yes Evaluation 2: other Evaluation 3: Other Evaluation 4: yes Evaluation 5: No Evaluation 6: No
Strong strategic allignment	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes Evaluation 6: yes
Additional comments:	Evaluation 4: Leadership involvement Evaluation 5: Leadership buy in / leadership commitment / more power & more effective decision making / Both configurations can be given clear accountabilities Evaluation 6: Partially, but it is not possible to be exhaustive.

6. Do you agree with the disadvantages

Configuration 1.1.1:

Unclear accountabilities	Evaluation 1: Yes Evaluation 2: Other Evaluation 3: Other Evaluation 4: yes Evaluation 5: no Evaluation 6: Other
Lack of hollistic view on decisions	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes Evaluation 6: yes
Lack of leaderhip	Evaluation 1: yes Evaluation 2: yes Evaluation 3: yes Evaluation 4: yes Evaluation 5: yes Evaluation 6: yes
Lack of strategic allignment	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes Evaluation 6: yes
Additional comments:	Evaluation 1: Lack of power towards the rest of the organization, risk of working on wrong topic/ extra need for resources Evaluation 2: Who is accoutnable for the decisions and are these decisions made based on majority or is someone specific responsible? Evaluation 3: who is accountable? Collaborative or single Evaluation 6: Partially, but it is not possible to be exhaustive.

Configuration 1.1.2:

Slow decision making	Evaluation 1: Yes Evaluation 2: other Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes Evaluation 6: yes
High human resource costs	Evaluation 1: No Evaluation 2: No Evaluation 3: No Evaluation 4: yes Evaluation 5: other Evaluation 6: other
Additional comments:	Evaluation 3: HR speed –>tabling decisions for steering committee expected no extra resource costs , Slow decision making totally a free especially with data mesh / mitigating by changing the steering committee structure Evaluation 5: Additional decisino making aproval needed, The team does only need to gather around once a month, so a disagreement with the high human resource costs

1.2: Federated governance councils

1. Do you recognize this design decision from your experience?	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: Yes Evaluation 5: yes Evaluation 6: yes
Additional comments:	
a. Could you imagine that this is an design decisions that has to be made?	Evaluation 1: Yes Evaluation 2: yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes Evaluation 6: yes
Additional comments:	Evaluation 5: give an explanation of different combinations of data governance councils
2.1 Do you recognize Configuration 1.2.1 from experience or could you imagine that this configuration occurs in practice?	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: Yes Evaluation 5: yes Evaluation 6: yes
Additional comments:	Evaluation 1: applicable if data mesh maturity is low
2.2 Do you recognize Configuration 1.2.2 from experience or could you imagine that this configuration occurs in practice?	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: Yes Evaluation 5: yes Evaluation 6: yes
Additional comments:	

3. Are there configurations missing?	Evaluation 1: Other Evaluation 2: no Evaluation 3: No Evaluation 4: no Evaluation 5: no Evaluation 6: no
Additional comments:	Evaluation 1: If there are multiple federated governance teams, there could be a governance council in between Evaluation 6: It needs to show in more detail what the different types of governance councils are.
4. Do you agree with the influential factors?	
Data mesh maturity	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Other Evaluation 4: yes Evaluation 5: Other Evaluation 6: yes
Organizational culture	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes Evaluation 6: yes
Organization size	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes Evaluation 6: yes
Organization complexity	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes Evaluation 6: yes
C-level management	Evaluation 1: Yes Evaluation 2: No Evaluation 3: Yes Evaluation 4: yes Evaluation 5: No Evaluation 6: yes
Additional comments:	Evaluation 2: Organizational structure Evaluation 3: different way around, a more mature data mesh is the steering committee less a decision maker, Data strategy Evaluation 5: Data mesh maturity is too broad to use as influential factor Evaluation 6: Ragulatory requirements
5. Do you agree with the advantages	
Configuration 1.2.1:	

Central allignment within a single team	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: Other Evaluation 6: yes
Additional comments:	Evaluation 2: everybody is informed and involved but there is a need for prioritization of resource for advice Evaluation 5: Can be better formulated

Configuration 1.2.2:

Focused decision making	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes Evaluation 6: yes
In-depth discussion on specific topics	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes Evaluation 6: yes
Additional comments:	Evaluation 3: Efficient contribution to the councils people that are needed are there, the rest not , no waste of time Evaluation 6: Overlaps between councils in multiple governance councils is not addressed. This will be a major issue.

