A Meshed Up Data Architecture Design

Thesis Report

MSc Complex Systems Engineering & Management

W.M. de Boer

Delft University of Technology

Deloitte.

A Meshed Up Data Architecture Design

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Complex Systems Engineering and Management**

Faculty of Technology, Policy and Management

by

Willemijn de Boer

Student number: 4447190

To be defended in public on July 7th 2022

Graduation committee

Chairperson	: Prof.dr.ir. M.F.W.H.A. Janssen, Section ICT
First Supervisor	: Prof.dr.ir. M.F.W.H.A. Janssen, Section ICT
Second Supervisor	: Dr. ir. B. Enserink, Section Policy Analysis
External Supervisor	: E. Bookholt, Deloitte
External Supervisor	: R. Van de Beek, Deloitte

PREFACE

Dear reader,

I feel proud and fulfilled when presenting to you my thesis on The Assessment of Organization Readiness for guiding Data Mesh Migrations. This report concludes my MSc Complex Systems Engineering and Management at the Faculty of Technology, Policy and Management at Delft University of Technology.

During the past 5 months I have immersed myself in one of the newest concepts in the field of data architectures: the Data Mesh. It was a subject that quickly captured my dedication and enthusiasm, because of its new, innovative and above all socio-technical complex character. The exploratory nature of this study, as well as the overarching need for 'best practices' from the application environment, ensured that I had the opportunity to talk and spar with many different experts from all around the world about this intriguing topic. People who know me well know that this is something that complements my abilities and gives me a lot of joy. Considering this, there had been no graduation project that would have suited me better than this one. I would therefore like to express my greatest thanks to Deloitte Consulting and in particular Erik Bookholt and Rik van de Beek for making this research possible. You gave me the idea and inspired me to bite into this research, while you were and remained enthusiastic about the progress we made together. In addition, the TV&A team at Deloitte welcomed me as family, and showed me in the short time available what a great team they are.
Additionally, I would like to express my gratitude to Marijn Janssen and Bert Enserink for their academic guidance throughout the process, and the ability to put me on the scientifically correct path every now and then when I kept thinking too much in practical solutions.

The moment has come for me to officially put an end to my student life, something that is close to my heart, but which I also look back on with warm feelings. In particular, these feelings go out to my dearest friends, my boyfriend, my roommates, and all the other people who made my time in Delft remarkable. Not to forget my parents and my sweet sister, who have always encouraged me to give the best of my ability and who gave me every opportunity to develop myself into the person I am today and the person I've aspired to be.

Dear reader, thank you for reading this report. I hope it provides you with the insights and perspectives you are looking for, and with the answers to the questions you now have. And don't worry, if you don't have the time to read more than one hundred pages - something that I can well imagine -, there is a scientific article at the end of this report that summarizes the content in just over nine pages. That gives you even more time to start the exciting process of shaping your prospective Data Mesh journey...

Let's see what the future brings us!

Kind regards,

Willemijn de Boer
Rotterdam, June 2022

EXECUTIVE SUMMARY

Due to the increasing amount of data processed within organizations, these organizations are increasingly experiencing shortcomings in their current monolithic data architectures and their capacity to facilitate these data operations. The particular limitations cause an increasing demand for a data architecture design that can address these limitations. A distributed and domain-oriented Data Mesh architecture promises to address the limitations of current monolithic architectures in terms of scalability, improved data accessibility, shortened lead times between operations and analytics and improved appointments of ownership and responsibility for data. These improvements are promised, based on the design of a Data Mesh architecture that is built upon the following four principles: (1) domain-oriented decentralized data ownership, (2) data as a product, (3) self-serve data platform, and (4) federated computational governance. This research focuses on the assessment of organization readiness for the migration to a Data Mesh architecture.

Considering the novelty of the Data Mesh concept, there is a lack of understanding of the required level of preparatory efforts to build readiness for a Data Mesh migration. Seeing that the process of migrating is not as simple as implementing a new tool or feature, it needs to be thought through carefully before it can be carried out (Furia, 2021; Schultze, 2020). Moreover, literature currently does not provide a suitable Data Mesh readiness framework, making it difficult for organizations to make a well-considered choice on when and how to transition (I. A. Machado, Costa, & Santos, 2022a). This research addresses this gap in knowledge by designing and developing a Data Mesh Readiness Model that assesses organizational factors and capabilities that are required within organizations to be able to start their Data Mesh migration. Moreover, the model can provide insights for these organizations about their current position with regard to Data Mesh migration, as well as outline potential areas of improvements. The main research question of this study is therefore:

"How can organizations assess their readiness for migrating to a Data Mesh architecture?"

When determining readiness for a Data Mesh migration, the Data Mesh Readiness Model (DMRM) as shown in figure 0.1 offers guidance in assessing the as-is situation of organizations. The model aims to mobilize decision-makers to start their migration by providing insights to what improvement areas to focus on. The DMRM was designed on the basis of theoretical and empirical research towards the identification of factors influencing Data Mesh readiness. These factors are reflected in a novel two-dimensional readiness model, which addresses both the organizational steps towards Data Mesh readiness and the technological dimensions of a Data Mesh architecture design. The model was applied within seven demonstration environments, on which a cross-case analysis was performed in order to be able to compare the performance of the model in practice between the cases. A visual representation of the Data Mesh Readiness Model is shown in Figure 0.1.

The case-study analysis concerning **Assessing the reason to change** shows that the demands of an organization that point to a Data Mesh architecture often come from different angles within the organization concerned. In practice, it is often a combination of the Data Mesh *needs* as identified in the above-captioned model. Generically, the demands stem from the organization's ambition to become more data-driven in the future.

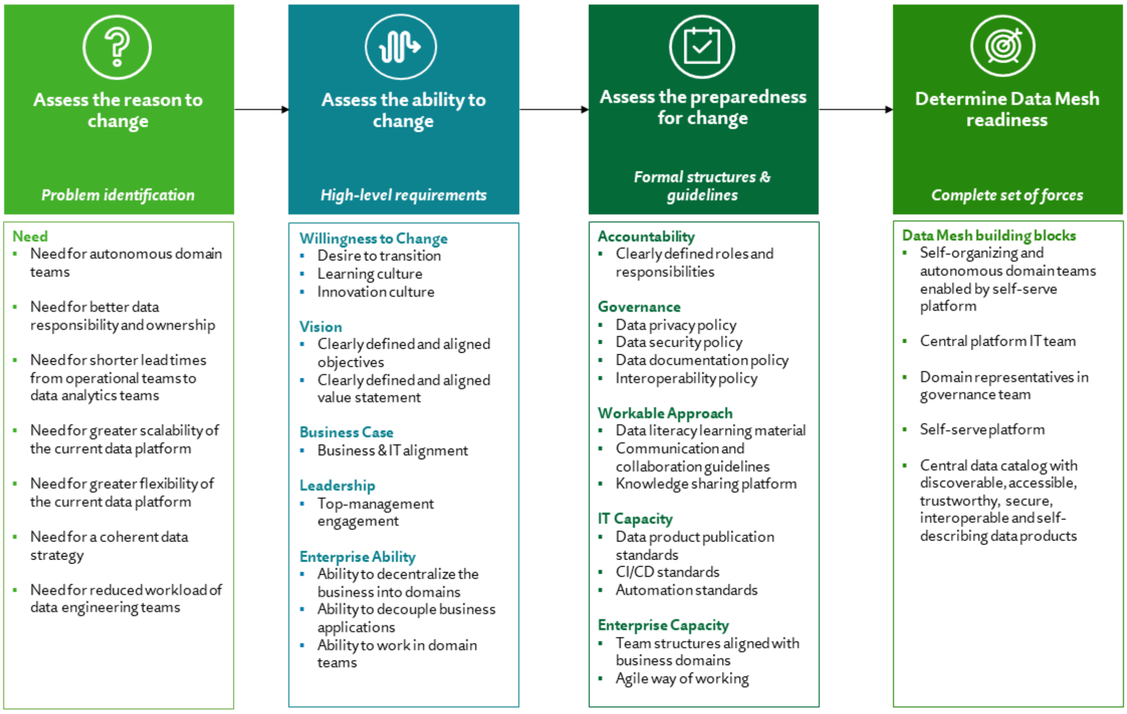

Figure 0.1: Data Mesh Readiness Model (DMRM)

In terms of **Assessing the ability to change** to a Data Mesh architecture, the case-study analysis shows that the organization needs to meet a number of requirements. One of the most important requirements is a detailed *business & IT alignment*. Specifically for a Data Mesh migration, the departments within an organization should recognize the value of each other's operations. In particular, they should recognize the value of the data that the departments use for their operations. The above largely requires the establishment of organisation-wide data awareness. The alignment mainly consists of shared awareness of the value and use of data for operations. In addition, among the factors *decentralization of the current monolithic platforms* and the *decoupling of the existing business applications*, many different perspectives exist among the participants of this research. Examples of these perspectives are to start with domains that are eager to be autonomous, and to separate the business application layer from the Data Mesh migration.

When **Assessing the preparedness for change**, the DMRM enables organizations to review whether the required formal structures and guidelines supporting a Data Mesh migration are in place. The case-study analysis discovered that this stage leans more towards an execution framework than an organizational readiness assessment. However, since this component contributes to the aim of the model, namely guiding organizations through the preparatory phases prior to Data Mesh implementation, it was decided to maintain it in the model.

Moreover, the analysis showed that cross-domain collaboration should be facilitated since it adds to the performance of the Data Mesh. Additionally, it has been discovered that the design of a governance model for a Data Mesh architecture lacks consensus among the participants. Therefore, it has been proposed in this study to conduct further research in this area. Furthermore, the analysis shows that the inter-domain dependencies in the organization should be well thought out before starting the Data Mesh migration. Finally, the case-studies show that the *alignment of the team structures with the business domains* supports the preservation of domain expertise in the teams.

When **Determining Data Mesh readiness**, a meaningful insight is that a Data Mesh migration should always be handled in an iterative way throughout the entire process. It appears difficult to get everything right in one go, which means that factors earlier readiness phases often have to be adjusted afterwards. Also, since a Data Mesh migration has great organizational impact on all departments, it should be approached in a structured and organized manner. A model as proposed in this study can provide guidance during this process.

Facilitating the testing and evaluation phase of this research, an operationalized artifact of the DMRM has been developed in the shape of a self-assessment questionnaire. Both the testcases and the evaluation sessions show that this self-assessment contributes to the usability of the DMRM, due to the factor explanations and actionable follow-up questions. The insights of these sessions include several future improvements of the DMRM and the self-assessment instrument, such as the expansion of the model to a broader business context and the design of an automated assessment that immediate actionable advice based on the assessor's answers.

This research extends the body of knowledge on the academic literature on Data Mesh architecture assessment methodologies by providing a two-dimensional readiness model that can be utilized to assess both the organizational capabilities as well as the technological implementations required for a Data Mesh migration. Moreover, this assessment can also be used as a self-assessment, offering new insights for customers of the assessment and guiding them in an iterative way through the process of preparing them for a Data Mesh migration. Reflecting on both the findings as well as the DMRM, it has been revealed that not all organizations are suitable for a Data Mesh architecture, since the architecture design mainly offers a solution for large and complex organizations that are digital native. Moreover, it seemed that it is not always necessary for an organization to implement all the Data Mesh principles as a whole to be able to retrieve value out of it. Organizations that remain critical towards the new architecture design shall decide for themselves which components address their specific challenges, based on the results and insights from the DMRM.

Due to the exploratory and novel character of this research, further research is strongly recommended. Firstly, a follow-up study could focus on broadening the demonstration environments. Secondly, it could focus on the design and development of a Data Mesh governance model. Thirdly, a study into the design of a detailed and actionable Data Mesh execution framework is recommended. Finally, further research could dive deeper into the societal and organizational impact caused by the large-scale implementation of Data Mesh architectures.

Overall, this study has discovered that migrating to a Data Mesh architecture is a profound and complicated process. This process must therefore be handled in a meticulous and structured manner. A governing artifact such as the DMRM can provide actionable handles for organizations that need guidance throughout this course. This research has lead to the understanding that a Data Mesh architecture mainly revolves around the organizational culture, architectural decisions for moving data, and the supporting governance structure. Therefore, the choice of technology is rather an executive decision than an organizational priority.

LIST OF ABBREVIATIONS

CD - Continuous Deployment
CI - Continuous Integration
DMRM - Data Mesh Readiness Model
DSR - Design Science Research
DSRM - Design Science Research Methodology
DWH - Data Warehouse
EA - Enterprise Architecture
TAM - Technology Acceptance Model
SOA - Service Oriented Architecture

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 CONTEXT

Currently, our society is in the middle of a digital revolution. This results in a bigger amount of data to be processed by organizations: research of IBM has shown that more than 90% of all global data has been collected in the last five years (Marr, 2018). As a consequence of the ever-increasing amounts of data collected by organizations, the concept of "Big Data" has been developed. This concept refers to the collection and processing of data for various use cases on a massive scale (Einav & Levin, 2014). Consequently, 95% of businesses cite the need to manage the large and unstructured amount of data as a challenge for their business (Kulkarni, 2019). In addition to the fact that more and more organizations have experienced these problems in recent years, it has also become increasingly clear what the value of insights from data can have for their performance. In order to come up with a solution to the problems, as well as to recognize this value, large investments have been made by organizations in the development and design of data platforms. The purpose of these data platforms has been to be able to manage and process the collected data in a cost-effective and time-efficient manner (Dehghani, 2019). Because these platforms are primarily designed to accommodate huge amounts of data, there was an increased need for organization-wide data architectures that can be build up around these platforms. The functioning of the platforms within these architectures should be made possible by technologies that can recognize the amounts and different types of data (Saddad, El-Bastawissy, M., & Hazman, 2020). For a long time, these enterprise data architectures were built as a single pipeline consisting of a single data processing unit, with the data generators at the beginning of the pipeline and the data operators at the end of it (Selmadji et al., 2020). This way of designing a data architecture is referred to as a monolithic data architecture design. However, as a consequence to the aforementioned increasing amount of data to be processed, the ever-growing amount of data analysis techniques, and the amount of functionalities it has to address, these monolithic architectures have shown their limitations in different areas such as maintenance and debugging, scalability, maintaining quality data, and ownership and responsibility issues (Khazaei, Barna, Beigi-Mohammadi, & Litoiu, 2016; Khazaei et al., 2016; Lawal Moshood, Ileladewa Adeoye, & Lawal Habibu, 2020; Saransig & Tapia Leon, 2019; Selmadji et al., 2020). Thereupon, these limitations result in an overload in the capacity of data teams in response to the growing needs of the organization (Dehghani, 2022).

In an effort to address these limitations, some developments have already taken place in the field of alternative data architectures. Within these developments, the shift from a monolithic towards a distributed architecture design has played a primary role (Kalske, Mäkitalo, & Mikkonen, 2018; Lawal Moshood et al., 2020). However, current implementations of these distributed architectures focus on a decomposition of the monolithic unit into different service units, while still little attention is paid to coordination of ownership and responsibility over the data (Newman & Kotonya, 2015) and scalability of the data platform in terms of the amounts of data to be processed (Fleury, 2021). These long-standing issues cause an unsatisfactory alignment between organizational needs and the functioning of the architectures instituted (Moses, 2020). As a result, there is a demand for an enterprise data architecture that addresses both technological challenges, as well as the organizational needs of data processing organizations (I. A. Machado, Costa, & Santos, 2022b).

To overcome the aforementioned limitations of current enterprise architecture designs, a relatively new architecture design has been introduced, that is built upon earlier distributed architectures but uses a domain-driven decomposition instead of a service-driven decomposition (Dehghani, 2019; Ray & Pal, 2020). In 2019, Dehghani (2019) identified this new domain-driven distributed architecture as a Data Mesh, although other scientific publications have suggested a similar approach (Ray & Pal, 2020; Waseem, Liang, Shahin, Di Salle, & Márquez, 2021). For the sake of simplicity, this research continues addressing this new architecture design as a Data Mesh architecture design.

A Data Mesh is a domain-driven distributed architecture design, that is all about "recognizing and identifying data domains in an organization, as well as constructing an architecture based on numerous components that make up these domains" (Hokkanen, 2021). The main goal of a Data Mesh is to address the problems that arise when scaling data operations in organizations in terms of data availability and accessibility (Dehghani, 2022). The domain-driven design characteristic of a Data Mesh architecture tackles issues when different business units are working on large projects, speaking different business languages and formulating different requirements, as can emerge in monolithic architectures (Braun, Bieniusa, & Elberzhager, 2021). The bounded context in a domain-driven design defines clear boundaries for every business domain in the organization, ensuring that domain experts and data engineers within that domain are working closely together (Braun et al., 2021). The architecture design is made up of different domain components that exist interoperably but independently of each other (Enyo-one Musa, 2021). Compared to existing distributed data architectures, the Data Mesh concept mainly revolves around an organizational adjustment of the existing data architecture. This adjustment is made within the teams (e.g. they become distributed domain teams) and the division of roles within the teams (e.g. data engineers and data owners) (Fleury, 2021). This distributed Data Mesh architecture is built upon earlier alternatives to monolithic architecture and promises a more independent, scalable, and interoperable architecture comprising of multi-functional teams and a domain-driven data architecture, aimed to address the organizational challenges that were neglected by earlier enterprise architecture designs (Enyo-one Musa, 2021; Ray & Pal, 2020).

Implementing a Data Mesh architecture promises to deliver significant benefits over other approaches, but the process itself is not as simple as implementing a new tool or feature. Moreover, in order to actually deliver these benefits, all relevant processes, structures, and regulations should be included in the transition to a Data Mesh architecture. This is a drastic and complex process for organizations that needs to be thought through carefully before it can be carried out (Furia, 2021; Schultze, 2020). However, this process is hampered by the lack of a clear understanding of the required level of preparatory efforts in order to build readiness for a successful Data Mesh migration. In addition, it is still unclear to some organizations whether a Data Mesh architecture can offer a solution for their specific needs. Currently, literature does not provide a suitable Data Mesh readiness framework, making it difficult for organizations to make a well-considered choice on when and how to transition (I. A. Machado et al., 2022a). This gap in literature is filled by a study that investigates the factors that influence this migration and the necessary capabilities needed within organizations to start this migration. The outcome of this research can serve as a way for organizations to test whether they are ready to migrate to a Data Mesh architecture, or which improvements they still need to implement to prepare themselves for this migration. In addition, this assessment will give them an indication of potential areas of improvement with respect to a future migration to a Data Mesh architecture.

The scientific relevance of this research can be found in the fact that very little research has been done on the concept of Data Mesh, and that consequently there is no scientific shared view on the design of a Data Mesh architecture and its capability requirements (I. A. Machado et al., 2022b). Moreover, limitations of current enterprise data architectures have been addressed in scientific literature (Khazaei et al., 2016; Lawal Moshood et al., 2020; Saransig & Tapia Leon, 2019; Selmadji et al., 2020), as well as the benefits of a domain-driven Data Mesh architecture over these other architectures (Dehghani, 2019;

Furia, 2021; I. A. Machado et al., 2022b; Moses, 2020; Ray & Pal, 2020), but a detailed overview of specific architecture requirements addressing these limitations is still lacking. This research aims to fill the theoretical gap dealing with the migration to a Data Mesh architecture and the way this architecture addresses the limitations of current enterprise data architectures.

The practical relevance of this study lies in establishing a ready-to-use readiness assessment determining the state of readiness of organisations to migrate to a distributed Data Mesh architecture. Using this readiness assessment, organizations can assess whether or not they should start migrating to a Data Mesh architecture, and what improvements need to be made in order to enable the migration. These insights can be used by organizations to determine the as-is status of their current data architecture, and whether the necessary degree of features such as scalability and resilience of the data architecture is sufficient for future use (Balalaie, Heydarnoori, & Jamshidi, 2016).

The proposed research is part of obtaining a master degree in Complex Systems Engineering and Management (CoSEM). A CoSEM master thesis is focused on designing in socio-technical systems. Research in the field of data and data management nowadays focuses mostly on the technical functioning of the data, however recent publications point out that there is a need for a more comprehensive sociotechnical systems view to realize effective and safe integration of data management and data handling into organizations (Stalla-Bourdillon, Wintour, & Carmichael, 2019). This research fits the CoSEM program, as it is focused on designing a framework that guides organization through the decision-making phase of deciding whether and how to move the organization into an organization-wide architecture transformation.

1.2 RESEARCH GOAL

1.2.1 Research Objective

This research focuses on the adoption of a Data Mesh architecture in order to address the identified limitations of current enterprise architecture designs found in scientific literature. This research aims to guide organizations towards a better understanding of their current data architecture status, and whether the key capabilities of this architecture meet the level needed to migrate to a Data Mesh architecture. This is translated into the design of a readiness assessment that measures a set of predetermined and scientifically founded capabilities and determines whether these capabilities meet the required level needed for a Data Mesh architecture. In addition, this assessment enables organizations to identify capabilities that are still subject to improvement before being able to migrate to a Data Mesh architecture. The research objective of this study is formulated as follows:

"To develop a framework that enables organizations to assess their readiness for migrating to a Data Mesh architecture."

1.2.2 Research Deliverable

The research deliverable of this study is a readiness model. According to Al-Omari and Al-Omari (2006), a readiness assessment measures an organization's ability to undertake a transformational change by means of a systematic analysis, while identifying potential challenges that might arise when implementing new procedures and structures within the current organizational context. The readiness model indicates whether an organization is ready to make the transformation, or whether it should enhance or improve some of the indicated required capabilities. A systematic literature review will be conducted into existing readiness models, in order to use those as a basis for the model that is designed during this study. The building blocks for the readiness model in this research will be the theoretically and empirically identified factors that are needed for migrating to a Data Mesh architecture design.

1.3 RESEARCH SCOPE

Typically, reasons for organizations to migrate to a distributed system come from issues originating from their existing architecture being unable to process its vast amounts of data (Zimmermann et al., 2018). Therefore, it can be presumed that the scope of this research focuses on organizations either experiencing these problems or foreseeing these problems to occur in the future.

1.4 RESEARCH METHODOLOGY

Based on the research objective, the main research question is formulated below:

"How can organizations assess their readiness for migrating to a Data Mesh architecture?"

1.4.1 Research Design

In order to formulate an answer to the research question, a design science research (DSR) proposed by Hevner (2010) will be used. This research approach is chosen as it allows for the development of technology-based solutions to a relevant and practical problem. It is a qualitative research approach in which the object of study is the design of an innovative artifact that serves as a solution for the identified problem (Johannesson & Perjons, 2021). In case of this study, a readiness model for Data Mesh architectures will be designed. In order to do so, a researcher should develop an artifact for a specific practice, then distill the experience and information to inspire a broader solution (Johannesson & Perjons, 2021).