6. Do you agree with the disadvantages

Configuration 1.2.1:

Need for facilitators	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: other Evaluation 6: yes
Relativly large team	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: other Evaluation 6: yes
Need for prioritization of decision making	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Other Evaluation 4: yes Evaluation 5: other Evaluation 6: yes
Additional comments:	Evaluation 3: competing priorities Evaluation 5: The disadvantages should be more high over difficult to oversee all the teams due to the size, agenda and expertise needed in the council

Configuration 1.2.2:

Risk of misallignment	Evaluation 1: Yes Evaluation 2: yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes Evaluation 6: yes
Slow decision making	Evaluation 1: Yes Evaluation 2: yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes Evaluation 6: yes
	Additional comments:

2.1 Data domain ownership	
1. Do you recognize this design decision from your experience?	Evaluation 1: Yes Evaluation 2: yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Additional comments:	
a. Could you imagine that this is an design decisions that has to be made?	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: Yes
Additional comments:	
2.1 Do you recognize Configuration 2.1.1 from experience or could you imagine that this configuration occurs in practice?	Evaluation 1: Yes Evaluation 2: yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Additional comments:	Evaluation 1: importance of informal communities of data product owners
2.2 Do you recognize Configuration 2.1.2 from experience or could you imagine that this configuration occurs in practice?	Evaluation 1: Yes Evaluation 2: yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Additional comments:	Evaluation 2: Both configs should have a single point of contact
3. Are there configurations missing?	Evaluation 1: No Evaluation 2: no Evaluation 3: No Evaluation 4: no
Additional comments:	
4. Do you agree with the influential factors?	
Domain complexity	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes

Organizational culture	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Human resource availability	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Data mesh maturity	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Additional comments:	Evaluation 1: Data mesh maturity, data domain definition Evaluation 2: Domain size and people size Evaluation 4: compleity is very important!!!

5. Do you agree with the advantages

Configuration 2.1.1:

Fast decision making	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: No
No need for person with a hollistic view	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: Other
Additional comments:	Evaluation 5: "No need for person with a hollistic view" can be phrased better

Configuration 2.1.2:

Clear accountabilities	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: Yes
Additional comments:	Evaluation 1: Bridge the gap between data product owners and the council, motivation , leadership Evaluation 2: Single point of contact for the domain Evaluation 5: difficult to find dedicated business roles, potential risk , Business allignment & Faster decision making

6. Do you agree with the disadvantages

Configuration 2.1.1:

Unclear accountabilities	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes

Lack of a decisoin maker	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Additional comments:	Evaluation 4: Lack of business allignment
Configuration 2.1.2:	
Need for person with domain owner capabilities	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Other Evaluation 4: yes Evaluation 5: yes
dependent on strong domain definition	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: No Evaluation 4: yes Evaluation 5: yes
Additional comments:	Evaluation 2: Increases complexity of roles Evaluation 3: Effort to motivate people to take domain responsibility Evaluation 5: Difficult to find people for the domain owner role, potential decision from domain owner / lack of motivation

2.1.2 Domain representatives' structure	
1. Do you recognize this design decision from your experience?	Evaluation 1: Other Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Additional comments:	
a. Could you imagine that this is an design decisions that has to be made?	
Additional comments:	
2.1 Do you recognize Configuration 2.1.2.1 from experience or could you imagine that this configuration occurs in practice?	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Additional comments:	Evaluation 3: to limit the people that
2.2 Do you recognize Configuration 2.1.2.2 from experience or could you imagine that this configuration occurs in practice?	Evaluation 1: No Evaluation 2: Yes Evaluation 3: yes Evaluation 4: yes Evaluation 5: yes
Additional comments:	Evaluation 1: you don't need a domain owner in this configuration (the domain manager is the bridge between the data governance council and the data product owners) / ondemand for specific topics now its is more understandable Evaluation 3: super organized and super small