1.4.2 Research Setting

In order to elaborate on the DSR approach, a Design Science Research Methodology (DSRM) be used. DSRM includes a six-step research process with relevant research issues at each step. The phases of the DSRM are (1) Problem Identification and Motivation, (2) Objectives for a Solution, (3) Design and Development, (4) Demonstration, (5) Evaluation, and (6) Communication. The sub-questions of this research are formulated in order to structure the research within the phases. The first phase, "Problem Identification and Motivation", acts as a preparation phase for the explorative research. The first sub-question, guiding the process through the first phase, deals with the general exploration of a Data Mesh architecture. By conducting an extensive literature review and performing exploratory interviews with experts, the main characteristics of a Data Mesh architecture are identified. The first sub-question is therefore:

***SQ1**: What are the characteristics of a Data Mesh architecture?*

The next phase is "Objectives for a Solution". Since the proposed deliverable of this research is a readiness assessment framework, the objectives of this framework must be determined. The objectives of the readiness assessment are organizational factors that influence the readiness of an organization to migrate to a Data Mesh architecture. These factors can then be used as assessment criteria in the final model, which enable testing whether the organization meets the requirements. A first version of these factors are taken from the scientific literature, and serve as the input for the first theoretical version of the final model. The second sub-question is as follows:

***SQ2**: What are the factors that influence readiness for migrating to a Data Mesh architecture?*

In the next phase, "Design and Development", the process of designing the artefact begins. For the development of the readiness model, an modified version of the maturity model development theory of Becker has been followed (Becker, Knackstedt, & Pöppelbuß, 2009). A visual representation of this artifact development model as used in this research is shown in figure 1.1 (Becker et al., 2009). This methodology is chosen since it is based on the DSRM as proposed as the baseline of this research. In preparation for the design of the readiness assessment, a systematic literature research is carried out into existing and readiness-assessment frameworks. After analysis and comparison, these frameworks can be used as a basis for the artifact to be designed. By complementing this meta-model comparison by the theoretical knowledge from the former phase and by the insights of several interviews with Data Mesh experts, the design constructs can be determined for the first empirically founded version of the Data Mesh Readiness Model (DMRM). In addition, refinement sessions will be organized to evaluate this version. The third and fourth sub-questions are therefore:

SQ3*: Which readiness assessments with regard to IT architecture transformation are provided in literature?*

SQ4*: How to design a model that guides organizations through assessing their readiness for migration to a Data Mesh architecture?*

During the fourth phase, "Demonstration", the designed artefact is demonstrated in real-life case scenarios. According to Mora, Gelman, Steenkamp, and Raisinghani (2012), efficiency, effectiveness and impact of artefacts are all context-dependent and can only be fully assessed after deploying the artifact in a practical setting. To facilitate this, an assessment instrument is being developed that makes it possible to perform the readiness assessment within organizations that want to migrate to a Data Mesh framework. This is enabled by conducting a multiple-case study at these organizations. The assessment instrument is an operational representation of the framework for use in practice. The fifth sub-question is:

SQ5*: How can the readiness model be operationalized for practical use?*

During the last phase, "Evaluation", the artifact is evaluated through context-different test cases and individual evaluation sessions. During the test cases, the model is deployed within four different scenarios: both with and without the researcher's presence, and within organizations with and without experience with Data Mesh architectures. In addition, based on the Technology Acceptance Model (TAM), which defines evaluation criteria in the form of perceived usability and perceived ease-of-use (Venkatesh, Morris, Davis, & Davis, 2003), the artifact from this research is also evaluated on these characteristics during the evaluation sessions. The results from both the test cases and the evaluation sessions serve to determine the applicability of the DMRM in practice. Therefore, the sixth sub-question is:

SQ6*: Is the designed readiness model and readiness assessment instrument applicable in practice?*

1.5 THESIS STRUCTURE

The research questions as proposed in the former section are answered by dividing the research into three parts: theoretical background, design & development, and demonstration & evaluation. These parts also structure the outline of this thesis. A visual representation of this structure can be seen in figure 1.2.

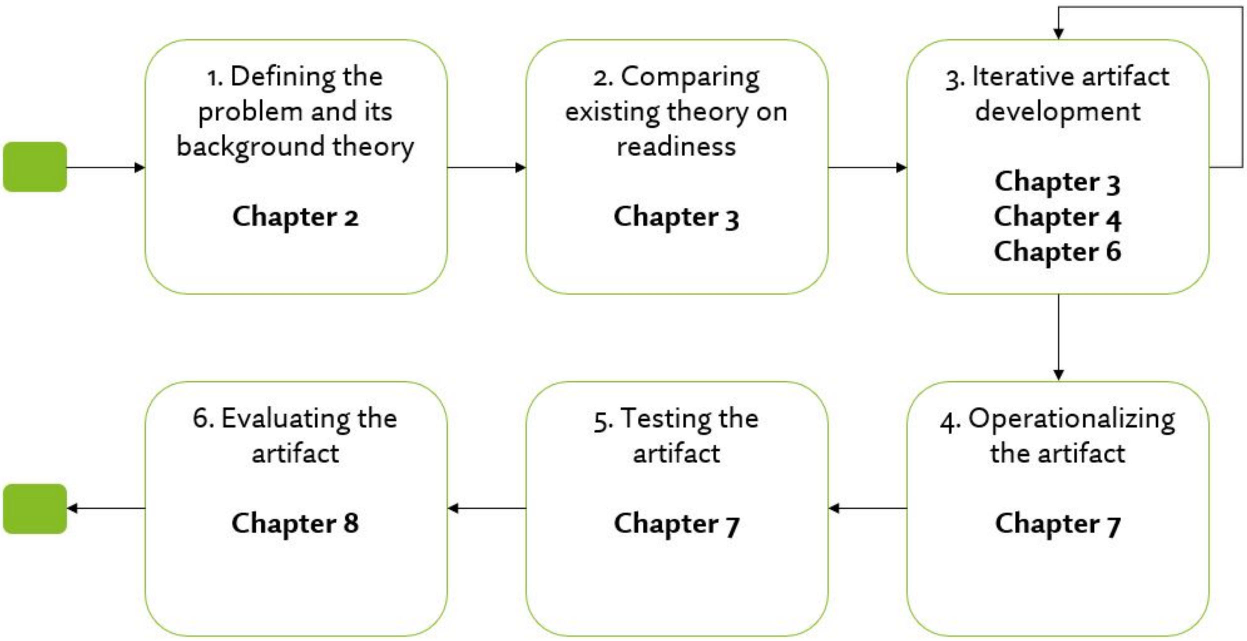

Figure 1.1: Artifact Development Model

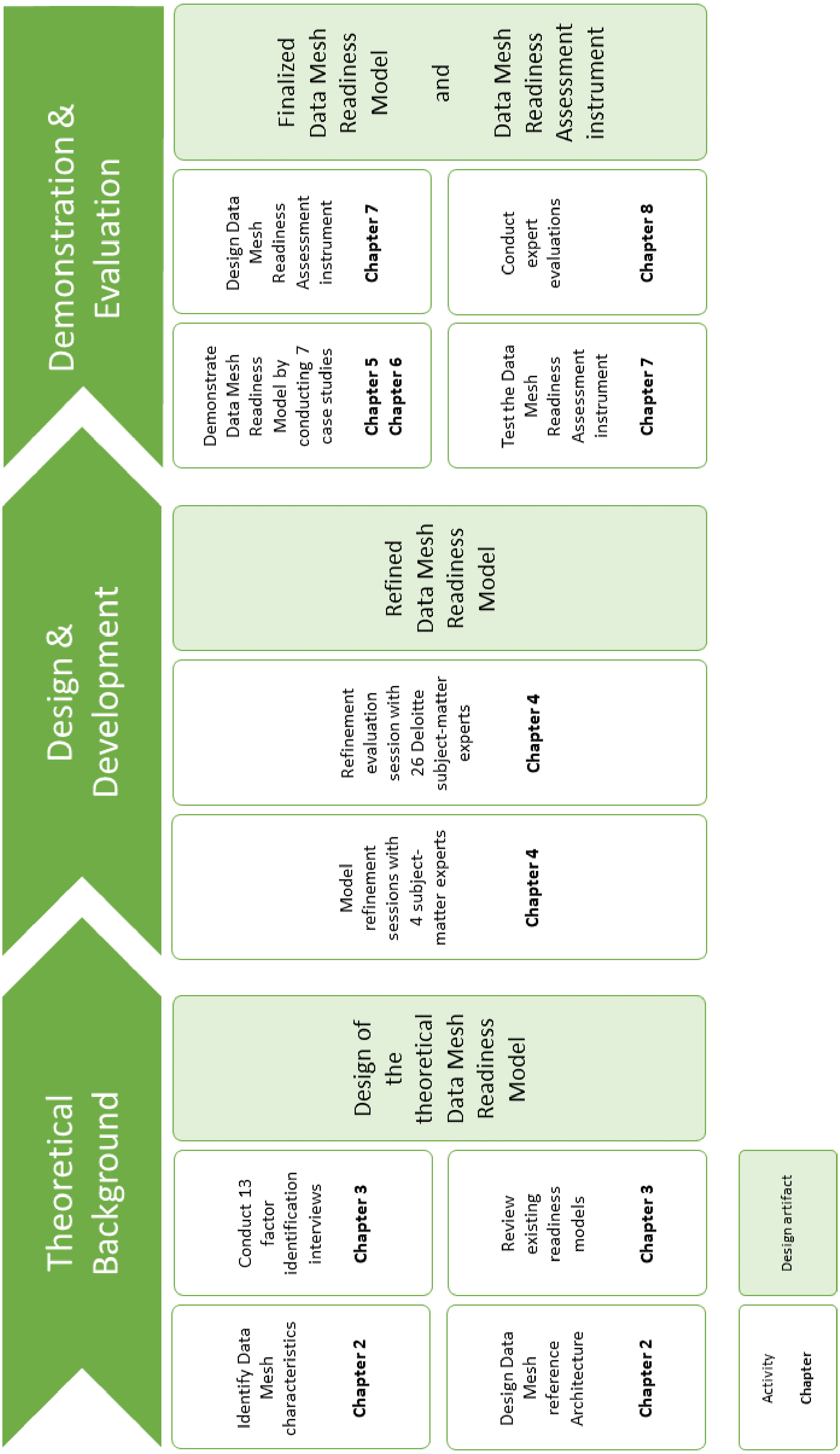

Figure 1.2: Research Flow Diagram

2 DATA MESH & RELATED CONCEPTS

In this chapter, concepts about and related to a Data Mesh are explained to provide theoretical background knowledge for reading through this research. The background knowledge is thus intended to serve as a theoretical foundation for the subsequent chapters in this report. Moreover, since the first research question addresses Data Mesh specific characteristics, this chapter also serves as a guideline for formulating an answer to this research question. The chapter is build up as follows: first, a generic explanation is given about data. Second, specific data architecture concepts are explained. Third, a Data Mesh architecture is explained in more detail based on its core principles. Fourth, the Data Mesh architecture will be positioned relative to existing data architectures, in order to discover the advantages and disadvantages of this new data architecture design. Lastly, overarching this chapter, a Data Mesh reference architecture is presented that visualizes the Data Mesh architecture design and its characteristics.

2.1 DEFINING "DATA"

2.1.1 Data as an Asset

Data a key asset in today's business environment. It is becoming increasingly important for organizations, as it creates competitive advantage by delivering insight in several analytics such as business performance and other performance indicators (Hagiu & Wright, 2020). It can be gathered by data generation teams, via applications and user-systems. It is not the data itself that carries value, but this value lies in the insights that analyzing this data provides. Subsequently, organizations have put efforts into gathering these insights and utilizing this information to base their decision-making on. For less experienced organizations, this could become a slow, costly and unscalable process (Hagiu & Wright, 2020). This process has been accelerated by the emergence of innovative IT advancements such as cloud computing, making it much easier for organizations to perform large-scale analyses on the information they have gathered throughout their business units and applications (Hagiu & Wright, 2020). Data has thus become an important foundation of digitization and has also been given the function of being an important business asset. In addition to being a driver of decision-making, data and its analytics have become much more functional, and are used by organizations throughout the entire business chain to improve their own business performance and gain a competitive advantage over its competitors (Hagiu & Wright, 2020). Subsequently, organizations have been investing more effort and funding in improving their data analytics and data management processes over the recent years, in order to guarantee the quality of the data itself and the insights it provides. Since the amount of generated data continue to increase, it is very important for organizations to have a solid and well-organized architecture that facilitates these processes.

2.1.2 Data Pipelines

Data Pipelines are the infrastructures that enable organizations to process data in multiple formats from distributed data sources with minimal human intervention. It starts with a data source and ends with a data sink, where data is moved by the data pipeline from one system or subsystem to another (Munappy, Bosch, & Olsson, 2020). There are two primary principles when it comes to transporting and processing data, in sense of processing the data in batch jobs, and storing it on the data platform. The terminology that is used to describe these principles refer to Extracting, Transforming and Loading (ETL) the data. This is a process that uses data pipelines to extract, transform and load data into a database that stores the structured data. The order in this process can be different: some architectures require data to be loaded into storage first and then transformed afterwards, storing unstructured data in the database. It depends on the data architecture which of these two processes can serve the business best (Engström, 2020).

2.1.3 DataOps

Data Operations, referred to as DataOps, takes its cue from DevOps, combining the ideas of integrating Development and IT Operations in order to accelerate delivery of changes and increase quality software (Ereth, 2018). DataOps uses the same philosophy, but applies it within the field of data engineering. The aim of this phenomenon is to emphasize continuous improvement and collaboration among the data teams within an organization. In addition, DataOps aims to boost automated data analytics within organizations and ensure high quality data. It can be described as a combination of technological operations, cultural benchmarks, enterprise system plans, and architectural arrangements (DataKitchen, 2021). As far as the data pipelines are concerned, DataOps tries to ensure that there is a high degree of automation in the various parts of the organization, where there is plenty of room for incremental change. Ereth (2018) uses the following working definition:

"DataOps is a set of practices, processes and technologies that combines an integrated and process-oriented perspective on data with automation and methods from agile software engineering to improve quality, speed, and collaboration and promote a culture of continuous improvement."

There is a reasonable scientific consensus that DataOps contributes to the flexibility and agility of organizations, and that therefore DataOps can be seen as a data foundation for data-driven organizations in the future (Atwal, 2020; Ereth, 2018; Munappy, Mattos, Bosch, Olsson, & Dakkak, 2020; Rodriguez, de Araújo, & Mazzara, 2020; Sahoo, 2019).

2.2 DEFINING DATA ARCHITECTURE CONCEPTS

2.2.1 Enterprise Architecture

An enterprise architecture (EA) is the technological and organizational infrastructure that enables the enterprise-wide sharing and processing of data. It is therefore not only an IT asset, but rather a strategic and organizational asset (Chen, Doumeingts, & Vernadat, 2008). The way how organizations arrange, secure, store and collect data can be governed by an EA (Enyo-one Musa, 2021). Enterprise Architectures can be seen as the data-skeleton of the organization that enables data analytics in order to conduct performance measures. An EA consists of many teams, competences, expertise and administrations, since it aims to arrange all the data assets around the data subjects of the enterprise. By administering strategic and tactical decision support, a complete and comprehensible EA keeps employees of the organization informed about its data arrangements and operations (Losey, 2004). An EA can therefore also be seen as a complementary architecture to an IT architecture, in order to govern system-wide organization and business context in which the IT units operate (Chen et al., 2008).

It is critical for companies to understand how data moves through the company, where it comes from, who processes it, and what information choices may be made on the basis of that data. The EA makes this public and hence gives a comprehensive picture of the organization's IT infrastructure. One of the major tasks of a strong EA is to create order from IT development and IT chaos. This task is fulfilled by the creation of an enterprise data model and an enterprise process model, both acting as the preconditions the EA should adhere to. An EA is then implemented as specified in these models, and over time structurally evaluated and iterated in order to maintain its successful output (Losey, 2004). Since there exists no single data architecture paradigm that fulfils all the needs of data workflows within an enterprise, a lot of complexities from the requirements of interoperability standards from the various business units of an enterprise arise when implementing a data architecture (Solano & Jernigan, 2012). An EA should both facilitate operational integration on all enterprise levels, as well as other enterprise considerations such as meta data, data security, access control and recovery strategies. It is therefore a very time-consuming and costly operation to implement an EA that integrates with all business operations at the operational level, as well as with other upstream and downstream business processes (Solano & Jernigan, 2012). It is understandable that the development of potential designs for an EA has received a lot of attention in recent decades. Due to the rapidly increasing amounts of data, types of data and functions of data within enterprises, a strong EA is a necessity to continue to carry out business operations (Lnenicka & Komarkova, 2019). It is therefore not surprising that the developments within EA designs are progressing at the same speed as the developments within the data management and analysis techniques themselves.

2.2.2 Monolithic Data Architectures

As a reaction on the first advent of data analytics as a means for business operations by organizations, It became necessary to adjust these firms' enterprise structures to the volume of data to be handled (Furia, 2021). The first sound EA was built around 1980, when the integration of data units and business operations was mostly addressed by implementing a central database and complementary interfacing systems (Vernadat, 2003). As an architecture with a central data team and a monolithic data management, performing all data operations from and to a single, centralized data platform, this architecture was the antecedent of the architecture that most organizations would eventually adopt inside their own EA (Enyo-one Musa, 2021). These data platforms frequently consist of a predetermined number of teams with specific expertise and responsibilities: data is generated by source teams, then flows into a large data storage that is maintained by data engineers, and data analysts extract this data from the storage to run their analyses on. This monolithic data platform may be thought of as a centralized storage system through which data travels through data pipelines. A typical monolithic data architecture consists of a User Interface layer that runs on the end-user's computer, a Business Logic layer and Data Access layer that process and analyse the data and finally a central DataBase that stores the data (Kalske et al., 2018). A high-level visual representation of this typical monolithic architecture design is shown in Figure 2.1. Specific forms of these monolithic architectures, such as siloed data warehouses and data lake architectures, have been the foremost ways of organizing the data architecture of organizations.

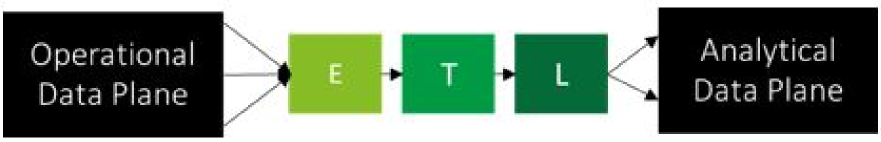

Figure 2.1: A high-level monolithic application architecture

Data Warehouse

One type of a monolithic data architecture is the data warehouse (DWH). The concept of a data warehouse dates from the 1980s, and is therefore one of the earlier variants. Characteristic of a DWH is that unlike other variants it contains structured data, which is extracted from different sources from the operational plane of the company and is processed by ETL transformations for analytical purposes. A

DWH is thus a clearly centralized data platform that serves as a central point of collection of structured data.

Data Lake

A few decades after the emerge of the DWH, around 2010, a new concept of a monolithic architecture was discovered, namely the Data Lake. A Data Lake differs from a DWH in that it contains unstructured data in raw format, collected from different sources and often in large volumes. The Data Lake contains unstructured data, which is transformed when the analytics are about to take place on the data. The Data Lake aims to integrate massive amounts of data into any format from any source of the organization, or even beyond. This was possible because major developments had already taken place in the orchestration of pipeline management in 2010, enabling the tasks of processing and modifying these huge amounts of data.

Limitations of Monolithic Data Architectures

Although these monolithic data architectures have been able to perform their function well for the last several years, they come with limited capabilities in terms of scaling, deployment and ownership distribution (Dehghani, 2022; Furia, 2021). To begin with, when the application grows in size, the monolithic application might be tough to adapt. This is because it is difficult to recruit new developers or replace team members in a growing architecture. A monolithic design inhibits developers and engineers from working independently since all components of the monolithic architecture are dependent on the performance of the other teams earlier in the pipeline (Namiot & Sneps-Sneppe, 2014). Moreover, since monolithic architectures are centralized platforms, typical issues associated with this centralization occur that have to do with the system becoming highly coupled and very hard to maintain. Since a change made to a small part of the application requires the whole monolithic system to be rebuilt, continuous deployment in large monolithic architectures is very difficult (Ponce, Márquez, & Astudillo, 2019). Furthermore, there is a persistent risk of a lack of ownership and responsibility, since various departments manage distinct data sets without any mutual communication. Because data creation, data transformation, and data analysis are all separate processes, there is little agreement on the purpose and quality requirements of the data (Furia, 2021). Due to the aforementioned limitations, change implementations in today's massive, complex, and rapidly expanding systems with monolithic architectures will be too sluggish and inefficient soon. (Krivic, Skocir, Kusek, & Jezic, 2018). A case study of Zalando, one of the largest online fashion shops, revealed that monolithic data systems had an obvious ownership problem: as the ones who generate the data, the producers have no knowledge what the intended goal of data analytics on this data is. On their side, data analysts have no idea where the data originates from, and there is a lack of overall ownership and accountability for the data sets (Schultze, 2020). Consequently, there is a lack of organizational scalability: as the volume or kinds of data grows, the platform's scalability suffers as a result of the central data storage bottleneck. Schultze (2020) shows that these issues appear in both data warehouses, as well as in data lakes, indicating that it is rather an organizational problem than a technical issue. Generically, it is affirmed in literature that monolithic data architectures are highly coupled and difficult to maintain (Ortiz et al., 2022). As a result, the central data platform faces the danger of cutting between domains, as domain expertise on these designs is dispersed among various technical teams on the platform. Because the duties for generating, converting, and analyzing expertise cut across disciplines, friction, a lack of data quality, and an inability to scale arise (Dehghani, 2019; Furia, 2021; Schultze, 2020). As a response to the aforementioned limitations of monolithic and centralized architectures, distributed architecture designs evolved, aiming to tackle these limitations and serving as a more scalable alternative.

2.2.3 Distributed Data Architectures

In an attempt to address the limits of monolithic data architectures, several studies emerged into breaking down the centralized architectures in smaller, decentralized components. These distributed data architectures promised better scalability and a solution for the many organizational challenges that the vast monolithic platforms brought with them. The shift from monolithic platforms to decentralized, more manageable and more scalable distributed platforms also came in response to emerging concepts dealing with continuous improvement, such as Agile and Scrum. These concepts require a continuous development of technologies and an iterative way of working, working with many different versions and iterations. To make this possible, it is important to be able to easily adapt the components of the architecture, which was not possible with large monolithic platforms.

A microservices architecture is one of the first alternatives to monolith architecture design. A microservices architecture is a distributed, modularized architecture made up of a collection of discrete, networked, and interoperable services, each with its own unique purpose and access to its own database (Ortiz et al., 2022). Each of these small services, in further particular, operates its own operations and connects with one another via lightweight mechanisms. They are based on business capabilities that might be used independently (Zdun, Navarro, & Leymann, 2017). The most major benefit of this architecture type over a monolithic design is agility, which is provided by modularized complex and resilient systems, allowing for speedier deployment of new solutions or upgrades to existing solutions (Nadareishvili, Mitra, McLarty, & Amundsen, 2016).

However, although microservices designs are promising in terms of scalability and deployment improvements, they still have not solved the issues around the lack of ownership and responsibility (Avci Salma, Tekinerdogan, & Athanasiadis, 2017). Because of the persistence of scattered domains, the organization's domain expertise is still spread throughout the data architecture as a whole, and the problem of lack of ownership persists (Hokkanen, 2021). As a result, a microservices design also leaves challenges in the field of efficient data utilization.

2.3 DEFINING "DATA MESH" ARCHITECTURE DESIGN

A Data Mesh is a domain-driven distributed architecture design, that is all about "recognizing and identifying data domains in an organization, as well as constructing an architecture based on numerous components that make up these domains" (Hokkanen, 2021). The main objective of a Data Mesh is to eliminate the challenges of data availability and accessibility at scale (Dehghani, 2022). The domain-driven design characteristic of a Data Mesh architecture advocates the establishment of independent domains in the architecture, working autonomously on their own data. This means that these domains are free to work on their own datasets and create their own data analytics, as long as they adhere to the global governance standards. Subsequently, this design aims to tackle issues when different business units are working on large projects, speaking different business languages and formulating different requirements for that project, as is happening in monolithic architectures (Braun et al., 2021). The bounded context in a domain-driven design defines clear boundaries for every business domain in the organization, ensuring that domain experts and data engineers within that domain are working closely together (Braun et al., 2021). The architecture is thus consisting of domain-driven components that operate interoperable and independently of each other (Enyo-one Musa, 2021). It is essentially an organizational change of current enterprise data architectures, rather than a technological transformation, in which the main focus lies on distributed data domain teams consisting of domain-specific data engineers and data owners (Fleury, 2021). This distributed Data Mesh architecture is based on earlier alternatives to monolithic architecture and promises a more empowered, scalable, agile architecture with multi-function teams and a domain-driven business structure, aimed at addressing organizational challenges that were overlooked by previous enterprise architecture designs (Enyo-one Musa, 2021; Ray & Pal, 2020). Figure 2.2 gives a simplified representation of a Data Mesh architecture design.

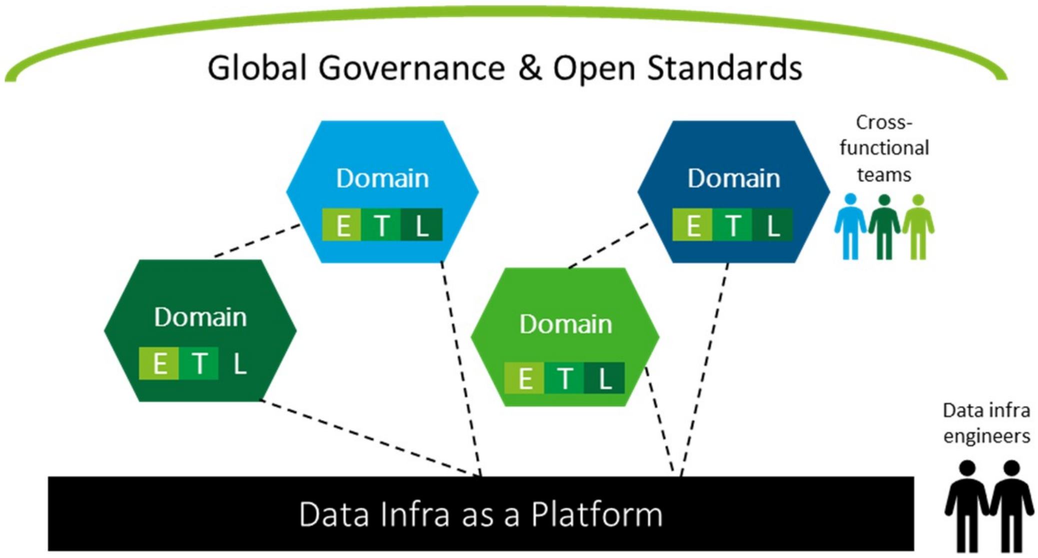

Figure 2.2: Data Mesh Architecture Design

This section dives deeper into the concept of a Data Mesh architecture design and defines its characteristics, design principles, advantages and challenges. Moreover, a reference architecture is designed in order to give a visual representation of the architecture design.

2.3.1 Characteristics of a Data Mesh architecture design

As proposed by Dehghani (2019), a definition of a Data Mesh is formulated as follows:

> *"An intentionally designed distributed data architecture, under centralized governance and standardization for interoperability, enabled by a shared and harmonized self-serve data infrastructure."*

A Data Mesh is built upon four core principles, which are a combination of existing ideas on Distributed Architectures and Domain-Driven Design (DDD) (Dehghani, 2022). The principles are discussed in the following part of this section.

Principle 1: Domain-Oriented Decentralised Data Ownership and Architecture

The first principle deals with decoupling and decentralizing the traditional monolithic data platform into a distributed variant. This decentralization is domain-oriented: that is, the decentralized components of the former large whole are selected on the domain they carry within the organization (Goniwada, 2022). The motivation behind this idea is about placing the ownership of analytical data for business domains close to the data itself, so that the processing, managing and analysing of the data can be done independently by the domain teams itself (Dehghani, 2022). The domains host and serve their datasets in an easily consumable way, while being close to the point of destination of the data (Serra, 2021).

When building this decentralized ownership into a data architecture, each domain must be able to independently implement its own data solutions, instead of having one centralized IT team that performs these deployments organization-wide. In terms of architecture, this means that each domain provides its own APIs, as well as an analytical data endpoint (Genovese, 2021). In this way, the domains can serve their own analytical data, without being dependent on other domains. Only when organizational dependency arises between domains, i.e. one domain depending on operational information from the other, an output port can be built in that serves simultaneously as input port for the other domains (Genovese, 2021).

Principle 2: Data as a Product

The second principle deals with the way data is viewed within the organization. Where data in the earlier data architectures was mostly considered to be a huge data repository, in a data mesh data is considered as a product that is ready-to-use for analytical purposes. Since the data in a Data Mesh becomes a prior business asset, the *Product Thinking* approach is applied to the data in a Data Mesh (I. A. Machado et al., 2022b). This approach sees data as a product, and this product needs to have a set of certain characteristics that maintain the quality of the data and the efficiency of the Mesh (Dehghani, 2022). These characteristics are in scientific literature formulated as the DATSIS principles: the data must be Discoverable, Addressable, Trustworthy, Self-Describing, Interoperable and Secure (I. Machado, Costa, & Santos, 2021).

- **Discoverable**
  For data to be discoverable, a central catalogue should exist within the company in which data owners must be able to register their data product in using some search engine, users can request access to the desired data products (Rigol, 2021).

- **Addressable**
  Having addressable data means defining standardized metadata that should be defined for every data product. In that way, data analysts are autonomous in finding and using the needed data, but data engineers have less interruptions from people asking where they can find specific data (Rigol, 2021).

- **Trustworthy and Secure**
  Trustworthiness can be captured in regularly checking the data quality and specific trustworthy characteristics of the data products within the enterprise. In order to adhere to security standards, these checks should be conducted automatically (Rigol, 2021).

- **Interoperable and Self-Describing**
  In order for datasets to be self-describing it is important that they use the same naming conventions and contain understandable metadata, which enhances the interoperability of the datasets (Rigol, 2021).

A data product usually is a published data set that can be accessed by other domains. Each data product can be managed independently of others, enabling them to be autonomous. In addition, the data products provide clearly and logically formulated data sharing contracts to ensure that the above-formulated characteristics of data products are guaranteed. Since these data products contain all the specifications needed to be managed autonomously within the organization, the components of a product also cover all the structural components needed to share data autonomously. For this, a product needs to contain (1) code on how to consume, transform and serve upstream data via pipelines, (2) data and metadata that can be served as events, batch files, relational tables or graphs depending on the nature of the domain data but maintaining the same semantic, and (3) an infrastructure components that enables the data product's code to be built, deployed, and executed (Genovese, 2021).

Principle 3: Self-Serve Data Infrastructure as a Platform

In order to enable the domain's cross-functional teams to share their data, a self-serve platform must be built into the Data Mesh (Dehghani, 2022). This platform is built and maintained by the central

IT organization, is domain agnostic and must enable users to surface data lineage across the Mesh. In addition, this platform enables users to control the full life cycle of individual data products, as well as to manage a reliable mesh of interconnected data products (Dehghani, 2019).

A few examples of these self-serve data infrastructure planes are proposed, such as a typical data product developer's main interface that manages the lifecycle of data products using simple declarative interfaces, or a set of global capabilities best offered at the mesh level in the form of a graph of connected data products (Genovese, 2021).

Principle 4: Federated Computational Governance

The last principle deals with the construction of a federated and global governance among the Data Mesh. Since one of the main limitations of the former monolithic architectures deals with a lack of responsibility and ownership, these two capabilities need to be clearly defined while still allowing room for interoperability, standardization and ecosystem-thinking. Interoperability allows other data teams to use data products in a consistent way, for example by having policies that define CSV files as the standard way to provide data. Standardization efforts are documented in order to discover and understand available data products consistently. An easy way to allow for this documentation is to provide an accessible document that contains a predefined set of information on the data-owner, location URL, descriptions of the CSV fields et cetera. Since the Data Mesh is distributed into single self-serve domain teams, these teams can make local and independent decisions on their own data processing and analyzing units. However, these decisions must adhere to global rules as defined by the global governance. In order to maintain interoperability and discoverability of data products and data sets, these global standards are defined by the ecosystem governance system in terms of how the data is defined (Genovese, 2021; Serra, 2021).

The organizational structure of this federated governance and global standards is challenging, but important for the functioning of the Data Mesh. The governance needs to set global rules on local data practices and decisions, and consists of principles underpinning the scope of the global governance. Moreover, the global governance is lead by a team that consists of members of the group from the domain, the self-serve platform and global compliance stewards that both have global and local incentives. Moreover, the global decision standards contain platform capability automating the decisions and computationally validating it continuously across all data products and domains (Dehghani, 2022).

2.4 DATA MESH RELATIVE TO EXISTING DATA ARCHITECTURES

Since the Data Mesh concepts are not widely presented in scientific literature, the comparison between the Data Mesh architecture concept and existing distributed architectures is made to elaborate on the position of Data Mesh relative to these existing concepts.

2.4.1 Differences Data Mesh Architecture and Existing Architectures

Starting with the comparison to a **microservices architecture**, the Data Mesh architecture idea has borrowed some major concepts from this architecture design. Similar to microservices, the Data Mesh architecture encourages polyglot technology solutions for each data product instead of for different services (Balalaie et al., 2016). In that way, the Data Mesh architecture can be seen as an extension of the characters of the well-established microservices architecture, build upon a domain-driven design in a distributed manner, but strongly focused on the domains and the responsibilities within the domains. The microservices architecture advocates the design of the data architecture based on the decomposition of the monolithic application into loosely coupled and independent components (Dragoni et al., 2017).

This idea originates from the earlier concepts of the **service-oriented architecture (SOA)**, which is the predecessor of the microservices design. However, within the microservices design, each service is operationally independent from other services and communication is only possible through published interfaces (Dragoni et al., 2017). In that, SOA is an enterprise-wide approach which considers the entire enterprise as scope for comprising the system, whereas microservices considers the application-level as scope (Blanco, Kotermanski, & Merson, 2007). The Data Mesh concept distinguishes itself here by using the domain level as scope. A visual representation of a Data Mesh architecture relative to other distributed architectures is given in Figure 2.3.

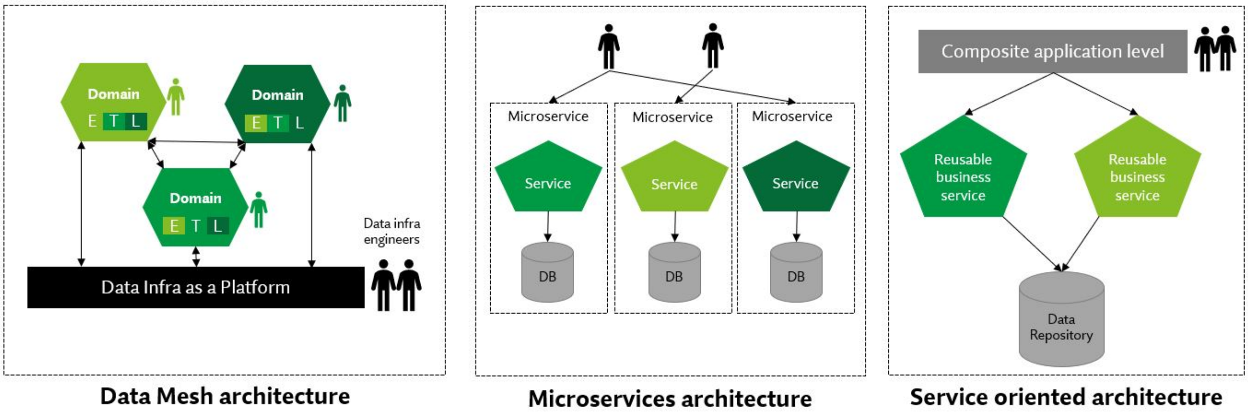

Figure 2.3: Various existing distributed data architectures

2.4.2 Benefits Data Mesh with Respect to Other Architectures

The benefits of Data Mesh have been widely reported on various gray literature sources. A Data Mesh architecture improves the accessibility of data, because the data products of domains are published in a discoverable way on a platform that is accessible to everyone in the organization. In addition, the data products must meet certain interoperability standards, so that the information contained in the data products is both understandable and usable.

In addition, the Data Mesh architecture aims to shorten the lead time between the operational data engineering teams, and the analytical data teams, simply by bringing them together in one domain. By bringing these teams closer together organizationally and performing the ETL work for only a domain-related amount of data, the bottleneck that existed in the central data platform will disappear.

Finally, the Data Mesh architecture advocates a clearer and more logical appointment of ownership over the data products over the architecture as a whole, by placing this ownership per domain-related data products with the domains themselves. Where the traditional centralized Data Lakes and Data Warehouses contain all data of the organization as a whole, and there was therefore no clear ownership of the data, it seems to be clearer in the Data Mesh which data belongs to which domain specifically (Saurabh, 2021).

2.4.3 Limitations of Data Mesh with Respect to Other Architectures

Although the limitations of the Data Mesh are much less discussed, there do exist challenges that arise in a Data Mesh specifically that are not or less present in the more traditional architectures. Firstly, there is a much greater need for data specialists within a Data Mesh architecture. Examples of these specialists are people who have the skill sets to deal with ETL operations, data tools and other technical actions that were initially done by the central data team. In a Data Mesh architecture, instead of a central team, these actions are performed by every domain in the architecture, simply requiring more of these people or an expansion of the skill set of the established people.

In addition, the Data Mesh architecture places a lot of emphasis on an independent approach to generat-

ing, analyzing and sharing data within the architecture, where these applications have previously always depended on a central data team. However, there will always be a need for a central body within the Data Mesh, and the boundaries between the dependencies of the central team and the independences of the domains themselves can sometimes be vague (Saurabh, 2021).
Finally, without proper alignment of the interfaces of the different business domains, there will exist unintegrated data silos in the domains themselves, because the domains will publish and share the processed data products only with the other domains. With these unintegrated data silos, there is a great risk of creating multiple copies of data if this data is also needed for applications within other domains, which in turn can be a problem in terms of data latency. The risk with these copies is that the moment a change is made close to the source of the data, it can cause major problems in the data quality of more downstream applications of this data within other domains. These cross-domain issues and dependencies will first have to be solved with, for example, extensive data virtualization solutions.
In conclusion, a Data Mesh architecture should primarily be seen as a socio-technical approach for solving contemporary problems with the accessibility, management and analysis of data within large and complex organizations. The organizational challenge primarily consists of changing the way of thinking, namely about data as products, and the way of working, namely with a high degree of self-service (Goetz, 2022). Although this is more of an organizational issue than a technological issue, technology is needed that can make this possible and prevent the organizational chaos in a distributed architecture. The aim of this technology should be to support and enable the various principles of the Data Mesh, while complying with a well-thought-out and partially federalized governance model that enables the Data Mesh and structures and organizes its decentralized nature.

2.4.4 Reference Architecture Data Mesh

Figure 2.4 shows a simplified reference architecture of a Data Mesh design. It shows an overarching federated governance structure containing global standards, which influences all independent domains in the Mesh. Within the domains, both operational data products and analytical data products are processed. The domains are served by self-serve data platform services that enable the enterprise-wide interoperable sharing of the data products. Within these domains, several data products exists. First of all, operational data is often ingested as raw and unstructured data, stored in some kind of database. Analytical data is the operational data after it has been cleaned and structured for analytical use. Moreover, domains can integrate data products from other teams as external data with the use of the aforementioned data governance policies. The final data product that is published by the domain is derived by aggregating all the data within the domain, after which it can be stored in the organization-wide data catalog.

2.5 CHAPTER CONCLUSION

This chapter presented the concepts about and related to the Data Mesh architecture design, in order to serve as practical background knowledge of this research. This background knowledge provided a theoretical foundation for the subsequent chapters in this research, and generated an answer to sub-question 1: *What are the characteristics of a Data Mesh architecture?*
The Data Mesh architecture is explained based on its core principles. Moreover, the Data Mesh architecture was compared with other existing architectures, specifically with the Service-Oriented Architecture and the Microservices Architecture. With domain level as a scope and a strong emphasis on independent domain teams, the Data Mesh architecture primarily distinguishes from these architectures on an organizational level.
With respect to benefits and limitations of the Data Mesh architecture relative to the other distributed architecture design, the benefits primarily lie in an improved accessibility of data, a shortened lead time between operational and engineering teams, and a clearer and more logical appointment over the data

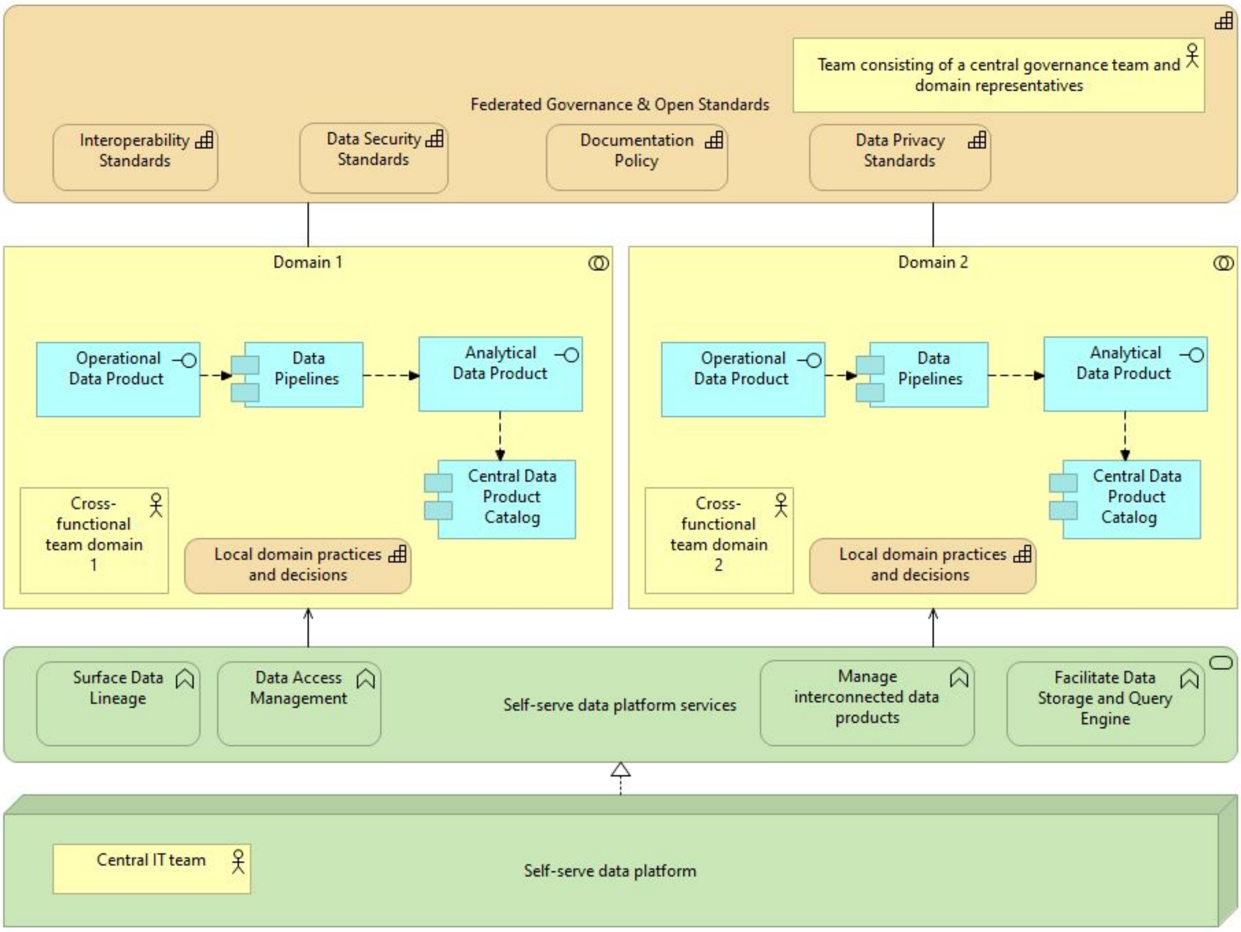

Figure 2.4: Data Mesh Reference Architecture

in the architecture. Additionally, the limitations of the Data Mesh architecture can be found in a greater need for data specialists, the persistent need for a central platform within the decentralized design, and the risk for unintegrated data silos in the domains.
Resulting from this chapter, it emerged that the Data Mesh architecture is mainly a socio-technical approach for solving contemporary problems with the accessibility, management and analysis of data within large and complex organizations. The Data Mesh architecture design is therefore more an organizational issue than a technological issue, but needs a certain set of technological applications in order to enable the operations within the decentralized architecture.
In the next chapter, a first version of the readiness model assessing organization readiness for a Data Mesh architecture will be designed based upon various research activities.

3 | DESIGN AND DEVELOPMENT OF THE MODEL

Building upon the theoretical background knowledge on Data Mesh architectures, a first version of the Data Mesh Readiness Model (DMRM) is designed in this chapter. The process of designing the model is structured in three parts.
First, existing literature is reviewed to identify generic factors that influence the readiness of organizations towards major IT transformations. Additionally, 13 exploratory interviews with Data Mesh subject-matter experts are conducted to complement this set with factors that specifically address readiness towards Data Mesh architectures. The identified factors serve as the building constructs of the DMRM.
Second, existing literature is analyzed to design a theoretical structure that captures these factors in a structured manner. These two activities end up with a first version of the DMRM.
Third, this first version of the DMRM is evaluated in practice, in order to conduct a last refinement on the model before proceeding to the next phase in this research. These refinements are performed on the basis of the insights from three individual subject-matter expert refinement sessions and the consensus from a panel session consisting of 26 Data Mesh subject-matter experts.

3.1 PART 1: CONSTRUCTS OF THE DMRM

The answer to sub-question 2 is captured in a conceptual framework, visually presented the identified dimensions and factors from scientific literature. This conceptual framework forms the base of the readiness assessment that will be empirically evaluated later in this research. In order to develop this conceptual framework in a structured manner, design choices have to be determined. In order to gain a clear overview on influencing factors within all facets of the organization, the conceptual framework must contain the following components (Fraser, Moultrie, & Gregory, 2002):

- a number of **dimensions** that address the various organizational facets;
- a number of **factors** that correspond to the dimensions;
- a generic description of the factors.

According to De Bruin, Rosemann, Freeze, and Kaulkarni (2005), the identification of domain components, i.e. dimensions, can be achieved through an extensive literature review. Therefore, the conceptual dimensions and factors that determine organization readiness towards Data Mesh adoption are extracted from literature in order to form a conceptual basis of the readiness assessment. By means of developing this basis, all facets of an organization that can be of influence on its readiness, including both organizational and technological factors, need to be included.

3.1.1 Dimension Identification from Existing Studies

First, the relevant dimensions were identified that address organizational facets within organization readiness. In order to structure the search for these dimensions, existing literature on readiness assessments was analyzed and investigated for similarities.

The existing studies that capture organization readiness within several dimensions were analyzed (Balasubramanian, Shukla, Sethi, Islam, & Saloum, 2021; Barham & Daim, 2020; Chanyagorn & Kungwannarongkun, 2011; Hussein, Mahrin, Maarop, & Abu Bakar, 2020; Joshi, Pratik, & Podila, 2021; Mirarab, Fard, & Kenari, 2014). From these studies, relevant dimensions capturing organization readiness were extracted and compared to each other. A complete table with these studies, their context and their identified dimensions can be found in Appendix C. In total, seven studies were analyzed.