3. Are there configurations missing?	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: yes Evaluation 4: yes Evaluation 5: yes
Additional comments:	Evaluation 1: there should be a configuration that present that only the data product owners are in the council for organizations that do not have data domain owners Evaluation 2: Another configuration could be one in which data consumers are present in the council Evaluation 3: The data owner role could also be in the council Evaluation 4: one without the data domain owner in the council and only data product owners Evaluation 5: bring an data product owner as SME in the domain
4. Do you agree with the influential factors?	
Organizational culture	Evaluation 1: no Evaluation 2: yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Organizaitonal Strategy	Evaluation 1: no Evaluation 2: yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: no
Organization size	Evaluation 1: no Evaluation 2: yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Organization complexity	Evaluation 1: no Evaluation 2: yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
C-level management	Evaluation 1: no Evaluation 2: no Evaluation 3: Yes Evaluation 4: yes Evaluation 5: no
Additional comments:	Evaluation 2: size en complexity belangrijk, C -level minder belangrijk Evaluation 4: organization size Evaluation 5: Less is more, try to look at the influential factors with the pros and cons
5. Do you agree with the advantages	
Configuration 2.1.2.1:	
Fast decision making	Evaluation 1: Other Evaluation 2: yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes

Additional comments:

Configuration 2.1.2.2:

Centralization of communication	Evaluation 1: Other Evaluation 2: yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: Yes
Data product specific problem discussions	Evaluation 1: Other Evaluation 2: yes Evaluation 3: Other Evaluation 4: No Evaluation 5: other
Additional comments:	Evaluation 4: overkill of people maybe only at the very very beginning Evaluation 5: potential derailing of committee slow decision making

6. Do you agree with the disadvantages

Configuration 2.1.2.1:

Effort required for good communication	Evaluation 1: yes Evaluation 2: yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes

Additional comments:

Configuration 2.1.2.2:

Slow decision making	Evaluation 1: Other Evaluation 2: yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Large federated data governance council	Evaluation 1: Other Evaluation 2: yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: Other
Additional comments:	Evaluation 2: not needed Evaluation 3: A lot of meeting prep Evaluation 5: Potential derailing of the council meeting because of data product specific details, slower decision making

2.2 Data product developer capability allocation	
1. Do you recognize this design decision from your experience?	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: Yes Evaluation 5: yes
Additional comments:	

a. Could you imagine that this is an design decisions that has to be made?	Evaluation 1: Yes Evaluation 2: yes Evaluation 3: Yes Evaluation 4: Yes
Additional comments:	
2.1 Do you recognize Configuration 2.2.1 from experience or could you imagine that this configuration occurs in practice?	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Additional comments:	
2.2 Do you recognize Configuration 2.2.2 from experience or could you imagine that this configuration occurs in practice?	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Additional comments:	
2.3 Do you recognize Configuration 2.2.3 from experience or could you imagine that this configuration occurs in practice?	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Additional comments:	Evaluation 5: domain specific knowledge is still important, a bit of knowledge is needed
3. Are there configurations missing?	Evaluation 1: No Evaluation 2: Other Evaluation 3: No Evaluation 4: no Evaluation 5: yes
Additional comments:	Evaluation 2: A hybrid model with both domain specific data product developers and a data domain developer pool
4. Do you agree with the influential factors?	
Domain complexity	Evaluation 1: Yes Evaluation 2: yes Evaluation 3: yes Evaluation 4: yes
Human resource availability	Evaluation 1: Yes Evaluation 2: yes Evaluation 3: yes Evaluation 4: yes
Capability availability	Evaluation 1: Yes Evaluation 2: yes Evaluation 3: yes Evaluation 4: yes

Additional comments:	Evaluation 1: Business IT allignment maturity Evaluation 2: Hoe meer links hoe meer data mesh Evaluation 4: Small organization size means pool, self serve data platform maturity/ Organization size, domain complexity is less important, HR availabilty in domains and organization Evaluation 5: het erbuiten laten van domain specific knowledge / maturity of the data strategy, If everybody knows the data strategy, this is what we are going to do and these are the priortities, then you can have a central pool and the decentralization. If this is not the case and you distribute the data product developers, then it will be challenging. And Business and IT allignment should be Business Data IT allignment.

5. Do you agree with the advantages

Configuration 2.2.1:

High domain specific knowledge	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
High efficiency	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Additional comments:	

Configuration 2.2.2:

Moderate domain specific knowledge	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
High feasibility	
	Evaluation 5: knowledge sharing
Additional comments:	Evaluation 1: Moderate efficiency Evaluation 2: High flexibility Evaluation 5: feasibility weglaten

Configuration 2.2.3:

Low need for human resources	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: no

High flexibility	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Additional comments:	Evaluation 5: Standardization in way of working levariging BP's

6. Do you agree with the disadvantages

Configuration 2.2.1:

High need for human resources	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Low flexibility	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Additional comments:	Evaluation 5: more specialised data product developing resources neede/ lack of standardization

Configuration 2.2.2:

Effort needed for prioritization	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Additional comments:	Evaluation 5: difficult prioritization and decision making

Configuration 2.2.3:

Need for prioritization	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Not scalable	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Other Evaluation 4: yes Evaluation 5: Other
Low domain knowledge	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: yes Evaluation 4: yes Evaluation 5: yes
Additional comments:	Evaluation 1: Low efficiency Evaluation 3: Monitoring the pool, resourece bottleneck, low efficiency Evaluation 5: not efficient allocation of resources

Table G.1 Continued:

3.1: Presence of Enabling team	
1. Do you recognize this design decision from your experience?	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Additional comments:	
a. Could you imagine that this is an design decisions that has to be made?	Evaluation 1: Yes Evaluation 2: yes Evaluation 3: Yes Evaluation 4: Yes
Additional comments:	Evaluation 1: This team is likely to be there, but the size of this team can depend on the organization
2. Do you recognize the configurations from experience or could you imagine the configurations?	
2.1 Do you recognize Configuration 3.1 from experience or could you imagine that this configuration occurs in practice?	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Additional comments:	Evaluation 1: Not a smart decision Evaluation 2: In config 3.1, one person can be responsible within the team
2.2 Do you recognize Configuration 3.2 from experience or could you imagine that this configuration occurs in practice?	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Additional comments:	Evaluation 1: Extremely important!
3. Are there configurations missing?	Evaluation 1: Other Evaluation 2: Other Evaluation 3: No Evaluation 4: no
Additional comments:	Evaluation 1: different sizes of enabling team
4. Do you agree with the influential factors?	
Human resource availability	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Organizational culture	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Organizational strategy	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: other

Additional comments:	Evaluation 1: Financial resources Evaluation 2: data mesh maturity in the team Evaluation 3: overall data literacy Evaluation 4: Maturity of the data mesh<−, maturity of the domains

5. Do you agree with the advantages

Configuration 1.1.1:

No need for the creation of a communication or enabling plan	Evaluation 1: Yes Evaluation 2: yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: other
No need for extra human resources	Evaluation 1: Yes Evaluation 2: yes Evaluation 3: Yes Evaluation 4: no Evaluation 5: Other
Additional comments:	Evaluation 1: Less costs Evaluation 2: more flexibilty Evaluation 5: faster decision making , enabling can be tailored towards the specific team

Configuration 1.1.2:

Clear responsibilities regarding capability development	Evaluation 1: Yes Evaluation 2: yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: No
Extra guidance in capability development	Evaluation 1: Yes Evaluation 2: yes Evaluation 3: Other Evaluation 4: yes Evaluation 5: yes
Additional comments:	Evaluation 3: drive guidance in capability development Evaluation 5: standardization, knowledge sharing, sharing best practices. Inside the domians people need to be motivated

6. Do you agree with the disadvantages

Configuration 1.1.1:

Less development of capabilities	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Other Evaluation 4: yes Evaluation 5: no
lack of change management plan	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes

Additional comments:	Evaluation 1: Lack of allignment in enabling initiatives, lack of motivation, weak change management Evaluation 3: difference in capabilities between domains and slow development
Configuration 1.1.2:	
Need for extra human resources	Evaluation 1: Yes Evaluation 2: No Evaluation 3: Yes Evaluation 4: yes Evaluation 5: yes
Slow development of capabilities	Evaluation 1: Yes Evaluation 2: Yes Evaluation 3: No Evaluation 4: yes Evaluation 5: No

G.2 Utility, understandability, efficiency and effectiveness validation interviews

TABLE G.2: The results of the expert interviews on understandability, utility, efficiency and effectiveness

Understandability		Min	Max	Avg
1	The hierarchy of design decisions is understandable	5	7	5,75
Comment	Evaluation 2: would be good to include an overview of 'governance components' and start with that (piramyd principle) Evaluation 8: Create different colors for different layers, 2) add to title: Conceptual model for the design options for data mesh organizational structure.			
2	The explanations of the configurations are understandable	4	7	5,31
Comment	Evaluation 1: Add more information about definitions & assumptions / really valuable to understand the asset 4–>6 Evaluation 4: bullet point text is better, less text more highlights Evaluation 7: the hierarchy of the federated governance teams is not clear, is it possible to have multiple governance councils without having a steering committee? Evaluation 8: scenarios instead of configuration, making it more consistent, steering committee unclear			
3	The differences between the configurations are understandable	5	7	6,00
Comment	Evaluation 8: making bold and visually represent the differecnce, allign the federated governance team in the visualization			
4	The (dis)advantages of the design options are understandable	5	7	5,88

Table G.2 Continued:

		Min	Max	Avg
Comment	Evaluation 8: Use brackets in the tool to represent potential affects/disadvantages for example: Low(er) decision making speed			
5	The influential factors are understandable	3	7	5,13
Comment	Evaluation 3: Refer to the feedback we discussed during the walk-through session we had. Evaluation 5: as discussed, try to specify them better, make them less generic Evaluation 7: add a definition of the influential factors is unclear Evaluation 8: add a definition of the influential factors			
6	The graphical representations of the organization configurations are understandable	5	7	6,38
Comment	Evaluation 3: Love the graphics!			
Additional Comment	Evaluation 7: be aware to remove dependencies of configurations if they are not dependent. Or give extra information through a footnote Evaluation 8: Data mesh transformation (organizatinal) instead of migration (technical) because you talk about the organizational restructuring, More consistent in federated teams and councils, add a definition of different teams and roles			

Utility		Min	Max	Avg
7	This guidance tool is easy to use during the design process Evaluation 3: It is quite easy to use but I think some of the configuration explanations could be a little shorter and sharper to enhance the user experience. Evaluation 7: the clicking through makes it really easy to use Evaluation 8: theoretical and practical / contents. It is really intuitive, nice that you can click through from decision to decision and it is possible to directly navigate to the design decision that you want to talk about with your client or internally with your team to discuss the different scenarios and formulate a recommendation upon this improves the usability. The theoretical vs practical slide could be more elaborated. Add one pager on how to use this tool, extend the current page. Make the disclaimer that it is to support to make the dicision instead of making the decision	5	7	5,88
8	This guidance tool is useful during the design process	6	7	6,50
9	This guidance tool is useful to distinguish different design decisions	5,5	7	6,13
10	This guidance tool gives design configurations that could help during the design process	5	7	6,13

11	The (dis) advantages are useful to support decision making in the design process Evaluation 8: The disadvantages are really organization dependent and there should be clearly mentioned why the advantage or disadvantage occurred	4	7	5,63
12	The influential factors are useful to support decision making in the design process Evaluation 2: The usability of the influential factors is very dependent of the organization, however they give a starting point for the discussion Evaluation 5: influential factors are too generig.Specifying thm better would increase the usability	4	7	6,00