In addition to the analysis of existing readiness assessment frameworks, the Technology-Organization-Environment (TOE) Framework is used as a basis for identifying the factors of the readiness assessment. This framework was developed by Tornatzky, Fleischer, and Chakrabarti (1990) and distinguishes three different dimensions of an organization's context on which the adoption of a technology can be tested: technological factors, organizational factors and environmental factors. Because this study is located in a multi-actor arena with different stakeholders, it is important to take into account all important dimensions in the readiness of an organization. The TOE framework provides an important guideline for drawing up these dimensions. In addition, it was decided to use this framework because it is independent of firm-size restrictions (Wen & Chen, 2010), which is important for this study because no restriction on the size of the organization was chosen. In addition, it is a framework that is widely used in scientific literature on implementation and adoption processes of technologies (Cruz-Jesus, Pinheiro, & Oliveira, 2019) and its use has been widely validated in the field of usefulness and reliability (Gangwar, Date, & Ramaswamy, 2015). In this study, the framework will be interpreted within the research context and will therefore be slightly modified, but will serve as a basis for establishing the dimensions of the organizational context in which the readiness assessment is designed.

After comparing and evaluating the existing readiness assessments and the theoretical TOE framework as presented above, six dimensions were formulated that structure the conceptual framework on Data Mesh readiness. Table 3.1 presents the final collection of dimensions including a description based on literature.

Table 3.1: Identified dimensions from scientific literature

Dimension	Description
Culture & Competences	The extent to which the culture and competences within the organization allow for migration to a Data Mesh architecture (Chanyagorn & Kungwannarongkun, 2011).
Responsibilities & Ownership	The extent to which formal channels and procedures enable employees to take on their supposed responsibilities and ownership over assets within the organization (Barham & Daim, 2020).
Strategy	The extent to which the strategy currently adopted by organizations is related to Data Mesh readiness (Avci Salma et al., 2017)
Governance	The extent to which the governance of organizations is suited for the migration to a Data Mesh architecture (Joshi et al., 2021).
Enabling Technologies	The extent to which technologies are implemented that support the migration to a Data Mesh architecture (Giebler et al., n.d.).
Central Integration	The availability of a central integration platform that enables organizations to properly adopt a Data Mesh architecture (Hokkanen, 2021).

3.1.2 Factor Identification from Existing Studies

Within the identified dimensions, several factors were extracted from literature that more specifically address the given dimension within a certain context. A visual representation of these factors is shown in figure 3.1. A detailed table on the identified factors and the underlying scientific sources is given in Appendix C.

Culture & Competences

The dimension Culture & Competences consists of factors that deal with the internal culture and the competences of the people associated with the organization. The way this dimension is designed is of influence on the success of migration (Hussein et al., 2020). It is the driving factor behind the motivation of transitioning to a different data architecture design. This dimension includes the following factors: a desire to transition (Chanyagorn & Kungwannarongkun, 2011; Hussein et al., 2020), the existence of a learning culture (Al-Ammary & Saleh, 2021; Bahadorpoor, Tajafari, & Sanatjoo, 2018; Barham & Daim, 2020; I. Machado et al., 2021), a change management strategy (Goniwada, 2022; Henry & Ridene, 2020; Priebe, Neumaier, & Markus, 2021), engagement of the top-management (Al-Ammary & Saleh, 2021; Bahadorpoor et al., 2018; Barham & Daim, 2020; Chanyagorn & Kungwannarongkun, 2011; Hussein et al., 2020) and a basic level of data literacy (Genovese, 2021; Goniwada, 2022; Hazel, n.d.; I. Machado et al., 2021; Oreščanin & Hlupić, 2021). The identification of these factors enables organizations to interpret the reason for the desire to migrate into a different design for their data architecture. Before starting to transition, organizations will have to indicate which of these factors are driving forces behind their willingness to transition into a Data Mesh architecture.

Responsibilities & Ownership

The dimension Responsibilities & Ownership dimension entails the following factors: clearly defined roles and responsibilities (Goniwada, 2022; Hokkanen, 2021; Loukiala, Joutsenlahti, Raatikainen, Mikkonen,

& Lehtonen, 2021) and clearly appointed data ownership (Barham & Daim, 2020; Goniwada, 2022; Hokkanen, 2021; I. Machado et al., 2021; Mirarab et al., 2014).

Strategy

As identified by both the TOE Framework, strategy and organizational factors play an important role when determining an organization's readiness for transformation (Cruz-Jesus et al., 2019; Lee, Kozar, & Larsen, 2003). The dimension Strategy entails therefor the following factors: the existence of clearly defined objectives (Barham & Daim, 2020; Chanyagorn & Kungwannarongkun, 2011; Hussein et al., 2020), a domain-oriented view on the organization as a whole (Goniwada, 2022; Gouigoux, Tamzalit, & Noppen, 2021; Hokkanen, 2021; Joshi et al., 2021; Loukiala et al., 2021; Oreščanin & Hlupić, 2021), the ability to identify current bottlenecks (Al-Ammary & Saleh, 2021; Hazel, n.d.; Hyperight, 2021; Mirarab et al., 2014), and an understanding in the degree of coupling between business applications (Goniwada, 2022; Hyperight, 2021; Loukiala et al., 2021; Mirarab et al., 2014).

Governance

The dimension Governance entails the following factors: an effective business and IT alignment (Gouigoux et al., 2021; Loukiala et al., 2021; I. Machado et al., 2021; Mehmandarov, Waaler, Cameron, Fjellheim, & Pettersen, 2021; Mirarab et al., 2014), clear communication guidelines (Goniwada, 2022; Mirarab et al., 2014), an effective data governance (Al-Ammary & Saleh, 2021; Barham & Daim, 2020; Chanyagorn & Kungwannarongkun, 2011; Hussein et al., 2020), and organization-wide standards for maintaining interoperability between business processes (Al-Ammary & Saleh, 2021; Genovese, 2021; Gouigoux et al., 2021; Henry & Ridene, 2020; I. Machado et al., 2021; Mehmandarov et al., 2021).

Enabling Technologies

While the transition to a Data Mesh architecture is mostly an organizational change, it also includes many associated technology practices. The dimension Enabling Technologies entails the following factors: accessible and discoverable data assets (Avci Salma et al., 2017; Barham & Daim, 2020; Joshi et al., 2021; Kotorchevikj, 2021; Loukiala et al., 2021; Solano & Jernigan, 2012), and the use of intelligent tooling (Joshi et al., 2021; Mirarab et al., 2014; Priebe et al., 2021).

Central Integration

The cental integration factors relate to the overall business integration processes that enable the integrated functioning of the data architecture of an organization. These factors relate to an organization's current state of readiness with regards to automated processes and operations that are required to take into account within a Data Mesh architecture design. The dimension Central Integration entails the following factors: an accessible data catalogue (Avci Salma et al., 2017; Joshi et al., 2021; Kotorchevikj, 2021; Solano & Jernigan, 2012), data quality management (Avci Salma et al., 2017; Joshi et al., 2021; Mirarab et al., 2014; Priebe et al., 2021), Continuous Integration (CI) and Continuous Delivery (CD) standards are identified as critical factors for the effective functioning of a distributed data architecture (Genovese, 2021; Gouigoux et al., 2021; Henry & Ridene, 2020; Joshi et al., 2021; Loukiala et al., 2021; I. Machado et al., 2021; Mehmandarov et al., 2021; Mirarab et al., 2014), and the use of DataOps for automated data operations (Goniwada, 2022; Gouigoux et al., 2021; Hokkanen, 2021; Joshi et al., 2021; Kotorchevikj, 2021; Loukiala et al., 2021; Mirarab et al., 2014; Oreščanin & Hlupić, 2021).

3.1.3 Factor Identification from Exploratory Interviews

Due to the exploratory grounds this research is based on, empirical data has to be gathered in order to refine the theoretical base of the artifact. Moreover, since the topic of Data Mesh architectures is

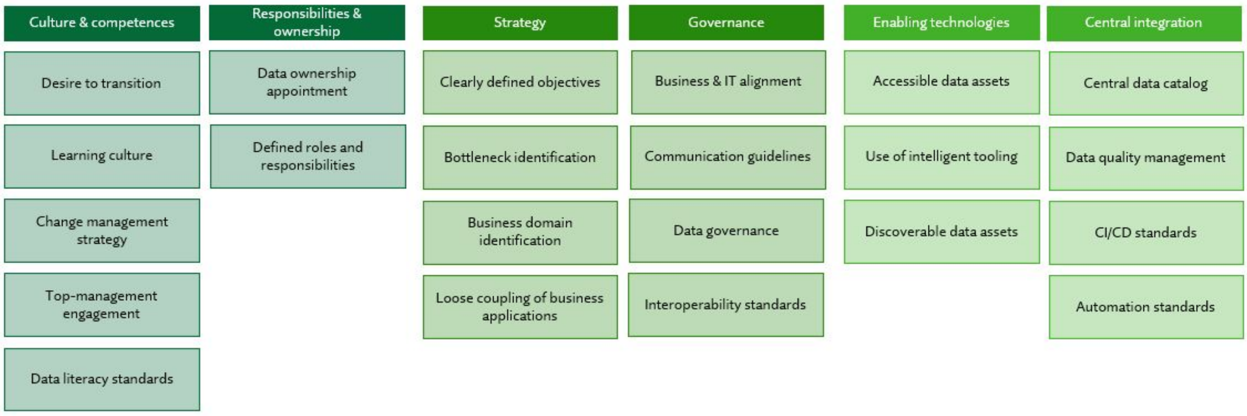

Figure 3.1: Overview of identified factors influencing Data Mesh readiness

very novel and little research has been done within this topic, empirical data is very valuable to get insights from the practical point of view. Moreover, by conducting these interviews, the dimensions of the theoretical model are reflected from a practical perspective. This creates a view on the model for organizations which can help to assess the readiness of an organization more specifically.

Setup of the Interviews

The interviews were conducted with individuals in different sectors and roles, all with experience in implementing Data Mesh architectures. These interviews were used to evaluate the factors extracted from the literature, and possibly to add unidentified factors. Among the interviewees, the extent to which they came into contact with data mesh implementations differed: some interviewees had only thought about it and written a process for it, other interviewees had already been involved several times with an actual implementation of a data mesh architecture. Because this research revolves around the development of an assessment tool that helps organizations to assess their readiness for migration, the emphasis during these interviews was also placed on the factors that are important for organizations in the preparation phase towards a data mesh architecture.
As also described in the more detailed interview protocol in Appendix D, the list of factors influencing organization readiness was run through with the interviewees and for all factors it was considered whether or not they were important to include in a readiness assessment. In addition, the interviewees were given room to reason from their own experience which factors were still missing from the list.

Interviewees

Table 3.2: Interviewees for refining theoretical model

Interviewee	Role	Operating Group	Years of Experience
A	Data Mesh Researcher	Engineering Research	5 years
B	Data Engineering Manager	Financial Services	6 years
C	Platform Product Manager	Public Services	10 years
D	Principal Scientist	Information Technology	20 years
E	Senior Manager	Financial Services	15 years
F	Chief Technology Officer	Information Technology	10 years
G	Senior Data Consultant	Information Technology	10 years
H	Managing Partner	Information Technology	7 years
I	Data Specialist	Public Services	8 years
J	Global Senior Director Data	Consumer Goods	5 years
K	Senior Customer Engineer	Information Technology	15 years
L	Principal Engineer	Consumer Goods	6 years
M	Principal Data Consultant	Information Technology	7 years

Additional identified influencing factors

Table 3.3 shows the additional factors that were identified during the exploratory interviews.

Table 3.3: Exploratory interview factor identification

Factor	Dimension	Interviewee source
Cultural support	Culture & competences	C, H, J
Innovation culture	Culture & competences	F, G, H
Business domain autonomy	Responsibilities & Ownership	B, G
Data engineer availability	Responsibilities & Ownership	A, B, F, G, H
Self-organizing domain teams	Responsibilities & Ownership	D, E
Central platform team	Responsibilities & Ownership	A, C, D, E, F
Domain representatives	Responsibilities & Ownership	C, D
Understanding of domain boundaries	Strategy	G, K, L
Identification of early-adaptor domains	Strategy	A, B, C, G
Agile way of working	Strategy	D, F, M
Data product publication standards	Data governance	A, C, L
Data democratization policy	Data governance	A, B, L
Data documentation policy	Data governance	C, J, M
Data literacy guidelines	Data governance	J, K
Monolith changeability	Enabling technologies	A, B, E, G, H, L
Monolith scalability	Central integration	A, B, E, G, H, L, M
Self-service platform	Central integration	C, I, K

Figure 3.2 shows the adapted version of the overview of the identified factors influencing data mesh readiness, including both the readiness factors as extracted from literature, as well as the identified readiness factors from the conducted exploratory interviews.

3.2 PART 2: STRUCTURE OF THE DMRM

The second literature review aims to answer sub-question 3: *Which readiness assessments with regard to IT architecture transformation are provided in literature?*

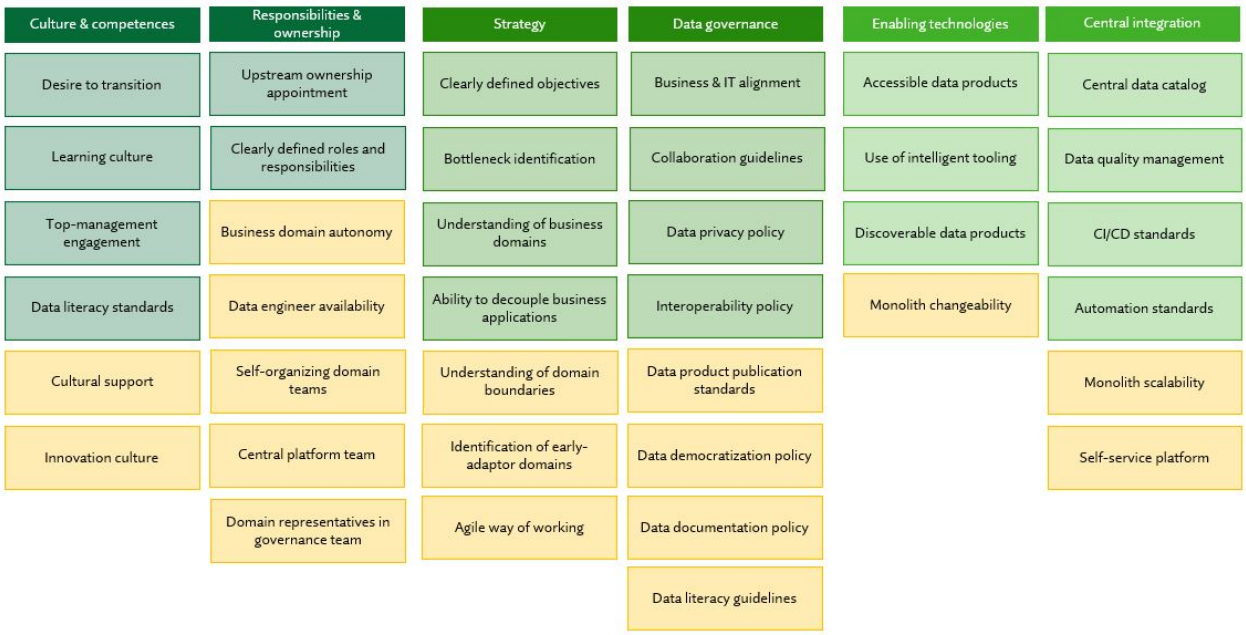

Figure 3.2: Adapted overview of factors influencing Data Mesh readiness

The successful establishment of a Data Mesh architecture depends on an organizations readiness and ability to adopt this architecture. Without proper readiness, it is likely that an implementation of a Data Mesh architecture will fail (Dang & Pekkola, 2016). Data Mesh readiness refers to an organization's assessment of how ready and prepared it is to adopt and establish a Data Mesh architecture within the current organizational context. A readiness assessment helps the organization to measure its readiness and identify potential areas for improvement (Jahani, Reza Seyyed Javadein, & Abedi Jafari, 2010).
Since the migration to a Data Mesh is not only a technological shift, but rather an organizational shift, its readiness deals with both IT architecture transformation as well as with organizational transformation. Therefore, two kinds of readiness models will be analyzed to design the theoretical DMRM. First, readiness models that deal with IT architecture transformation and second, readiness models that deal with organizational transformation.

3.2.1 Existing readiness models for IT architecture transformation

When reviewing scientific literature for existing readiness models that deal with IT architecture- or enterprise architecture transformation, many sources refer to the Business Transformation Readiness Assessment as adopted by The Open Group (TOGAF, 2018). This readiness assessment is build upon the work by the Canadian Government and its Business Transformation Enablement Program (BTEP) (Weisman, 2004). This assessment determines the Readiness Factor Dimensions that impact the organization when transforming to a new IT architecture. After identification, these readiness factors can be assessed by means of urgency, readiness status and degree of difficulty to fix. Another study conducted by Jahani et al. (2010) propose a method to assess readiness levels for enterprise architecture readiness within organizations, based on their own algorithm. However, due to the time constraints of this research project, this study will not be used as theoretical base for the development of the readiness model. Other IT architecture assessment studies that are provided in literature are merely used for assessing the functioning of the architecture, rather than assessing the readiness of an organization to migrate to a new architecture (Niemi & Ylimäki, 2007; Pruijt, Slot, Plessius, Bos, & Brinkkemper, 2012; van der Raadt, Bonnet, Schouten, & van Vliet, 2010; Vasconcelos, Sousa, & Tribolet, 2007; Velitchkov, 2008).

Therefore, the TOGAF Business Transformation Readiness Assessment, and foremost its Readiness Factor Dimensions, will function as theoretical base for the DMRM. These Readiness Factor Dimensions are

Need, Vision, Willingness to Change, Vision, Business Case, Leadership, Enterprise Ability, Accountability, Governance, Workable Approach, IT Capacity, and Enterprise Capacity.

3.2.2 Limitations on existing models for IT architecture transformation

A limitation of the TOGAF Business Transformation Readiness Assessment is that it functions as a static assessment on all levels of readiness, neglecting the organizational shift that is necessary when migrating to a Data Mesh architecture. Their assessment gives a good impression on what readiness factors can be addressed in order to assess readiness for migrating to a Data Mesh architecture, however, it mostly focuses on implementation measures than on organizational readiness for change. Moreover, it neglects a step-wise approach to assessing whether a Data Mesh is a suitable solution for the organization's problem identification, whether the organization is able to migrate to a Data Mesh architecture and whether it is actually ready for the migration. Since the migration is bound to a socio-technical context, both technical implementation challenges, as well as the socio- organizational challenges as described before need to take into account.

In order to capture the organizational complexity of this migration as well, organizational change models will be taken into account in the next subsection for designing the theoretical DMRM. A visual representation of this design is shown in figure 3.3

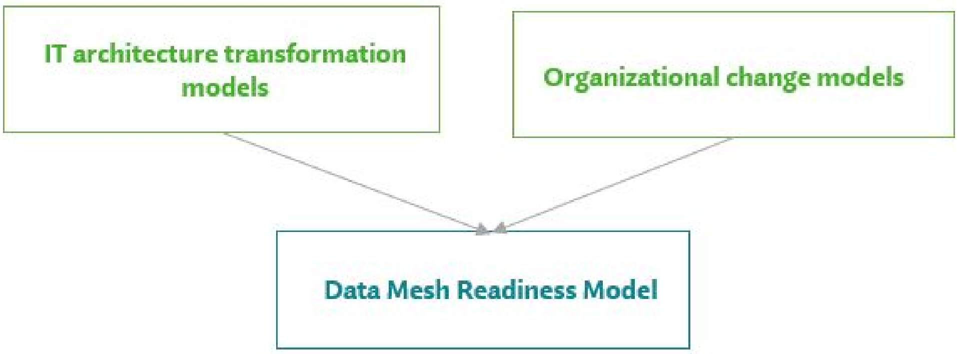

Figure 3.3: Schematic overview of the design of the theoretical model

3.2.3 Existing readiness models for organizational change

First of all, Buono and Kerber (2009) provide an overview of measures that need to be taken into account when addressing organizational change. These measures include developing an understanding and acceptance of the proposed change, a willingness and ability to change, building a change-supportive infrastructure, creating a change-facilitative culture, and ensuring ongoing stategizing. Moreover, they suggest that when the socio-technical uncertainty of the change is high, that the organizational readiness for change should be guided. Second, Pellettiere (2006) developed an organization self-assessment to determine the readiness for a planned change, by means of assessing an organizations ability to change. This ability to change has to do with contextual variables such as Vision, Mission, Core Values, and Culture. The results show that an organizations ability to change both gives a strong indication of the organizations readiness for change. Third, Vakola (2013) define readiness towards organizational change by means of Willingness to Change, Ability to Change, and Capacity to Change. Moreover, they draw a distinction between Individual Readiness and Group Readiness. Fourth, Agarwal and Prasad (1997) suggest that organization readiness should be assessed by starting with the Business Value, and subsequently assessing the current state and the projected future state of the organization. Moreover, Okorie Awa, Ukoha, and Emecheta (2012) argue that the needs and the drivers behind the change should be assessed in order to assess organization readiness. Lastly, Rosas and Camarinha-Matos (2009) identify a distinction between readiness and preparedness, and see preparedness as a facilitating precondition for readiness. Table 3.4 gives a presentation of the analyzed prior studies.

Table 3.4: Existing readiness models guiding organizational change

Author	Context	Readiness levels
Buono and Kerber (2009)	Organizational change	Understanding Acceptance Willingness Ability Change-supporting infrastructure Change-facilitative culture Strategizing
Pellettiere (2006)	Organizational change	Ability to change Readiness to change
Vakola (2013)	Organizational change	Willingness to change Ability to change Capacity to change
Agarwal and Prasad (1997)	Organization readiness	Business Value Current state Projected future state
Okorie Awa et al. (2012)	Organizational readiness	Needs Drivers Readiness
Rosas and Camarinha-Matos (2009)	Organizational readiness	Organization preparedness Organization readiness

Need, Ability, Preparedness and Readiness

In order to clarity on the different steps to determine organizational readiness as identified by several organizational change models, the distinction between the process-steps Need, Ability, Preparedness and Readiness for change needs to be made. First of all, change efforts need to achieve momentum and a sense of urgency. In order to establish this, a **need** for change has to be defined that touches upon this sense or urgency and **problem identification** (Smith, 2005). Moreover, the **ability** to change is seen as a prerequisite for entering a state of preparedness and eventually readiness, since this ability is a reflection of the organizations culture and value propositions, in sense of **high-level organizational requirements** which need to be recognized (Waterman Jr, Peters, & Phillips, 1980). Moreover, according to Rosas and Camarinha-Matos (2009), an organization is prepared if a set of **preparedness** conditions are met. These conditions are **formally defined rules and guidelines** that need to be in place in order to be prepared for change. This step has to be taken prior to readiness. Moreover, **readiness** entails the **complete set of forces** needed to perform the proposed change, by means of organization capabilities and technological functions (Harrison, 2014).

A conceptual representation of the various steps to determine organization readiness can be found in figure 3.4

3.2.4 Combining models on IT architecture transformation and organizational change

Based on the studies above, a first version of the DMRM has been developed. Due to the strong organizational and cultural nature of the migration to a Data Mesh architecture, a step-wise approach of first assessing the need, second assessing high-level contextual ability to migrate, third assessing formal preparedness for migration and lastly determining actual readiness is adopted. Moreover, with regards to the

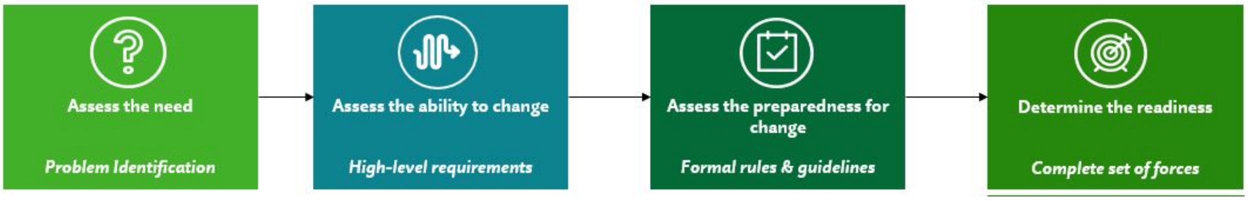

Figure 3.4: Organizational readiness assessment steps

technological nature of enterprise architecture transformation, the readiness factors as determined by the Business Transformation Readiness Assessment were mapped into their specific readiness levels.

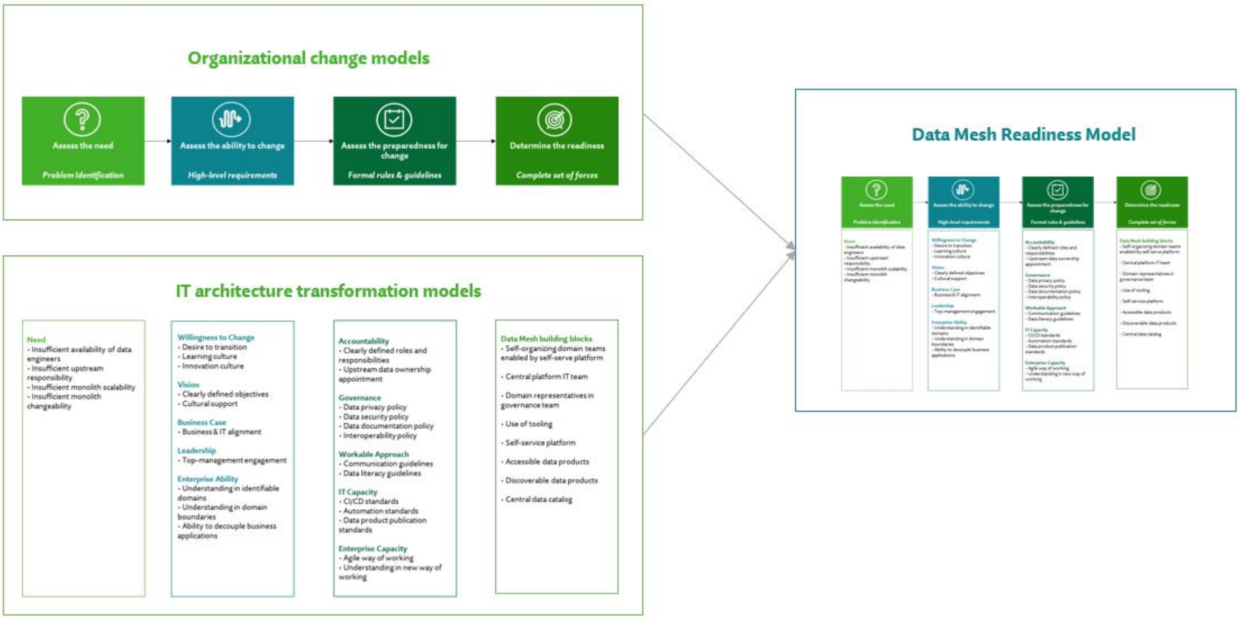

Figure 3.5: Systematic overview of the design of the theoretical model

Based on the description of the Readiness Factor Dimensions by TOGAF (2018), as well as on the definitions of the readiness levels Need, Ability, Preparedness and Readiness, the TOGAF Readiness Factors were divided over the various readiness levels. A blueprint of the theoretical DMRM is given in figure 3.6.

Mapping the identified factors to the DMRM

To finalize the design of the theoretical DMRM, the identified influencing factors from earlier in this section were mapped onto the DMRM blueprint. During this mapping, the definitions of the TOGAF Readiness Factor Dimensions as well as the different organizational readiness levels were taken into account.

This resulted in a readiness model with two axes: the horizontal organizational axis assessing the various steps of determining organizational readiness, and the vertical technological axis assessing the various Readiness Factor Dimensions. The factors as identified in the conceptual model were, with respect to their relation to these axes, mapped onto the DMRM. The model can be seen in figure 3.7. As stated by Stoianova, Lezina, and Ivanova (2020), a readiness assessment must satisfy the requirement of completeness, i.e. take into account all aspects of the factor dimension under consideration. Since the concept or readiness is vague, it is hard to ensure completeness of the assessment model. In order to address this reliability issue, the proposed DMRM will first undergo a series of subject-matter expert sessions, where after the model will be demonstrated in a multiple-case study setting. Both research

Figure 3.6: Blueprint of DMRM

steps allow for refinement and adjustment of the model and its containing readiness factors, in order to improve the completeness of the artifact.

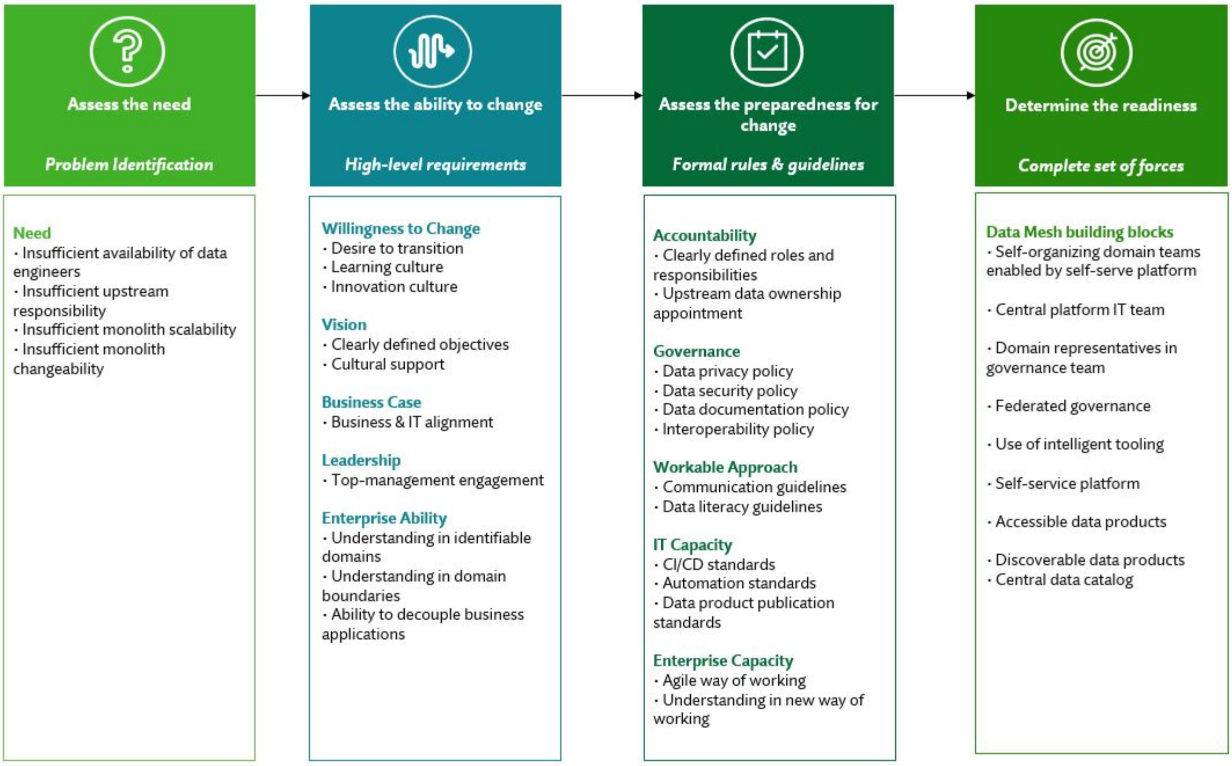

Figure 3.7: First Version of the Data Mesh Readiness Model (DMRM)

3.3 PART 3: REFINEMENT OF THE DMRM

In this part, the theoretical model as designed in the previous parts is refined on the basis of several subject-matter expert refinement sessions. Because the theoretical model was build up on many different sources of information (e.g. existing studies on readiness, exploratory interviews, existing studies on organizational change and existing studies on technology readiness), it was decided to make an extra iteration in the design and development of the model before it is demonstrated in practice. This design iteration is performed based on practical experience of the participants of this research step, so that the model is both theoretically and empirically founded before it is used in the next steps of the research.
First, the theoretical Data Mesh model is refined on the basis of refinement sessions with four different subject-matter experts. Subsequently, a refinement evaluation session was organized in which the refined model is evaluated on the basis of the suggested additions during the interviews. The practical value that is gained during this empirical research step helps to make the theoretical model more generically applicable for organizations in a practical context. This section ends with a refined version of the DMRM to be used for the following steps in this research.

3.3.1 Subject-Matter Expert Sessions

In order to refine the theoretical model, a number of subject-matter expert sessions are organized. These people are able to test the functioning and completeness of the model against their own experiences with assessing readiness for Data Mesh migrations, and are thus able to refine the model as a whole. In order to mitigate risk for bias, experts both from Deloitte as well as from external businesses and organizations were asked to join these sessions.

Interview 1: Subject-matter expert 1 (Deloitte - Data Mesh analyst)

During the first refinement session, the DMRM was walked through and provided with feedback. The subject-matter expert suggested to bring back the original factor dimensions into high-level People, Process and Technology categorization in order to give extra contextual clarity to the factors that are assessed in the model. Moreover, he suggested to think of a scoring mechanism when designing the Readiness Assessment Instrument, in order to indicate what dimensions the assessed organizations need to work on for future improvement.

Interview 2: Subject-matter expert 2 (Extern - Principal Engineer)

During the second refinement session, the interviewee mentioned that for them the lack of data engineers is not a real problem. Instead, they lack a coherent *data strategy* which results in the fact that a lot of data engineers are doing redundant work. This is a result of an alignment problem, more than it is a problem with the availability of data engineers. As a result, the interviewee stated that the list of Needs for a Data Mesh architecture could be expanded with *an insufficient data strategy*. Moreover, the interviewee addressed the importance of *continuous delivery* in an organization as a way to address the gap between data engineers and data analysts. Moreover, the interviewee emphasized the importance of painting the picture of organizational chaos as a means for assessing its ability to change. Lastly, the interviewee stated that the *insufficient availability of data engineers* should be rephrased to a *need for reduced cognitie load of data engineering teams*, in order ease the interpretation of the factor. Lastly, the interviewee stated that the *lack of domain knowledge* for data engineers that are positioned in new domains is automatically disentangled by putting the people together in a domain team and involving them in the normal domain operations.

Interview 3: Subject-matter expert 3 (Extern - Co-founder & Senior Consultant)

During the third refinement session, the major topic of discussion was the part of the model that guides assessing the need of organizations to migrate to a Data Mesh. Firstly, the interviewee stated that the need is not really covered by an *insufficient availability of data engineers*. This interviewee stated that for some organizations, these data engineers are available, but they do not adhere to an overarching *data strategy*. To some extend, this aligns with the statements of subject-matter expert 2 in subsection 3.3.1. He suggested to focus more on the addressing of the data platform bottleneck, than on the availability of data engineers. Moreover, this interviewee identified *long lead times* between stating a request to data operations and the data analysts receiving the results as an important driver behind an organization's need for a Data Mesh. This lead time is a specific example of the data platform bottleneck. Lastly, this interviewee identified an alignment of the *team structures* within the organizations with the business domains as an important prerequisite for an organization's readiness towards a distributed architecture.

Interview 4: Subject-matter expert 4 (Extern - Data Strategist)

Subject-matter expert 4 was interviewed during the fourth refinement session. First, the interviewee suggested that *Assessing the need* should be more centered on *Return on Investment* and *Data Products*. Moreover, the interviewee clarified that when talking about the *need for flexibility* and the *need for scalability*, these needs should be further defined by a lack of architectural flexibility to handle lots of use cases and a lack of architectural scalability when these platforms become very expensive to run. The interviewee therefore suggested to rephrase these factors to a *need for greater flexibility of the data platform* and a *need for greater scalability of the data platform*, in order to improve interpretability of the factors. Moreover, he proposed *an understanding in the identifiable domains* could be rephrased to *the ability to decentralize*, as that would also clarify the interpretation of the factor. Lastly, the interviewee addressed that factors concerning *employee skills* should be added to the model, as the migration to a Data Mesh requires learning a lot of new and improved skills.

3.3.2 Refinement Evaluation Group-Panel

In order to evaluate on the refinements resulting from the above described sessions, as well as to further improve the model, an intermediate evaluation session was organized to discuss the refinements as made so far. This session was attended by 26 Data Mesh subject-matter experts from Deloitte Denmark.
The evaluation session started with a general introduction to the topic of Data Mesh. Additionally, this research was elaborated on, as well as the DMRM and the exploratory interviews that have led to it.Lastly, the adaptations to the DMRM as suggested during the refinement sessions were discussed.

- The panel stated the importance of creating the domain teams and defining their boundaries. It was identified during the session that aligning domains together with the business structure is not always possible, especially not when there are teams that cross business domains. Examples of these teams could be teams maintaining ERP source systems or subsystem domain teams. Therefore, when decentralizing a formerly monolithic architecture, it was agreed on to structure the domain teams in line with the business structure as suggested by subject-matter expert 3, unless the system oriented teams are not forgotten in this approach.

- The importance employee skills and knowledge sharing opportunities, as suggested by subject-matter expert 4 was agreed upon. A way to establish this factor in practice could be for example by incorporating a knowledge sharing platform within the organization.

- The panel agreed upon the addition of the need for shorter lead times to the model, since they identified this need as one of the major drivers behind the desire to migrate to a Data Mesh architecture.

- The importance of CI/CD standards and automation standards was evaluated. The panel stated that these standards are not per se required, but would really be a *nice to have* feature within a Data mesh architecture. Especially with respect to shorter lead times, it would be important to include automation standards within your organization, since this automation would be key to liberate the resources within domains while scaling, instead of enlarging the maintenance efforts.

3.3.3 Refinements to the DMRM

Table 3.5 shows the refinements resulting from the refinement sessions and the evaluation session in this chapter.

Table 3.5: Results of the refinement sessions

Refinement	Source
Add the factor to the model: [Team structures aligned with business structure] in the *Assess the preparedness for change* phase	Refinement session 3
Add the factor to the model: [Knowledge sharing platform] in the *Assess the preparedness for change* phase	Refinement session 4
Add the factor to the model: [The need for shorter lead times] in the *Assess the reason to change* phase	Refinement evaluation session
Change the importance of the factor: [CI/CD standards] to *Nice to have*	Refinement evaluation session
Change the importance of the factor: [Automation standards] to *Nice to have*	Refinement evaluation session

3.4 CHAPTER CONCLUSION

In this chapter, the DMRM is designed. The design process of the model was build up in three parts, structured by the corresponding sub-questions.

The first part aimed to answer sub-question 2: *What are the factors that influence readiness for migrating to a Data Mesh architecture?*. By reviewing existing literature and conducting exploratory interviews, a conceptual framework was developed containing influencing factors on Data Mesh readiness.

The second part aimed to answer sub-question 3: *Which readiness assessments with regard to IT architecture transformation are provided in literature?*. After reviewing literature, a list of existing studies was extracted on technology readiness assessments. These studies formed the base of the structure of the DMRM. Since the analysis within this research shows that Data Mesh readiness emerges from both the technological dimension of the organization, as well as from the organizational side, the DMRM has been designed upon a combination of technology readiness models and organizational change models. The two-dimensional structure of the DMRM ensures that both the organizational side of the migration as well as the technological implications can be taken into account. Finalizing, the factors as identified in the first part were mapped onto their corresponding organizational readiness-step in the DMRM. Consequently, a first version of the DMRM emerged, as shown in figure 3.7. This model is used to formulate an initial answer to sub-question 4: *How would a model look like for organizations which want to assess their readiness for migration to a Data Mesh architecture?*.

The third part of the model design process consisted of model refinement activities, in order to evaluate the outcomes of the former parts and to gain a stronger empirical foundation for the design decisions. The refinements were made based on the insights from several refinement sessions and the consensus from one group-session. Overall, the refinement activity in this research serves as additional iteration step on the previous version of the DMRM. The refined version of the DMRM, as shown in figure 3.8, concludes this section and will be used for demonstration and testing purposes in the following research activities.

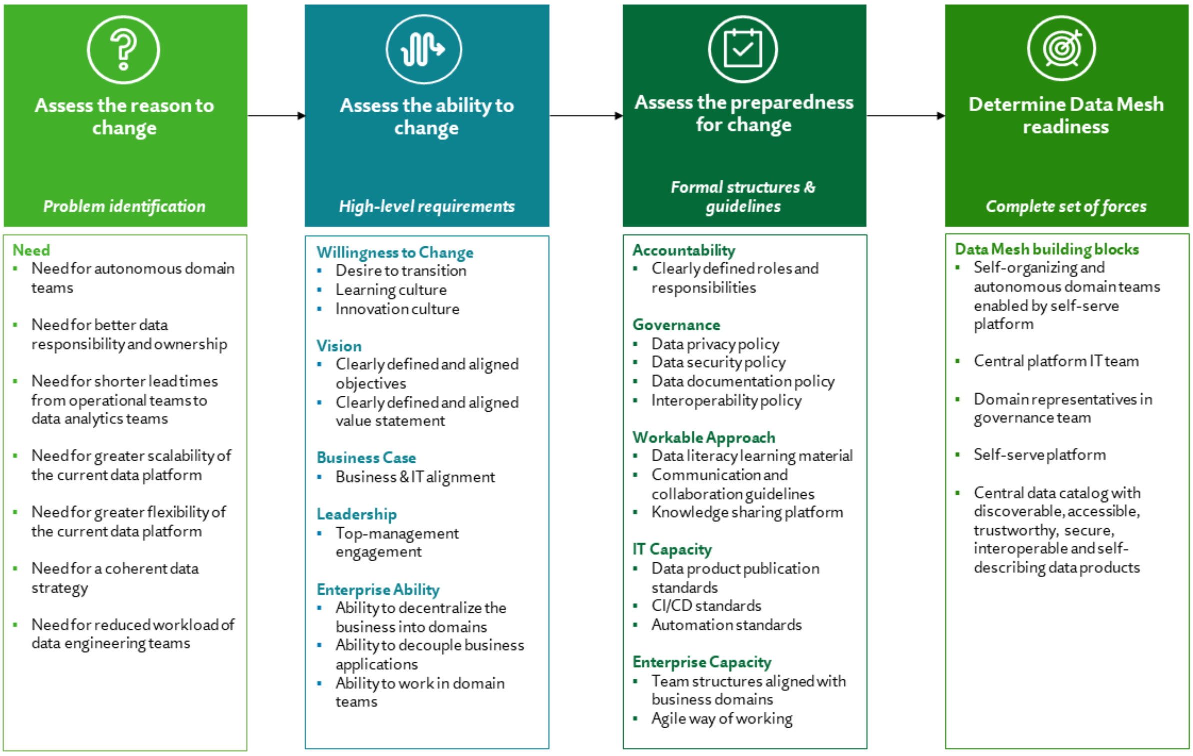

Figure 3.8: The Data Mesh Readiness Model (DMRM)

4 DESCRIPTION OF THE DMRM

In this chapter, descriptions are given on the structure of the DMRM, the factors in the DMRM and the way to use the DMRM in practice. First, the structure of the DMRM will be elaborated on, providing the purpose of and an explanation on its two-dimensional design. Second, the factors within the DMRM are described in detail. The follow-up questions that are formulated for every factor initially serve as support for the definition of the factors. Moreover, during the upcoming testing activity, they serve as guideline for the self-assessment function of the DMRM. Finally, attention is paid on the way in which the model can be used. This section explains the purpose and functioning of the model.

4.1 DESCRIPTION OF THE STRUCTURE OF THE DMRM

The DMRM is structured onto two dimensions: on the horizontal axis the DMRM is build up in several **organizational steps** towards Data Mesh readiness, and on the vertical axis the DMRM consists of several **factor dimensions** influencing the particular readiness steps.
Due to the complex nature of a Data Mesh architecture in both organizational and technological areas, the design of the model provides structure in the long-term migration process of organizations towards Data Mesh readiness.
In addition, because it became apparent during the formulation of the knowledge gap that it is not always clear to organizations whether a Data Mesh addresses their respective needs, the first step of the DMRM consists of problem identification that must be completed before the process is started. If this problem identification shows that a Data Mesh architecture does not sufficiently address their needs, it is not necessary for the organization to resume the next assessment.
The final step of the DMRM covers the overall technological and organizational readiness for a Data Mesh architecture, based on the structural principles of the Data Mesh concept. A remark that should be made here is that it is not necessary for all organizations to implement every specific principle in detail: going through the DMRM will show which parts of the organization need more attention than other parts.

4.2 DESCRIPTION OF THE FACTORS IN THE DMRM

This section describes the factors from the Data Mesh Readiness Model. In order to enhance comprehensibility of the factors, the descriptions also contain follow-up questions that help the participants of the assessments to assess their own factor capabilities.

4.2.1 Assessing the reason to change

Need

- **Need for autonomous domain teams**: Indication whether there is a need for domain teams to be able to do their own data analysis. Can they be independent in how they address their own domain analytics? Are the teams experiencing organizational and technical dependencies? Is the accountability of the data close to where the data is produced?

- **Need for better data responsibility and ownership appointment**: Indicate whether there is a need for greater knowledge over the data that is worked with, in terms of the fact that the people that work with the data are the people that know the data. Do the data engineers know the value and quality of their own data sets? Are the domain experts involved in delivering their data to the data analysts? Do the data teams understand what they contribute to in a business context?

- **Need for shorter lead times**: Indication whether there is a need for shorter waiting times for the analytics team to have their requests to the engineering team processed.

- **Need for greater scalability of the current data platform**: Indication whether there is a need for the current data platform to process bigger amounts of data through the data platform in the near future. Moreover, indication of the possibility and ease of adding business units to the data platform? What is the possibility and ease of growing the resource utilization of the current data platform?

- **Need for greater flexibility of the current data platform**: Indication whether there is a need to ease making changes on the existing IT system/data platform. Is the platform build on premise or is it easily accessible by others? Is the platform changeable, or is it too complex to be changeable? Is there siloed information or siloed data in the platform?

- **Need for a coherent data strategy**: Indication of the existence of a coherent data strategy. Are the data engineers rebuilding the same platforms? Are the data engineers and data analysts aware of each other's needs and purposes? Is there a coordinated, common data strategy? Is data an important asset of the organization? Is decision-making informed on data?

- **Need for understanding of the domain complexity**: Indication whether there is a need for the lessening of the current domain complexity, in terms of the existence of comprehensibility of inner domain processes and understandability of the end to end processes. Do the domains understand what they contribute to? How many data sets exist in the domains? Do the domains implement their processes independently and in a siloed manner?

- **Need for reduced workload of data engineering teams**: Indicate whether there is a need for healthier working conditions for the data engineering teams. Is the team always full? Is the data engineering team causing a bottleneck in the data processes? Are the data engineering teams satisfied with their workload (i.t.o. stress levels)? Are the data teams making their deadlines (i.t.o. quality of work)?

4.2.2 Assessing the ability to change

Willingness to Change

- **Desire to transition**: Indicate whether there exists a commonly shared desire to transition to a Data Mesh within the organization as a whole and among the individuals within the organization. Are they willing to take the efforts needed to establish this transition? Are the existing teams eager to be autonomous? Is there a willingness to participate in the Data Mesh migration?

- **Learning culture**: Indicate whether there exists a culture that supports learning. Are the people motivated to learn new procedures, such as working Agile? Is there management involvement in establishing this culture? Are there resources available to support this culture? Do the organization's mission and vision statements support this culture?