Efficiency		**Min**	**Max**	**Avg**
13	Using this tool would reduce discussion time about the organization structure Evaluation 3: I think this could be a really helpful tool to help facilitate discussions during workshops. Evaluation 6: Using this tool would probably reduce the design team, but discussion time would still be the same as this tool could only be a framework for discussion. The needs, influential factors and maturity levels of each organization are unique and due discussion will be required even with the given design options. Evaluation 8: It would definitly contribute to the efficiency of the organization design of a data mesh	4	7	5,88
14	Using this tool would reduce the design time of the organization structure Evaluation 3: 100% - your tool sets out the key parameters and key decisions required to be taken into account. Evaluation 7: It gives you a structure of the decisions and helps you with making decisions faster	5	7	6,00

Efficiency		**Min**	**Max**	**Avg**
15	Do you think that using this tool affects the effectivity/success of the organization structure design process? Evaluation 3: I do – I think this creates important parametres for this discussion and provides a good prompt (via the configuration options) to guide discussions on the decisions needed to build a governance model for data mesh Evaluation 4: depends on take action Evaluation 8: it adds to the completeness	5	7	5,63
16	If so, would it increase or decrease?			

Evaluation 7: I don't forsee any problems that could occur and result in worse decisions when you use this tool Evaluation 8: This is not a comprehensivy set of design decisions, therefore it is important to use these decisions as a starting point but not as a comprehensive checklist for the full organizaiton design. Financial components on organizational structures are also very important to make the process more effective. Funding of the roles and the teams are also decisions that are really imporant to make. 100% complete is not possible

Any other remarks?

Evaluation 1: The framework can be used with the external client, what are the options that we have, what are important decisions and what are the advantages and disadvantages, easier to define the role and easier who to pick out the person that fits this role. This framework can really help with designing the structure of the teams and meetings. Without having such a tool at hand,there is a lot of brainsormein and comparing. with this tool it is easy, we can open it and discuss it which accalerates the process because many thoughts are already included in the tool. This tool could absolutely help with the architects. This tool support to define the structure,compare the configs. The tool helps with defining the structure. It does not support identying the right people. Set the direction and the project that it is targeting.

G.3   Refinement table

TABLE G.3: Table with the refinements, their sources, the requirements that they contribute to and the affected component

	Refinement	Improve correctness according to support of evaluation interviews	Removal of visual dependencies between design decisions	Extend the theoretical vs practical slide	Extension of the tool guide	Addition of information about advantages and disadvantages	Addition of color gradient in advantages and disadvantages	Increase consistency in names of teams	Introduction of bullet points for the configuration explanations	Addition of cover slide	Addition of role definitions	Addition of influential factors definition	Addition of color differences in option table	Addition of new navigation buttons for new slides
Source	EI1	x					x			x				
	EI2	x								x				
	EI3	x												
	EI4	x							x					
	EI5	x						x				x		
	EI6	x												
	EI7	x			x					x	x			
	EI8	x	x	x	x	x	x	x		x	x	x	x	
Requirement	FR1													
	FR2													
	FR3													
	FR4													
	FR5													
	NRC1	x												
	NRC2	x												
	NRC3	x												
	NRC4	x												
	NRUN1								x		x	x		
	NRUN2												x	
	NRUN3		x						x		x			
	NRUN4													
	NRUN5					x	x							
	NRUN6											x		
	NRUN7													
	NRUT1				x	x	x		x	x				x
	NRUT2				x	x			x	x				x
	NRUT3													
	NRUT4												x	
	NRUT5					x	x							
	NRUT6													
	NRUT7					x	x							
	NREI1					x	x							
	NREI2													
	NREI3								x					
	NREI4					x	x							

Table G.3 Continued:

Component	NREE1												
Addition									x	x	x		
1.1				x									
1.2				x									
1.3				x									
1.4				x									
2.1													
2.2													x
2.3			x										
2.4													
3.1													
3.2													x
3.3												x	
3.4													
X.1	x												
X.2													x
X.3	x						x	x					
X.4	x	x					x						
X.5	x				x	x	x						
X.6	x												
6.7													