- **Innovation culture**: Indicate whether there exists a culture that supports innovation. Do the people have interest in working with data more efficiently? Are the people open for adapting to new technologies such as DevOps and DataOps? Is the organization as a whole data-driven? Do the organization mission and vision statements drive this culture? Are there resources available to support innovations? Are there organizational processes or facilities implemented that facilitate this culture? Are people experiencing long decision paths when implementing innovative projects?

Vision

- **Clearly defined objectives**: Indicate whether the objectives for migration to Data Mesh are clearly defined and aligned with the people in the organization. Why does the organization wants to migrate to a Data Mesh? What are the needs that are addressed by the migration to a Data Mesh? Does there exist common awareness about these objectives? Are there performance measures in place for measuring the progressions?

- **Clearly defined value statement**: Indicate whether the value of a Data Mesh to the organization is clearly stated and aligned with the people in the organization. What is the value that it will bring to the organization? Is there an idea on the returns on investment? Do there exist reflections on past projects that brought value?

Business Case

- **Business & IT alignment**: Indicate whether the data operations and processes are aligned with the business needs. What are the most important business requirements, and are they in line with the data strategy? Is the data team aware of their contributions to the business performance? Do the business users wish to be involved in the curation and processing of data from source applications to the reports that they use? Has the organization benchmarked itself with respect to similar organizations? Is there an understanding aout the impact that IT has on business processes? Is there an understanding about the amount/existence of organization strategic goals and requirements supported by IT strategic goals? Are Business and IT aligned with prognosed future growth? Is there an undertsanding of the digital capabilities needed to support the organization's business strategy? Will the Data Mesh architecture improve the organization's current Business and IT alignment?

Leadership

- **Top-management engagement**: Indicate to what extend the top-management of the organization is engaged with the migration to a Data Mesh. Does the migration align with their interests? Do the business leaders understand the Data Mesh? Do they understand the need for a Data Mesh? Have they allocated time and resources to the development of the Data Mesh migration?

Enterprise Ability

- **Ability to decentralize the business into domains**: Indicate to what extend the organization is able to decentralize its central architecture (in terms of processes and organizational design) into different domains. Does the organization understand the domain driven design principles? Can the organization map its business into different domains? Is there an idea on the domain boundaries? Are the boundaries distinct and explicit? Is there a reference design on the perceived decentralization of the organization? Is there an idea on the preferred degree of decentralization? Is the organization able to shift (a part of) the centralized decision-making into decentralized decision-making (e.g. across multiple teams?) Is the organization able to shift (a part of) the centralized roles, functions and tasks into decentralized roles and responsibilities?

- **Ability to decouple business applications**: Indicate to what extend the existing business applications (in terms of data processing) can be decoupled into the domain-oriented teams. Can the organization appoint the existing tables and pipelines to domain teams? Can the existing monolith be split up into smaller systems? Is the organization very much application focused, so that they don't break well into domains? Are the business processes very much centralized in ERPs? Is the organization able to establish a decoupled application architecture that allows each component of the application to perform its tasks independently in the domain teams? Can the components of the business application remain autonomous of each other? For organizations using pre-built applications: can these applications independently run across the Data Mesh (e.g. across multiple autonomous teams)?

- **Ability to work in domain teams**: Indicate to what extend the domain teams understand the new domain-oriented way of working. Do the data analysts, that were primarily focused on getting insights from the data, understand how to maintain data pipelines? Do the domain teams understand their extra responsibilities? Do they understand the concept of a Data Mesh? Do they understand the data-as-a-product thinking? Do they understand data products? Do they understand how to work together?

4.2.3 Assessing the preparedness for change

Accountability

- **Clearly defined roles and responsibilities**: Indicate to what extend the roles and responsibilities are correctly described. Will all the domains have a data product manager? Will the domains have data engineers, and are they given space to spend more time understanding the data? Do there exist data ownership guidelines? Does there exist no claimed ownership over the same applications? Are the roles and responsibilities of the team members in line with their personal skillsets?

Governance

- **Data privacy policy**: Indicate whether there exists a data privacy policy.

- **Data security policy**: Indicate whether there exists a data security policy.

- **Data documentation policy**: Indicate whether there exists a data documentation policy, that states how to document what a data product means and how to define a domain.

- **Interoperability policy**: Indicate whether there exists an interoperability policy, that states how to enable interoperability between domains and their domain related data-products. Does this policy allow for the fast consumption of data products?

Workable Approach

- **Data literacy learning material**: Indicate whether there exists learning material that enhances the data literacy among the people in the organization, in order for them to work autonomously on data products without the need for data engineers. What tools do they use for their data products and how do these tools work? How to understand the data? How to publish the data products? Does it state how to make data-based decisions? Does it support understanding, learning, analyzing and managing data products? Does it create overall data awareness among the organization?

- **Communication and collaboration guidelines**: Indicate whether there exist guidelines that enable and ease communication and collaboration between the various domain teams. How to enable cross-domain collaboration through data exchange? How do you collaborate on a data product, how do you work together to build a data product?

- **Knowledge sharing platform**: Indicate whether there will be knowledge-sharing facilities in place that enable the education, upskilling and sharing of knowledge within the organization. Is the platform accessible and interactive? Does the platform support Data Mesh engagement? Does it include information, guidelines, policies, and instructions on Data Mesh architectures?

IT Capacity

- **CI/CD standards**: Indicate whether there exist CI/CD standards that enable continuous integration and continuous delivery. Do these standards support the efficiency of the performance of the Data Mesh? Are these standards frequently measured in terms of cycle time, change failure rate and deployment frequency?

- **Automation standards**: Indicate whether there exist automation standards for standard software procedures like data quality checking, pipeline creation, etc. Is the organization able to automate software and data processes? Do these automations improve the efficiency of the performance of the Data Mesh?

- **Data product publication standards**: Indicate whether there exist standards that state how and when to publish data products. What is the standard publication format? What needs to be in a data product? Are there technology standard with which people should build a data product?

Enterprise Capacity

- **Agile way of working**: Indicate whether the organization supports and encourages an Agile way of working throughout the organization.

- **Team structures aligned with business domains**: Indicate whether the domain teams are business aligned, int terms of that they have responsibility for a certain part of the business. Will the domain teams not be too far away from the business cases? Does the organizational structure support a Data Mesh way-of-working?

4.2.4 Determining the readiness for a Data Mesh

Data Mesh building blocks

- **Self-organizing domain teams enabled by self-serve platform**: Indicate whether the organization has established self-organizing domain teams, that are able to cross-collaborate with other domain teams and that are enabled to autonomously work on their own data products via the self-serve platform.

- **Central platform IT team**: Indicate whether the organization has established a central platform IT team that bears responsibility over performing and maintaining the self-serve platform.

- **Domain representatives in governance team**: Indicate whether the organization has established a central governance team consisting of domain representatives, that administrates the policies and standards required for the performance of a Data Mesh.

- **Self-serve platform**: Indicate whether the organization has developed a self-serve platform, maintained by the central platform IT team, that enables the domain teams to work autonomously on their data products and collaborate with other domain teams.

- **Central data catalog**: Indicate whether the organization has developed a central data catalog which enables the sharing, discovering and accessing of trustworthy, secure and self-describing data products throughout the organization.

4.3 DESCRIPTION OF THE USE OF THE DMRM

The DMRM can be used by organizations to self-assess their Data Mesh readiness, and to identify improvement opportunities based on the results of the assessment. It therefore allows for organizational learning, aiming to facilitate faster decision-making regarding the presented improvement-areas of the participating organization. When using the DMRM, it could be beneficial for organizations to choose a group of diverse participants with different roles within the organization. Due to the broad scoping of the model, ranging from strategical and organization-wide dimensions to operational and technology implementation dimensions, readiness for a Data Mesh architecture could be assessed within all these different layers of the organization.

Due to the process and lengthy nature of a Data Mesh migration, it is recommended to use the assessment as a guideline throughout the entire process. Since it is unrealistic to assume that a Data Mesh migration can be completed within one iteration step, the step-by-step assessment model can be used to assess the progression of the different steps. It is also possible to use the assessment in later readiness steps to assess earlier steps, to improve or distort implementations that have already been done. In other words, it is recommended to take the assessment repeatedly during the Data Mesh migration process. In this way, the as-is state of the organization and the aspired to-be state can be assessed frequently and in a structured manner, so that the ambition levels that the organization is seeking to achieve can be determined time and again. In this way, after each assessment, the gap between the as-is state and the desired state of the organization can be analyzed, on the basis of which an improvement path can be formulated.

The DMRM should primarily be seen as an organizational guideline for structuring and organizing a multi-dimensional Data Mesh migration, rather than as a quantifiable assessment tool. The reason for this is twofold. First, a Data Mesh migration often depends on the specific organization in which it is performed, and it is therefore difficult to give a standardized quantitative rating to the degree of factors present. Secondly, it is beyond the aim of the assessment to present a one-off rating: after all, this number does not provide sufficient insight into the independent improvement areas and could lead to the risk of overseeing specific organizational challenges.

In order to strengthen the value of the DMRM, a set of standardized recommendations is required to generate an advice based on the outcome of the assessment. These standardized recommendations are determined based gathering best-practice approaches from participating participants in this study. The consensus that can be created on the basis of these best practices will be presented at the end of this research, after gathering all the necessary empirical data. These best-practice recommendations can be used by the users of the assessment to shape their improvement path in a specific sense. The set of recommendations for the assessment users, as gathered throughout this research, can be found in Appendix A.

5 DEMONSTRATION OF THE DMRM

In this chapter, the DMRM as designed and described in the previous chapters is demonstrated among various real-life demonstration settings. The conduction of the demonstrations is designed as a multiple-case study design, whose design and case-study design decisions are elaborated on in the beginning of this chapter. Subsequently, all demonstrations are presented independently of each other, after which the results of the demonstrations are presented in an overarching overview. Moreover, the comments, opinions and insights of the participants will be presented as additional findings. These comments are presented in italics between quotation marks. This chapter ends with an analysis on the complete set of demonstrations, in order to draw conclusions on both the performance of the DMRM within specific contexts, as well as the contextual influence on Data Mesh readiness.

5.1 DEMONSTRATION METHOD

The demonstrations of the DMRM in practice will be set up as a single-participant multiple-case study. First, the decision to choose a case study design will be elaborated on, and second, the decision for a single-participant design will be explained.

5.1.1 Case Study Design

A case study approach is chosen, since this method supports the development of the implementation model by assessing the practical value of a theoretical model. This practical value helps to make the generic theoretical model specific for Data Mesh readiness assessments. Moreover, case studies are relevant when there are exploratory grounds within the research topic where existing contemporary theories are not applicable (Flyvbjerg, 2006). Considering that this research is about a new organization of the data architecture design of organizations, there is a new topic that is still fully under development. This concerns a contemporary phenomenon, a research area in which, according to Yin (2017), it is customary and relevant to choose a case study research design.
Additionally, this research was carried out as part of the graduation within the CoSEM master's program. Typical research topics within CoSEM look for the possibility of designing a solution within a technologically and organizationally complex context. A case study is suitable for such an investigation because it leaves room for building a new theory by building on its application in practice.
During this research, case studies were conducted by collecting data through interviews and documents from various cases. Since data is collected from multiple sources in a multiple case-study design, the conclusions that can be drawn up from the study allow for the extension of the existing knowledge base on Data Mesh readiness (Yin, 2017). The model as designed in the previous chapters in this report will be demonstrated among these cases, in order to evaluate and iterate on the design of the model based on the collected data. In order to assure consistency among the interviews, a case study protocol was defined and followed. This case-study protocol can be found in Appendix E.

5.1.2 Single-Participant Case Study Design

The strategic selection of cases enhances the quality of case studies (Flyvbjerg, 2006). For most of the cases within this research, the case study was conducted with a single participant. The reason for this design decision lies within the fact that the the purpose of these case studies is not primarily to create generalizable conclusions, but rather to collect as many different attitudes towards the topic and perform a cross-analysis on these attitudes. Therefore, the case study results are likely to be limited in terms of generelizability, but offer a contribution in terms adding knowledge accumulation on Data Mesh concepts in a descriptive and phenomenological way (Flyvbjerg, 2006). Based on the aforementioned purpose, it was decided to extend the case study design through the conduction of interviews among a greater amount of different organizations and context, rather than a greater amount of internal perspectives within the participating organizations.

5.1.3 Case Selection

In order to strategically select cases for this research, several requirements for the case studies have been defined. These requirements are:

- **Size**: Since the complexity of migrating to a Data Mesh lies partly in the organizational challenges that it entails, organizations that are selected for the case study should have over 1,000 employees

- **Experience**: In order to be able to assess readiness for a Data Mesh architecture, it is required that the organization has some experience with implementing a Data Mesh, or else with gathering the background knowledge about Data Mesh architectures. Using the readiness assessment requires at least a basic knowledge of the concepts and the functioning of Data Mesh architectures. Organizations that have no conceptual understanding about Data Mesh architectures are less able to assess their own capabilities with respect to readiness for it.

- **Implementation**: In order to address the entire scope of implementation levels, the case selection should entail a variety in current Data Mesh implementation levels. In order words, the organizations that are selected must vary in the stage of Data Mesh implementation they're at currently, varying from not having started the implementation up until having finalized a complete Data Mesh implementation.

- **Role**: The availability of interviewees affects the possibility of conducting an interview at an organization. Selecting an interviewee within a particular organization is done based on the role they have within the organization. Due to the technological and data-related characteristics of Data Mesh architectures, interviewees that have common knowledge of the current state of the data architecture of the organization are selected.

According to Yin (2017), each case must be carefully selected so that the individual case either predict similar results or predict contrasting results but for anticipated reasons. For this case study, organizations were selected from different industries.

Based on the requirements, six companies have been selected for the case study that fulfill all requirements. In this thesis, they are referred to as Organization N, Organization O, Organization P, Organization Q, Organization R, Organization S and Organization T. Table 5.1 an overview is given of the selected organizations and their interviewees.

Table 5.1: Organizations selected for case studies

Case	Interviewee	Operating group	Number of employees	Data Mesh implementation experience
1	Interviewee N	Information Technology	1,500	Yes
2	Interviewee O	Public Services	7,000	Yes
3	Interviewee P	Financial Services	20,000	Yes
4	Interviewee Q1 & Q2	Information Technology	1,000	As consultants
5	Interviewee R1 & R2	Technology Consulting	4,000	As consultants
6	Interviewee S	Information Technology	1,000	No
7	Interviewee T	Logistics	7,500	Yes

5.1.4 Quality of the Case Studies

The quality of this case study design and the empirical evidence that can be generated from it, can be assessed based on reliability and on three different types of validity (Riege, 2003).

Construct validity ensures that the concepts that were studies are correctly investigated (Ferreira, Andrade, & Almeida, 2020). By means of enhancing the construct validity of this research, the conducted interviews were transcribed and coded in order to be able to draw ex-post conclusions. These verbatim interviews allow for the cross-case analysis of particular quotations and specific perspectives among participants, in order to enhance the quality of the empirical base of evidence (Griggs, 1987). Moreover, multiple sources of evidence were gathered throughout the process, since the case studies were conducted among many different contexts based on the selection criteria (Riege, 2003).

Internal validity ensures that the conclusions that were drawn up from the obtained data are adequate (Ferreira et al., 2020). The internal validity of this research was enhanced by discussing the case results with the participants after the interviews in order to check for this adequacy. Moreover, the results among all cases were cross-checked after the conduction of all case studies in order to assure internal coherence of all findings (Yin, 1994).

External validity ensures that the accumulated findings from the case studies can be used for the representation of the studied phenomenon as a whole (Ferreira et al., 2020). This external validity has been improved by comparing the findings of the case studies with the other findings as found earlier in the study. In this way, the results of the case studies can be used to strengthen the overall scientific and empirical contribution to the body of knowledge within the stated scope of the study (Klein & Myers, 1999).

Lastly, **reliability** deals with the probability that another researcher would arrive at similar results with the same research (Ferreira et al., 2020). The reliability of this research is enhanced in two ways. First, a set-up of semi-structured interviews guided by an established case study protocol was used (Yin, 1994). Second, the case study interviews were recorded during their execution, which also contributes to the reliability of the study (Nair & Riege, 1995).

5.1.5 Analysis of the Case Studies

After conducting all case studies in this research, the collected data is analyzed. The data analysis will take place in several steps, which are shown below.

First, all participants were asked general questions to get an idea of their experience with Data Mesh concepts and their organizations' current level of implementation. After these questions, the research model was presented to the participants, so that the structure of the model and the factor dimensions could be seen. First, the participant was asked whether the model and readiness steps seemed correct and complete. Then all factors in the model were run through. For each factor, it was assessed whether the factor was already applied in the organization and what the perceived importance of this factor was. The reason for these indications was also noted. Finally, the participant had the opportunity to share comments about the model with the researcher.

Secondly, the case study reports and assessment results of the participants were elaborated and compared with each other, so that the similarities and differences under the existing perspectives can be found. This led to a summarized overview of all results gathered during the interviews.

Thirdly, this overview was analysed. Because the results also included the perceived degree of importance of all participants, as well as the level of practical implications of these factors, this overview contained several insights and perspectives. The interesting or contributing insights and perspectives are then processed as an exact quote in the research report.

5.2 DESCRIPTION OF THE DEMONSTRATIONS

This section presents the conducted case studies, by means of a detailed description of the DMRM demonstration within that particular case. The cases are described by their industry sector, their number of employees, and their Data Mesh experience. The last term distinguishes cases based on whether or not they have experience in setting up a migration path towards a Data Mesh architecture in the organization.

5.2.1 Case 1: Interviewee N

Information Technology — 1,500 employees — Data Mesh experience

Assessing the need for a Data Mesh

The need for a Data Mesh arose from the fact that there were technical and organizational issues within the organization that needed to be resolved. There was a challenge that multiple domains existed within the organization, but they were all struggling because they had different definitions of the term 'person'. This resulted in the fact that the domains were unable to work interoperably with each other's data. From a technical perspective, one would define a person as a customer, while the other would define a person as an individual. These definitions differed across the domains, so it was unclear in the interaction between different systems. In addition, they had also been a company that had grown through acquisition and had several disparate groups that had implemented things independently for a long time. So they all had individual silos that could hardly cross-collaborate with each other. They had a central data team that became a major bottleneck in this situation, not necessarily from a technical perspective, but especially from an organizational perspective that team became a blocker to getting things done. The central data team had more organizational power than they should have, which meant that decision-making fell short.

Assessing the ability to migrate to a Data Mesh

One of the most important factors of this phase was the ability to think about data as products, not projects. If this ability does not exist, there is no point in migrating to a Data Mesh. This organization already had strong product thinking about data, which made that mental leap very straightforward. As a result, only some technical training was required for the current product owners. In addition, the ability to decentralize the current monolithic organization was very decisive here. This organization already thought in terms of domains, because they maintained a domain map of their organization. This map helped them to set boundaries in terms of ownership, responsibility, and the groundwork for what individual teams do decide on and what they do not decide on, so that the different domains would not step on each other's toes. In addition, the influence of top management engagement has greatly influenced this organization. The project came to a standstill for a while when a new CTO was hired who was not supportive of the migration, due to different motivations and directions. As a result, they came to a standstill in the early implementation phase. A final factor that was important within this organization, but which is not reflected in the model, is the scale of the organization. Some organizations are simply too small for a Data Mesh, and would be better suited by sticking with a monolithic model. Only when the complexity of the problem the organization is trying to solve becomes too great for a monolith, it makes sense to think about a Data Mesh.

Assessing the preparedness to migrate to a Data Mesh

As for the preparedness of this organization, they had a lot to do with setting up standards and guidelines. Especially since it was often forgotten that when changes are made upstream, sometimes downstream and cross-domain use cases can be broken. With regard to the preparedness for a Data Mesh, it was very important for this organization to set up standards and guidelines that structured this. One of the examples of this was the CI/CD standards as a mechanism to do this. In addition, the data product publication standards have proven to be very important in the sense of not only having an API for the structure of the data, but also about how the data product itself may evolve over time. It turned out to be important to model that in the contract of the data product itself, so that the data product becomes trustworthy for other domains to use.

Determining the readiness for a Data Mesh

As for the complete set of Data Mesh building blocks, this has not yet been fully implemented in the organization. The reason for this is mainly that it is still very chaotic from an organizational point of view and the governance is not yet fully aligned across the various domains. A step in the right direction has proved to be not to keep the federalized governance model completely federalized, but to continue to have a centralized governance team that can enforce the new rules and guidelines across the organization as a whole. Finally, it has become apparent within this organization that a Data Mesh can certainly address as number of their needs, especially with respect to their organizational challenges, but that its implementation is a lengthy process.

5.2.2 Case 2: Interviewee O

Public Services — 7,000 employees — Data Mesh experience

Assessing the need for Data Mesh

The need for a Data Mesh architecture for this organization was mainly due to the excessive independence of the various data-generating domains. The domains worked too independently of each other, and there was no awareness about how to prepare own data for use by other domains. This was mainly due to the fact that they did not think in the context of processes, but mainly continued to think in their own business domain. The groups within the organization did not talk to each other, nor did they think about the consumers of their own data. As a result, the consumers of the data were too dependent on the source domains and therefore not autonomous enough.

Assessing the ability to migrate to a Data Mesh

The ability to migrate to a Data Mesh at this organization came from a correct business and IT alignment. These two were initially not aligned, in the sense that the business was not aware of the data required for certain objectives, and the data teams did not know what business value was in their data. It took a while before there was a realization within this organization that consensus about the data was needed to extract more value from the operations. They therefore started mandating more data awareness and data literacy across the entire organization, so that everyone understood the basic principles of data. In addition, the ability to migrate also turned out to lie to a large extent in enabling the domain-people to become more data aware, in order to avoid the problem of having too few data engineers. In addition, decentralizing the organization was a big task. Because it proved difficult to find the right level of abstraction from a business perspective, it turned out to be more important to think from service design and processes than from purely business oriented domains. Finally, the top management engagement within this organization played a major role, as they were very engaged with the migration from the start and therefore provided the resources to make it happen.

Assessing the preparedness to migrate to a Data Mesh

The preparedness for a Data Mesh for this organization came mainly from the construction of a data infrastructure that made it easy for data engineers to make a good product for their consumers. Preparedness was therefore not necessarily measured by implementing the correct CI/CD standards or automation guidelines, but more by mapping out the ideas of a more consumer-oriented way of producing data. There was a fear within this organization that if too many technical standards and operations were implemented too quickly, it would deter the necessary way of working within a Data Mesh. That is why they started small and only implemented the necessary infrastructure and thinking that brought them in the direction of the Data Mesh way of working.

Determining the readiness for a Data Mesh

Almost an entire Data Mesh has been implemented within this organization. It turned out to be an iterative and time-consuming process, which had to be approached in a very structured and precise manner. In addition, there are still ongoing measurements of certain KPIs such as data quality checks and other efficiency measurements, to determine whether the new data architecture meets the objectives and needs. Within this organization it turned out to be important to consider the influencing factors for the migration to a Data Mesh at different importance levels, so that not all prerequisites had to be in place in order to be able to get started.

5.2.3 Case 3: Interviewee P

Financial Services — 20,000 employees — Data Mesh experience

Assessing the need for Data Mesh

The biggest need this organization had was to improve the quality and trustworthiness of the data. IN addition, the organization had to deal with a large overload of the data engineering teams. As a result, these overloaded teams gained a motivation to migrate to a Data Mesh architecture. Finally, there was a desire to become more data-driven, which also brought the business teams on board for the change.

Assessing the ability to migrate to a Data Mesh

With regard to the ability to change, it was important for this organization to formulate clear objectives that matched the needs of the organization. This helped with getting the rest of the organization on board. In line with this, they worked on a clear value statement and a Data Mesh Proof of Concept, with which they could prove the value of the Data Mesh architecture. With this proof of concept it turned out to be important not only to focus on the technology, but especially on the organizational shift that would come with it. Finally, the existence of a learning culture was of crucial value. The organization has given the people time to learn and evolve over a longer and iterative period.

Assessing the preparedness to migrate to a Data Mesh

With regard to preparedness, this organization attached great importance to designing a suitable governance model. In this, they strongly believed in the power of automation to make tasks easier for the domains, but also for the central governance team. It also turned out to be important to think carefully about the way in which you set up the domain teams, so that the responsibilities are correctly distributed. Example roles included a data owner on the business side, data custodian on the engineering side, and a data steward in the middle, with the data steward acting as the one who understands the data, but also acting as a liaison outside the domain.

Determining the readiness for a Data Mesh

One of the key components of the Data Mesh within this organization was the self-serve platform. This platform should serve the majority of the organization in a proper way, and therefore would initially be built very generically to be able to be used by the majority of users within the organization. The platform should also provide good guidance on transforming and automating the data and its processing, so that configuration within the domains could be made easier. The performance indicators of this platform were mainly within an acceleration of the deployment of data services. Finally, the aim of the platform was to gain a better picture of who consumes data, and why, through data virtualization and federation.

5.2.4 Case 4: Interviewee Q1 & Interviewee Q2

Information Technology — 1,000 employees — Data Mesh consultants

Assessing the need for Data Mesh

Organizations that are guided by this interviewee often have to deal with a large number of domains with a lot of complexity, which causes problems with regard to the scalability of the organization. In addition, during problem identification, there is often a great need to become more data-driven, so that decisions can be based on data. It also emerged from the case study that it is not always about assessing the specific needs of a company during problem identification, but that the value that organizations want to get from a Data Mesh often also plays a major role in this. According to this interviewee, the desire to enable decision-making on the basis of data is the main motivation.

Assessing the ability to migrate to a Data Mesh

First, it was important for these organizations that the value that the Data Mesh will bring is clear and understandable for everyone at all levels of the organization. This made it possible to introduce the data-driven culture top-down. This interviewee also indicated that the importance of proper education within organizations is very important. Finally, it has been found that organizations that have already gone through a major data-driven migration, such as a cloud transition or a DevOps transition, are in a better position compared to data mesh migration than organizations that have not yet experienced it.

Assessing the preparedness to migrate to a Data Mesh

When the preparedness for a Data Mesh migration is assessed within organizations, it is important that the new roles and responsibilities become apparent, according to this interviewee. This was done by setting up global contracts defining these responsibilities. The priorities were laid down therein in consultation with the domains, so that everyone was aware of them. The bounded context concepts, such as those derived from previous distributed architecture designs, were strongly taken into account.

Determining the readiness for a Data Mesh

This case study has shown that there is a difference between assessing readiness for an organization and preparing the organization for a Data Mesh architecture. Readiness mainly has to do with meeting the right organizational and cultural requirements that make such a migration possible, where the preparation goes more into a migration to a Data Mesh specifically. Where these two pillars can be combined in one model as is done in this study, it is important to clarify prior to the assessment what Data Mesh-specific knowledge of the assessor is required to be able to assess the concepts. In addition, it is important to know what stage they are in with regard to Data Mesh migration of the organization being assessed.

5.2.5 Case 5: Interviewee R1 & Interviewee R2

Technology Consulting — 4,000 employees — Data Mesh consultants

Assessing the need for Data Mesh

The organization that was guided by this interviewee mainly had to deal with domain teams that were not independent enough. As a result, a need arose for a new way of data sharing that puts the responsibility in the domains themselves, so that they could perform their own analyses on their own data. Until now, this responsibility mainly rested with the central data team. In addition, there was a need for the ability to have a platform infrastructure that was customized to their own needs. Finally, there was a very centralized data strategy, which was actually good. The domains needed to define their own data strategy, but they still have to adhere to the central strategy of the organization. Finally, the data engineering teams lacked business context. For example, they were tasked with performing data transformations for the domain teams, but they had no idea what would happen to that output.

Assessing the ability to migrate to a Data Mesh

With regard to the desire to transition, at the organization in question this mainly came from the domains. This also turned out to be important, because these domains also have to take responsibility for the data. In addition, it turned out that there were too few skills within some domains to create and maintain data products. It was therefore very important that there was an overarching learning culture, in which there was room for upskilling certain people in the domains. This upskilling has proven to be important when making the transition to the Data Mesh. In addition, it was difficult to make a clear value statement within this organization. In terms of returns on investment, it proved difficult to monetize the performance of the Data Mesh. Therefore, an attempt has been made to express it in the downward trend in lead times, and the speed at which data products were generated. Because the drive within the organization was to become more data-driven, these returns on investment were more important than the monetary values that come out of it. Finally, in terms of the decentralization process, many of the domains already existed. However, the problem was that these domains lacked autonomy from an IT perspective, so these domains also needed the resources to hire people who had the necessary skill set. had. It turned out to be important to see per domain whether they had the autonomy for a Data Mesh from an IT perspective.

Assessing the preparedness to migrate to a Data Mesh

The next phase of the migration to a Data Mesh was mainly characterized in this organization by defining the new roles and responsibilities within the organization and the domains specific. In addition, it has also proved important to properly define the responsibility of the central platform team, because it has to provide certain components of the Data Mesh to the domains. Within the domains there must be clear responsibilities, such as that of a data product owner. In addition, there had to be clear ownership of the data products that the domains publish, in order to make cross-domain collaboration possible. With regard to governance, it turned out that this federalized governance model has to be built on the organization in question. It is important to enforce certain components, such as the privacy and security of data products, from a central team. The way in which the domains themselves came to their data products and how they collaborated was left to the domains themselves to regulate. It turned out to be a challenge, from a technical perspective, to deliver the components of the infrastructure to the different teams so that they could actually adhere to the central policies.

Determining the readiness for a Data Mesh

This case study has shown that it can be important for organizations preparing themselves for a Data Mesh migration to ensure the required skills among its employees. These skills can be improved by a well-established learning culture. Moreover, it turned out to be of positive influence on Data Mesh readiness to organize the specific roles and responsibilities within the future domain teams in a structured manner.

5.2.6 Case 6: Interviewee S

Information Technology — 1,000 employees — No Data Mesh experience

Assessing the need for Data Mesh

The organization in this case study had no experience with implementing a Data Mesh architecture. During the assessments of a need for a Data Mesh architecture, it emerged that an organization must have a lot of experience with processing and using data to be able to see the added value of a Data Mesh. In addition, this interviewee found it difficult to determine the essential difference between Data Mesh architectures and other distributed architectures. The interviewee could imagine that there is a desire to become a data-driven organization and that this can be filled with, for example, a migration to a Data Mesh architecture.

Assessing the ability to migrate to a Data Mesh

As for the ability to migrate to a Data Mesh architecture, this interviewee mainly foresaw problems with splitting the monolith into decentralized domains. In addition, with regard to the decoupling of existing applications, an issue arose around the connection of application with the Data Mesh architecture. The interviewee stated that a sophisticated and mature data virtualization solution should be implemented that can ensure that data can be shared across the different domains without the need to copy the data. Finally, this interviewee emphasized the importance of education and training to introduce the new way of working across the organization as a whole.

Assessing the preparedness to migrate to a Data Mesh

Because this organization had absolutely no experience with implementing Data Mesh solutions, assessing the preparedness for a Data Mesh turned out to be premature. The interviewee agreed upon the factors as reflected in the model, although doubts were raised about the federalized governance model. According to the interviewee, this model would only be possible within the organization if there was also a central governance body that could enforce the global rules and guidelines across the various domains. In addition, this interviewee foresaw that CI/CD solutions and degrees of automation were not necessary for performing a Data Mesh, but could be better identified as extra features that would make the way of working more efficient.

Determining the readiness for a Data Mesh

With regard to determining the readiness for a Data Mesh architecture, this organization was still at an early stage. It has become clear from this case study that with organizations at such an early, or perhaps even non-existent, stage it is difficult to go through the entire chain of readiness steps as shown in the model. It would therefore be better according to this case study to use the model iteratively during the process, in order to identify areas for improvement and to gain insights about the progression of the migration. Because migrating to a Data Mesh architecture mainly requires organizational efforts and is therefore a long-term project, an iterative use of the readiness model could provide structure and guidance during this process.

5.2.7 Case 7: Interviewee T

Logistics — 7,500 employees — Data Mesh experience

Assessing the need for Data Mesh

Within this organization there were two main needs that led to the implementation of a Data Mesh architecture. Firstly, there was too little ownership and understanding of their data. It was unclear what the data meant and how engineers and analysts could collect certain data. As a result, there was little confidence in the data. Secondly, it was difficult to make changes to the existing data platform. You couldn't simply add a new data set or integrate two different data sets because it took way too much time and effort. Finally, within this organization there was not a lack of a data strategy, but a need to attain a data strategy that suited the organization.

Assessing the ability to migrate to a Data Mesh

A major challenge for this organization was translating and advocating the change across the organization as a whole. It turned out to be very important to leverage the right people who had the right influence on the organization. In addition, it appeared that the top management was very engaged and understood the value of the Data Mesh, which allowed them to encourage the change from the executives. This increased the desire to transition within the organization as a whole. In addition, they had large R&D teams, so that the value of innovation was also strongly recognized in the organization. One of the other challenges for this organization was the right Business & IT alignment. It was difficult to move away from the centralized team. The decentralized approach required a new perspective on the central team, in the sense that they were seen as enablers for executing the tasks within the Data Mesh architecture, rather than as executors of the tasks themselves.

Assessing the preparedness to migrate to a Data Mesh

In terms of preparedness, this organization has developed a program that would redesign the entire organization, processes, in and around the domains and the teams, involving the entire business. This program was run in tandem with the entire organization, allowing it to act as a foundation for the migration process. This program also included the new rules and guidelines, and the standardized Data Mesh-specific processes. With regard to governance, its implementation was a lengthy process, which in particular showed that enabling interoperability and collaboration was very important. In addition, standards for publishing data products have also proven to be very important, so that the reliability of the data was improved and people did not have to reinvent the wheel all the time.

Determining the readiness for a Data Mesh

With regard to the Data Mesh readiness, the entirety of building blocks mainly worked together if you can do it as an organization as a community. Standards and guidelines are needed for this, and time had to be invested in supervising and structuring the migration. They started with providing only the initial domains guidelines with standards, after which they looked at how the domains would grow in the new way of working. Good use of structures and guidelines helped in this regard.

5.3 RESULTS OF THE CASE STUDIES

Table 5.2 presents the results of the assessment as performed at the case study organizations.

Table 5.2: Assessment results of the conducted case studies

Case	Readiness	Assessment Results
1	Almost ready	**Current readiness phase**: 3 - Assessing the preparedness for migration **Proposed improvement area**: the biggest challenges currently still being within the institutional implementations, i.e. the right guidelines and governance strategies require the most attention so far. **Further recommendation**: design a governance model that fits the structure of the organization and implement it with the participation of all employees.
2	Ready	**Current readiness phase**: 4 - Determining readiness **Proposed improvement area**: identifying teething problems and detailed structures within the institutional setting. **Further recommendation**: the assessment shows that certain measurements are still needed in the coming years to determine whether the new architecture can meet its objectives. These measurements can be implemented by means of automated quality checks and other efficiency checks.
3	Ready	**Current readiness phas**e: 4 - Determining readiness **Proposed improvement area**: identifying the performance of the Data Mesh architecture within the cultural setting. **Further recommendation**: recommendations mainly lie in clarifying the division of roles within the teams and maintaining the strong learning culture and knowledge programs within the organization.
4	Almost ready	**Current readiness phase**: 3 - Assessing the preparedness for migration **Proposed improvement area**: steps can be taken with regard to the contractual determination of interdependencies and collaborations between domains. **Further recommendation**: to use the assessment iteratively throughout the processes with their customers, because the migration to a Data Mesh architecture is a chaotic and lengthy process that could use the guidance.
5	Almost ready	**Current readiness phase**: 3 - Assessing the preparedness for migration **Proposed improvement area**: the results showed the importance of improving the practices of clearly dividing roles and responsibilities, and establishing a learning culture. **Further recommendation**: a recommendation based on the assessment would be to iteratively improve the governance model based on its performance in practice. It is recommended to set up collaboration guidelines and interoperability policies that can formally guide this process.

6	Not ready	**Current readiness phase**: 2 - Assessing the ability to migrate **Proposed improvement area**: a number of needs could be identified that specifically point to a Data Mesh, but also a number of generic needs that had to do with the desire to become more data-driven. **Further recommendation**: a recommendation based on the assessment would be to start by formulating clear problems and objectives, and based on that make a choice whether or not to migrate to a Data Mesh. If a migration option is chosen, it is recommended to start with the second phase - e.g. the Ability to migrate - and carefully determine how the current monolith can be decomposed. The assessment can then provide guidance throughout the process for determining the next steps in the migration process.
7	Ready	**Current readiness phase**: 4 - Determining readiness **Proposed improvement area**: the results showed the importance of formal policies and guidelines in the coherent functioning of the architecture. The points for improvement that emerged from the assessment mainly concerned improving the governance model on the basis of improved interoperability between the various domains and the structures around them. In addition, the importance of taking a step-by-step approach to a Data Mesh migration emerged, because otherwise there is a risk of losing strength within the organization. **Further recommendation**: it can be recommended to complete the assessment again after a few months, when more information is available about the performance of the Data Mesh architecture. Based on this, new improvement areas can be identified.

5.4 INSIGHTS FROM THE CASE STUDIES

The case studies started with general questions about Data Mesh architectures. These questions were intended to find out to what extent the interviewee has encountered a Data Mesh implementation, as well as common barriers for organizations to start with a Data Mesh migration. In addition, these questions allowed to determine the interviewee's perspective on Data Mesh architectures.

5.4.1 Data Mesh in practice

Nearly all interviewees of the case studies indicated that Data Mesh architectures are still very rare in practice. This may be because the organizations are not yet ready for it, or because they do not know how to get started.

> *"I have been working on a Data Mesh implementation model. I think I might have one that will work, but I am also finding that a lot of organizations think they're not quite ready yet."* (Interviewee N)

> *"I'd say that there is not a standard way of implementing a Data Mesh in organizations. I try to look at all the principles and see what fits best for the organization."* (Interviewee R)

5.4.2 Insights about assessing the need

Assessing the need

With respect to assessing the needs for a Data Mesh architecture, the case studies have shown that many of these needs are interrelated. The need that emerges most often and can therefore be considered important is the **need for greater autonomy for domain teams**. However, the autonomy is related to the **need for more upstream responsibility**, because ultimately there is an overarching need within organizations for better understandability, discoverability, trustworthiness and quality of the data to be consumed. Organizations want to work more data-driven, and therefore be able to base their decision-making on data that is reliable for this. Interviewee X underlines this new way of working:

> *"In terms of whether there's the need for domain teams to be able to do their own data analysis, I definitely think that this is important. I think this is mostly not the case in companies that go into Data Mesh. Domain teams are not independent enough. Domains need to be able to do their own analysis and make their own data products."* (Interviewee X)

Interviewee T also indicates that it is important to be able to trust the data:

> *"The lack of ownership in particular gives bigger problems because people just don't trust the data. They don't trust the data that is given to them."* (Interviewee T)

Closely aligned to this, a more practical need is the **need for shorter lead times from operational teams to data analytics teams**. During the case studies, this was mostly identified as the 'pain felt by the data teams'.

> *"I think this is really closely related to the need for greater scalability. And why is there a need for greater scalability? Because there is a bottleneck in the centralized data team. They are not able to fulfill the requests that they are getting, so the lead times are too long. So for sure, to me, this is one of the most important needs."* (Interviewee R)

Of the other identified needs, a few emerged during the case studies. In addition, it appeared that it is important to align the interpretation of the needs with the assessment:

> *"For us, it was more the case that we had too much autonomy for the teams. They were too independent of each other, and they didn't think of the other when determining the data they needed. There was no awareness of how to prepare this data so that the other teams could use it in an efficient way."* (Interviewee O)

When assessing the **need for greater flexibility of the data platform** and the **need for greater scalability of the data platform**, this is identified as something that is needed among organizations.

> *"They also had been a company that had grown through acquisition and had several disparate groups that had implemented independently for a long time. They had these individual silos that didn't cross talk very well. They were in the situation where you end up with siloed information, siloed activities and duplication of efforts. It is very hard to make changes to these kind of systems."* (Interviewee N)

> *"I think this builds up on beneath the need for greater autonomy, the more flexible the data platform, the less lead time there is, and the easier it is to consume the data products. Data*

engineers need a data infrastructure that meets their needs, so it is easier for them to make good data products for their consumers." (Interviewee O)

"Yes, the platform needs to be flexible and that is what most platforms are not currently. However, within the Data Mesh, there also has to be standards, which shouldn't be changed. So this is a tough one to assess." (Interviewee R)

Finally, the **need for a data strategy** and the **need for reduced cognitive load for data engineering teams** are two needs that are more difficult to find consensus on among the interviewees. In the first case, this is mainly because there is a data strategy within organizations, but that strategy is often not necessarily the best strategy for that specific organization. This can lead to inefficiencies in the business process or decision-making, which may lead to a need for a more structured approach:

"I think there is this need for a data strategy that fits the organization. The company fails to identify what is the right strategy for itself. And when they actually do have a strategy, they need to actually implement it. I think this operationalization is what is lacking." (Interviewee T)

In addition, some case studies showed that there were not necessarily too few data engineers in the organization who have too high a cognitive load, but that these teams have too little knowledge of the purpose of the data they process. Although there are indeed organizations that suffer from a large overload in the data engineering teams, part of the problem is addressed by giving the engineering teams more business context:

"I think there will never be enough data engineers when you keep on not sharing the domain knowledge. And I think you need to enable data engineers to understand just so much from the domain that they can make wise decisions. So this need is clearly addressed by a Data Mesh architecture. You need the cross-functional teams. I come from the semantic data modeling team, and we try to put semantics or be very careful for how we call column names or tables. Because it's a way of communication. By thinking carefully about how you call things, you can reduce the cognitive load for those who need the data." (Interviewee O)

"Sometimes data engineering teams, they don't have business context at all. They're still required to do data transformation of domain teams. They make many efforts to maintain and transform data that they don't know of, and of which they don't know any business context. So yes, this is a big driver for a Data Mesh." (Interviewee R)

Concluding, it seems that most of the needs that point towards a Data Mesh architecture as a solution are intertwined. During the assessments, it was hard to tick all the boxes, since the problems felt by the organization mostly point towards the same direction. The model was used for targeting specific parts of the organization by encouraging the organization to think closely about the exact location of its experienced bottlenecks. Moreover, it was concluded from the case studies that the needs in the model should be ranked by importance, in order to guide the organization towards the Problem Identification and the main problems encountered.

5.4.3 Insights about assessing the ability to change

Willingness to Change

Starting with the **desire to transition**, it seemed important to identify where this desire comes from within the organization. From the demonstration of the readiness assessment, it turned out that it is important to find this desire throughout the entire organization, so bottom-up as well as top-down. The desire can also be motivated, for example by introducing organization-wide programs and workshops. According to interviewee R, it is most important to have this desire to transition coming from the domains:

> *"To me, it is important to identify where this desire comes from. It can come from a centralized team that decides that a Data Mesh is the way to go, or it can come from the domains itself, demanding a greater autonomy. I think, the main important thing is that it comes from the domains. The domains need to take this responsibility."* (Interviewee R)

Interviewee Z, on the other hand, believed that this desire to transition was best spread from the executive level:

> *"We started talking with the executives, in order for them to understand the concept. When they started understanding why you would need a Data Mesh and data products, they spread the word to other people in the organization, both publicly via videos and announcements, as during their meetings. The ball started rolling, and the desire to transition became bigger and bigger."* (Interviewee T)

Moreover, the existence of a **learning culture** and an **innovation culture** was discussed during the case studies. Although these terms were often used interchangeably in the case studies, the importance of education and freedom of experimentation was often emphasized. This can be achieved, for example, by established R&D teams, knowledge sharing platforms, and having executives who push innovation into the company. It turned out to be important to realize that many people in the organization had to be upskilled for the migration, and that this would take time and effort.

> *"I would assess learning culture in terms of upskilling the domain teams. What you see sometimes is that these domains, they don't actually have the skills to create and maintain data products. And that is actually a big deal. So for me, learning culture, and desire to transition are really close to each other. If you want to learn, you will want to transition into something where you have greater autonomy."* (Interviewee R)

> *"I think one of the first things that I've observed is that they need to start thinking about data as products. It's an innovative approach towards what organizations are doing right now, but it's crucial for a Data Mesh to function."* (Interviewee N)

> *"We introduce data through culture, through education. You have to educate people, via workshops or lectures. You also have to encourage people to be willing to learn and invest time in it."* (Interviewee R)

Vision

When discussing the factors **clearly defined objectives** and **clearly defined value statements** during the case studies, it turned out that most organizations have their main objective for a Data Mesh architecture formulated as *a desire to become more data-driven*. However, it is important to define these objectives and values in a concrete and executable manner, in order to achieve objectives from different points of views within the organization. Eventually, the overarching objective is mostly to take better decisions based on insights from the data, but due to the lengthy process associated with a Data Mesh migration, it is important to set concrete goals from all corners of the organization that are also feasible.

"At the end of the day, everybody wants to become data-driven. And everybody wants to make decisions based on facts. But the problem is that that is the objective of any data strategy, not only the one for a Data Mesh. What we really need to do is to see what we are trying to achieve from different point of views. In terms of high level and low level objectives, and strategical and tactical objectives, which should happen over time." (Interviewee T)

"You should define the objectives based on the needs. These needs define the end goal that you try to move forward to." (Interviewee R)

"In my opinion, it is what leads people to decide to go for a Data Mesh, because they understand the value of it for their data processing." (Interviewee O)

Business Case

Concerning the **business & IT alignment**, the problem was often experienced that the business side and the IT side of the organization were very far apart in terms of goals, means and data awareness.

"In my company, it wasn't very much aligned. The business side and the IT side both really needed to understand that we're now going to deal with data as a product. That really took a while. The organization gradually understood that they had to make teams mandate more data awareness, in order for the Data Mesh to work. There is this growing awareness, but you have to understand the basic principles of data." (Interviewee O)

"We've made a concerted effort to ensure that if this alignment isn't already there, we work to demonstrate the value of the project itself. This can actually come through metrics. We found that key performance indicators are a good way to influence change in business. Alignment often boils down to how you are measuring your success." (Interviewee N)

According to the case studies, this gap can be filled in various ways, for example through value demonstrations, workshops or spreading the word. The case studies of Interviewee P and Interviewee R even showed that assigning authority to different domains automatically narrows this gap.

"There's always this problem with business and IT being on different pages. Technology is worried about stuff like security, which is why they want all others to access their data. Business on the other sight wants to access the data, but are receiving the wrong data. You just need everybody from the entire organization to think about it, so everybody is organized around domains. These domains will help gain alignment between business and IT." (Interviewee P)

"To me, it depends on the organization. I've seen companies in which architects started to build some enormous database, in which they evangelize everything they need, thinking that everybody wants to move towards this. But then the business is not even really aware that this is happening. The other way around, you have the domains claiming that they want to have more responsibility, whereas they're not given any since they don't have that authority. So this is something that is really important." (Interviewee R)

Combining the different views from the different case studies, it seems that the business & IT alignment is something that needs to be put in place in order to be ready for a Data Mesh. Especially within large organizations, it is important that the business side of the organization understands the value of data. There are various strategies and ways to try to align the business and IT sides of organizations, but the common thought is that they need to be aligned and that that challenge needs to be thought through very carefully.

Leadership

According to all case study interviewees, **top-management engagement** is a prerequisite for Data Mesh readiness. Since all interviewees were able to indicate almost immediately whether this engagement was there or not, and then indicated that it was very important to have top management engagement, there has not been much discussion about this factor.

> *"It is all about whether your business leaders, as in your executives and non-technical people, understand it. If they're not excited about it, then you haven't completed the first readiness state properly. If they really understand the need, the transition can succeed."* (Interviewee N)

> *"The players on both sides need to be on board. For example the head of the domains, in more of a business capability. So the head of strategy, or people working really closely to strategy and want to become more data driven, they need to understand why data is a benefit for them to become more data driven. But on the other side, you'll also need the CTOs, they need to buy into this whole contract."* (Interviewee R)

> *"For us, it was a bottom-up approach, because the pain was felt on the bottom. I think that top-management is too far away from everyday life to really feel it. But we had top-management who were really good listeners, and they understood our problems. Those words have to merge."* (Interviewee O)

Enterprise Ability

One of the more complex aspects of Data Mesh readiness is assessing the enterprise's **ability to decentralize the business into domains**, and **the ability to decouple the existing business applications**. All in all, it turned out that it is a very iterative process, and that you can initially start from the existing business domains. However, there will often still be a shift in the boundaries of the domains and the precise tasks and responsibilities between the domains. In addition, it can be difficult to decouple the existing applications in a decentralized way. It can come in handy to use a good data virtualization mechanism, to enable the reuse of data sets from domains. In addition, the central platform and the central IT team play a major role in decentralization and decoupling, because they must be able to enable the interoperability of the various components within the Data Mesh. All in all, many different approaches have emerged from the case studies to tackle these challenges, but it has been identified by all case studies as a very important step in determining the readiness for a Data Mesh architecture.

Some case studies indicated that it was very difficult to divide the business up in domains:

> *"Dividing your business into domains itself is difficult, especially in defining where to put the right level of abstraction. I think the processes are always much more important than the domains. I have a tendency to think more in service design and processes, rather than domains. To me, the right way of organizing would be in terms of processes."* (Interviewee O)

Others suggested that the domains might even already exist within the company:

> *"I think sometimes, the domains exist already. And so I think companies have the ability to create these domains. However, especially traditional companies, have the problem that IT has been centralized. So traditional IT is very centralized, and is not catered in decentralizing in the domain-driven design way. I think the business is already more decentralized in this*

> *manner, but IT has traditionally been centralized. To me, that's where the challenges lie."* (Interviewee S)

This interviewee used existing principles from the domain-driven design approach to structure the process of decentralization:

> *"We realized that the technical architecture should follow the business architecture. The organization has domains, they are the natural implementation of the business architecture. The point is that we wish that the technical implementation should follow this business architecture. We used the idea of bounded context from the domain-driven design principles as major design aspect of domain distribution. It should some kind of prerequisite for Data Mesh to have some idea on the decentralization of your business."* (Interviewee Q)

With respect to the decoupling of business applications, some case studies pointed out that this decoupling process is a prerequisite for the performance of the Data Mesh:

> *"We basically want to provide self-serve infrastructure to the different domains. So these domains will have their own resources in the cloud, to make their own data products. And these data products need to be completely decoupled, because they belong to that domain. There will be some sort of standard interface where other domains can consume these data products. It is very important that these applications can actually be decoupled, but also that they can be consumed in a standard way."* (Interviewee R)

While others believe that there are alternative ways to address this challenge:

> *"If you provide the right data virtualization and a platform that enables the accessing and sharing of data without having to make copies, completely decoupling applications is not necessary to achieve the correct cross-collaboration of domains."* (Interviewee S)

Lastly, there needs to be an **ability to work in domain teams**, especially with regards to the people that are going to work in the domains.

> *"In the past, these people were consumers of data. They were able to consume their data in an ad hoc manner, but they didn't own any data products at all. And now, we're giving them the resources to actually own the data product as a whole, and I think this is a very big shift. It is a very important but challenging task to teach them the ability to work in a data domain."* (Interviewee R)

In conclusion from the case studies, the ability to adapt the design of your data architecture to the decentralized nature of a Data Mesh can be one of the biggest challenges when preparing for a migration to a Data Mesh architecture. It is important for organizations to assess this ability before embarking on the journey as it is one of the biggest tasks in the process. During the demonstration of the readiness assessment, it appeared that organizations find it difficult to estimate this ability for themselves, and that the set of follow-up questions was often used to guide them in this assessment.

5.4.4 Insights about assessing the preparedness for change

Accountability

With respect to accountability, an identified factor is having **clearly defined roles and responsibilities**. This means that every domain should have a data product owner or data product manager, in one of the domain roles. These roles and responsibilities should be correctly described, which can be offloaded with data ownership guidelines and Service Level Agreements.

According to interviewee T, these roles and responsibilities should be defined from the beginning on in order to avoid confusion with respect to the rightful owner of the data:

> *"When I see this factor, I think we could have done this better ourselves from the start. We have started taking temporary ownership of data products, as a trial. We figured out how to find the right people for the right data products. The problem with that is that when you work with temporary ownership, there are still many cases where the owner of the data does not know everything about the data itself. As a result, the data product can still not be used properly. I don't believe that having data engineers in the domains is a prerequisite from the start, as long as you have defined ownership of the data well and clearly. The person knows the use cases for his own data and guarantees its quality."* (Interviewee T)

Interviewee U states that their organization uses performance indicators to see whether the quality of the data is maintained correctly by its owners:

> *"In order to test the right accountability, you could measure the outcome of the domains. For example what you do with any other activity in organizations, like KPIs for data quality. Domains should be independent in how they achieve this quality, but you need to have some kind of common standard of measuring quality, in order to see if this ownership is correctly appointed."* (Interviewee O)

Contrary to some case studies, interviewee R argues that it is in principle the responsibility of the domains to define the roles and responsibilities, as long as they continue to comply with the set global rules and guidelines set by the central teams:

> *"Apart from the roles and the responsibilities in the domains, it is also important to define what is the responsibility of the centralized platform that still provides components to these domains. That's on a higher level, but within the domains there also should be clear responsibilities. What is the data product owner? What does the rest of the teams do? However, the way they address this appointment, is up to the domains itself. They should have some degree of freedom in order to decide on their inner processes, as long as they adhere to the standards of the Data Mesh."* (Interviewee R)

In conclusion, it can be stated that when assessing the degree of preparedness for a Data Mesh migration, the correct appointment of roles and responsibilities must always be taken into account. How this is arranged within the domains, either by pre-defining those roles and responsibilities, or by leaving this appointment to the domain itself, is something in which the practices of the case studies performed differ.

Governance

With respect to governance, the exploratory interviews have shown that this could be designed in several ways. For example, a governance forum could be established, in which data producers and data consumers can be brought together in order to be part of the decision making. However, the governance in a Data Mesh should be distributed onto the domains with some sort of rules, guidelines and policies what their data products should adhere to.

Interviewee S believes that federalized governance must have some or some degree of central leadership:

> *Within a federalized governance model, there must always be a centrally organized team that can enforce the global rules. Without this central team it is impossible to impose specific policies and procedures throughout the organisation. The federated character of this model then mainly covers the inner workings of the domains and the way in which they operate within their domain, as long as they comply with global standards.* (Interviewee S)

Interviewee N, on the other hand, indicates that according to his organization, care should be taken with the amount of central leadership, because according to them this could lead to new bottlenecks:

> *"If the governance was very centralized and top down, the Data Mesh wouldn't work. What will happen is that the central governance group will become the bottleneck to the overall system, because it is a centralized resource that doesn't scale with the rest of the organization. So the governance has to be distributed as well."* (Interviewee N)

Interviewee R indicates that there is a balance between centralized and decentralized governance, and that the degree of democratization in particular has proven to be important for success:

> *"So you could have two separate teams that are working on governance, but with different goals. On the one side, you have the centralized data platform team, who define the guidelines of what the standards should be of the data products. And then on the other hand, you have the federated governance team members within the domains, who make sure that these rules, guidelines and standards are enforced within the domains as well. But these teams need to be working together very closely, in a fully democratized way."* (Interviewee R)

Both the exploratory interviews and the case studies showed that especially the interpretation of the federalized governance model differs among all participants in this study. This difference therefore also indicates that the way in which governance within a Data Mesh should be set up is not yet completely clear and understandable, and that there is still plenty of experimentation going on. The consensus that did exist among the case studies is that at least certain policies, standards and guidelines are needed to enable sharing and access of data products between the domains in a trustworthy manner.

Workable Approach

One factor within the dimension Workable Approach is **communication guidelines**. These guidelines define how to work together to build a data product, in a multidisciplinary way. They also align cross domain collaboration through data exchange. Within the case studies, it resulted in different perspectives among the interviewees.

> *"To me, this collaboration is already defined through the interoperability policy and the documentation policy. I think the collaboration doesn't really need guidelines, but it needs a place to be able to collaborate with each other. Since domains have some interdependencies, yet still it is only the final products that they expose."* (Interviewee O)

> *"Domains do need some sort of guidelines in which they understand what is the best way of doing things, and maybe what sort of skills they need to become data driven. But for sure, this isn't something that should be enforced, since it is still up to the domains to decide on how to collaborate and communicate."* (Interviewee R)

On the other hand, with respect to **data literacy guidelines**, these guidelines define how to work on the infrastructure that the organization provides to support the Data Mesh. Moreover, these guidelines show how to work with the tools that are used within the domains. According to the interviewees from the case studies, data literacy is mostly about upskilling the people in the domains.

> *"I think that workable approach is mostly about skills and literacy. You need some data training, some necessary skill set. How do we build up these skills?"* (Interviewee O)

Moreover, some case studies related the workable approach within the Data Mesh with the existence of a **knowledge sharing platform**, that was implemented in order to support the learning culture:

> *"Implement knowledge exchange sessions, in which the engineers and analysts can discuss what they are doing within the domains, what technologies they are using, and the lessons learned from the Data Mesh performance."* (Interviewee R)

IT Capacity

With respect to the Workable Approach, there were varying perspectives on the level of automation and standardization there should be on the way of working within the Data Mesh. Most case studies approached having **CI/CD standards** and **automation standards** are identified as nice-to-have features, but not necessary for the performance of a Data Mesh:

> *"It is up to the domains whether they want to use CI/CD. In order to help the domains that wish to have these, the central platform team could provide some of these standards, which can be easily implemented within the domains. But it is going to be up to the domains to decide whether they want to use them."* (Interviewee R)

> *"CI/CD standards and automation standards really improve the way of working for a Data Mesh, but they shouldn't become a barrier for people starting to implement a Data Mesh. These features would be very nice to have, but they aren't necessary for the working of a Data Mesh and shouldn't scare people willing to implement it."* (Interviewee O)

However, the case study of Interviewee N showed that these CI/CD standards were the enabling mechanism for the development of the Data Mesh performance:

> *"What usually is not considered, is what happens when I change a column in a data product that is used by consumers. New standards need to be thought of to prevent breaking use cases for downstream users. This really comes back to the thinking of data as a product. To us, CI/CD was a mechanism that accomplished that."* (Interviewee N)

Likewise, Interviewee T agreed upon the importance of CI/CD standards in a Data Mesh architecture:

> *"If you don't do CI/CD, and foremost Continuous Deployment, you are going to be really slow. Especially when you need to manage multiple data products, you will need continuous deployment. The same thing holds for automation. Whenever you want to scale, you need automation in place. And Data Mesh is all about scalability. Maybe the first month that you are trying out the Data Mesh you don't care too much about the automation. But the moment that you start scaling, you need it."* (Interviewee T)

Lastly, according to all case studies, having **data publication standards** is identified as crucial for the performance of a Data Mesh:

> *"At the end of the day, the data product publication standards are the most important, to make sure that the publication of the data products enables interoperability between the domains. It is up to the domains you they go about automating and deploying this publication, but the publication standards have to be met."* (Interviewee R)

Enterprise Capacity

With respect to Enterprise Capacity, one of the factors that was identified was the ability of an **agile way of working**. The exploratory interviews identified this factor as important, since it speeds up the time to market of specific publications, something that is needed within a Data Mesh.

> *"My own personal views is that anything you design is wrong, either currently because you didn't fully understand the problem or requirements, or eventually because something changes. A company has to be agile, it has to be constantly moving forward and evolving."* (Interviewee N)

> *"I think an agile way of working is needed for anything. But actually even more important is the ability to enable the domains to make their own decisions. So if the domains decide to not work agile, that's up to them."* (Interviewee S)

Lastly within this factor dimension, was having **team structures aligned with the business domains**. These teams are already business aligned in terms of that they're having a responsibility for a certain part of the business. Moreover, this brings them closer to the business case, which might enhance responsibility and motivation for the business processes. The case studies have shown that this is not the case for every organization, because the perspectives of the interviewees differed greatly on this factor. The case study with interviewee R showed that most teams are business aligned, but that there are always teams left in the mesh that are not aligned with the business.

> *"I think this is an interesting view. I don't think this counts for every case. For example, the consumer aligned domains can be really business focused. But what about the source aligned domains, like the teams that maintain the systems, who don't really have a business domain. They're actually maintaining the ERP or CRM system. These teams could be a domain team as well, but they could also be part of the central IT team. But they are very important, since they are the provider of the data."* (Interviewee R)

Interviewee T believes it is necessary to align the domain teams with the business to preserve domain knowledge:

> *"You have to have some degree of this alignment in a sense, because you want to have specific domain knowledge in your domains. This requires a team structure alignment with the business. With this domain knowledge you don't have to figure out every time how to tackle certain domain-specific issues. There is one team that is not aligned with the business, and that is the team that builds the platform. They serve the entire community."* (Interviewee T)

5.4.5 Insights about determining the readiness

Data Mesh Building Blocks

The Data Mesh building blocks are **self-organizing and autonomous domain teams enabled by a self-serve platform**, having a **central IT team**, having **domain representatives in the governance team**, having a **self-serve platform** and having a **central data catalog with DATSIS (i.e. discoverable, addressable, trustworthy, self-describing, interoperable and secure) data products**.

5.5 CROSS-CASE ANALYSIS

5.5.1 The needs of organizations

When assessing Data Mesh readiness, the case studies have shown that the motivation and need for a data mesh architecture within organizations often come from different angles. While one organization mainly needs more autonomy for the business domains and shorter lead times between requests, the other organization mainly needs structure, responsibilities and general data awareness. It seems that the more data mature and data driven an organization is, the more it tends to the former. Organizations that want to become more data-driven often use that need as a driver for a Data Mesh architecture. It has therefore proved difficult during the case studies to tick all the needs as specified in the DMRM within one organization. In practice, it often comes down to a combination of different needs, and the needs from the model do help organizations to become more aware of these specific needs.

A special insight from this phase is that although Data Mesh is often compared to other distributed architectures such as a Microservices design, none of the case studies had implemented such a Microservices design prior to the Data Mesh in the organization itself. This indicates that while already having a decentralized architecture is probably a good starting point for expanding to a Data Mesh architecture, while in practice that is often not the case. One reason for this could be that the drivers behind the implementation of a Microservices architecture often came from the software development angle and are also often implemented as a technological solution focused on the interoperability of services (Krivic et al., 2018), while a Data Mesh architecture is almost completely separate from a technological solution and is usually evoked from a organizational motivation behind data issues. The two different distributed architectures thus function as other total solutions for other challenges within organizations, and should therefore also be used for different purposes.

5.5.2 The ability of organizations to migrate

The ability of organizations to migrate to a Data Mesh architecture is mainly determined in the model in response to high-level requirements. These requirements are mainly within the cultural and organizational context, because during the exploratory interviews it often turned out that these are important preconditions for a successful Data Mesh migration. During the case studies, these factors were fairly easy to identify. What emerged is that there were many different perspectives and interpretations about the way in which organizations decentralize their central data platform towards a distributed architecture.

Getting business and IT aligned

One of the current discussion points within this topic is the alignment of business and IT. Although many different organizations approach achieving this alignment in different ways, there is a consensus that having this alignment is very important in terms of readiness for Data Mesh architectures. Most case studies showed that in many cases this alignment does not exist: the business does not see the value of IT and vice versa. One way to bring these groups closer together turned out to be possible to discuss data awareness and the value of data for the organization, so that especially on the business side of organizations more value is seen in the use of data for decision-making. The moment they understand that having and understanding data is an important requirement for the performance of the business, it is easier to implement data-driven transformations within the organization. It turned out to be important to align the tasks of the data teams within organizations with what the business sees as the most important conditions for good performance.

The process of decentralization

Regarding the process of decentralizing and splitting the existing monolithic data platform in the organization, some of the case studies showed that it can be good to start with domains that are eager to become autonomous. In this way, these domains can serve as proof of concept, and also provide support for the migration within the organization. The domains most hindered by the centralization of the data architecture also have sufficient motivation to help with the iterative and lengthy process. In addition, there was little consensus among the case studies about the best way to approach the decentralization process: while one interviewee believes that the existing business domains already exist and it is therefore easy to identify domains, the other interviewee believes that this is a lengthy and painful process that often has to be redone. The conclusion that can be drawn from this is that it is in any case a process that needs to be paid a lot of attention when determining readiness for a Data Mesh architecture. It seems that at least a structured and well-thought-out approach helps to prevent or reduce chaos during the decentralization process.

Decoupling existing business applications

Especially from the critics of the Data Mesh architecture there is doubt about the connection of the Data Mesh architecture with the application level within organizations. Within various case studies, existing applications were compared with the solutions that the Data Mesh offers or insufficiently offers. By enabling the discussion on this point, it became apparent that it is very organization dependent on what choices are made within the Data Mesh for certain enabling technologies that make the operation possible. It turned out to be important to realize that the concept of a Data Mesh mainly revolves around the organization, architectural decisions for moving data and governance, and that the choice of technology should mainly be seen as an executive decision rather than an organizational priority. Organizations make different choices about which technology to use, while still similarly complying with the organizational philosophy that the Data Mesh concept advocates (I. A. Machado et al., 2022a).

5.5.3 Preparing for a Data Mesh architecture

During a number of case studies, doubts arose whether the step of assessing preparedness really belongs to a readiness assessment, or whether it leans more towards execution instead of readiness. Within this study, it was decided to keep the preparedness component in the assessment, because it can also be useful to determine the extent of preparation when you want to assess your readiness for migration to a Data Mesh architecture. In addition, the case studies have shown that good preparation for the migration can contribute to a better adaptation of the organization to the Data Mesh architecture. The preparedness as measured in the DMRM covers the formal rules and guidelines that must be established in order to meet the institutional boundaries of a Data Mesh.

Communication and collaboration

Within the case studies, a link to collaboration often arose during the discussion about communication. It is therefore not only about the move towards a model in which someone takes ownership of the data product, but also about enabling collaboration on it in a multidisciplinary way. Questions that arose related to the requirements for this collaboration and how you can work together to build a data product. A shared opinion within this dimension is that bringing different disciplines together contributes to a large extent to communication and collaboration, simply by reducing the distance between the disciplines. In addition, having certain communication guidelines can contribute to communication between domains that are very far apart, but this is not a prerequisite for preparing for a Data Mesh migration. It is often about getting enough common ground among the domain teams so that they understand each other what they are doing, and upskilling the people that lack knowledge about it, in order to enable the entire domain to work autonomously. The way in which domains work together internally is basically up to the domains themselves to determine, due to the autonomous nature of the Data Mesh architecture.

Governance

A much discussed point during the case studies was the way of designing the governance model in a Data Mesh architecture. Some organizations initially opt for a minimally federalized governance model, in which only a number of global rules are enforced that ensure that the performance of the Data Mesh will in any case remain within the applicable governmental rules. Other organizations chose to design the entire federalized governance model from scratch, including a central governance team, so that all institutional preconditions were in place before starting the migration. A consensus that can be found within this dimension is that there should in any case be standards about publishing, sharing and accessing data between the domains in order to guarantee interoperability. In addition, many organizations leave the way in which data products are built to the domains themselves, as long as their output complies with the guidelines. With regard to Data Mesh readiness, it turned out to be important to involve as many people as possible within the organization in the decision-making process when drawing up this governance model, so that a large organizational base can arise for the new democratized governance model.

CI/CD and automation standards

A dimension where it was more difficult to find consensus among the case studies was the IT capacity and data maturity of organizations. Standards regarding CI/CD and automation were seen by some interviewees as necessary for the functioning of the Data Mesh, while other interviewees saw these standards as a 'nice-to-have' feature. An analysis of the different cases shows that with regard to this dimension it can be stated that organizations should not be deterred by these maturity standards. A certain degree of automation and CI/CD does contribute to an efficient and fast performance of the new distributed design of the data architecture, but it is not an organizational capability that must be established from the start. When assessing the readiness of an organization for the migration to a Data Mesh, it is therefore not a priority to have these standards in place, because it is not a precondition for Data Mesh readiness. It was decided to keep these factors within the model, but with a lower priority, because when using the DMRM iteratively, these factors can be implemented over-time within an organization to improve overall data maturity towards readiness for the Data Mesh migration.

Inter-domain dependencies

When assessing readiness for a Data Mesh migration, the issue of how to deal with inter-domain dependencies often arose. Because the domains often work together on data and insights from the data, it is important that this collaboration can be reliable and structured. It could happen within organizations that one domain is dependent on the other domain for certain data, but that the prioritization of this data is not the same between these domains. To increase the reliability of the cross-collaboration within the organizations, many organizations chose to draw up release management contracts and Service Level Agreements, which specifically state who owns the data so that consumers know what to expect from the data product.

Team structures alignment

A final analysis of the case studies is about the way of structuring the teams. The case studies have shown that it is valuable to structure the domain teams according to the existing business domains, in order to maintain domain expertise in the new Data Mesh domain teams. However, the case analyses and refinement sessions have also shown that it is not always possible to structure these teams aligned with the business domains themselves. One reason for this is that there are often teams within the organization that are not business aligned, such as the maintainers of ERP or CRM systems, where the people who understand these complex systems and can extract data from them are often scarce resources. The question arises whether you want these teams in domains or in a central IT team. A clear answer to this issue has not emerged from the case studies, so it depends on the organization whether you want these teams in the central IT team, or whether you hold the domains responsible for the extraction of its own data.

5.5.4 Determining Data Mesh Readiness

Determining readiness for a Data Mesh architecture appears to be a complicated process, especially due to its organizational complexity. In any case, the case studies have shown that a static assessment in the form of a checklist does not work, because it often involves a lengthy process in which a full Data Mesh with all its principles and building blocks can often not be implemented in the first instance. In contrast, it is merely an iterative process, in which a Data Mesh is formed within an organization, building block by block, over a time-span of several years. The purpose of a readiness assessment is therefore mainly identified in these case studies as a guideline for structuring and organizing the process as a whole, and for forming insights into the organization's strengths and weaknesses in relation to the migration. With regard to the last readiness phase, namely that of a complete set of Data Mesh forces, it has turned out that a complete assembly of these forces hardly occurs in practice. This representation within the DMRM should therefore mainly serve as direction and an example of the principles of a Data Mesh to be implemented, without these all having to be crossed off at once.

5.6 CONTEXTUAL ANALYSIS

The case studies in this study were conducted in organizations with experience with Data Mesh architectures, as well as organizations without this experience. The organizations with experience then had these to varying degrees: there were organizations that consisted of consultants who help other organizations with a Data Mesh migration, as well as organizations that have tried to implement a migration internally. Although the number of organizations studied must be much larger to draw a more reliable conclusion about the contextual influences on Data Mesh Readiness, tentative conclusions can be drawn from the analysis in this study.

First, organizations without knowledge or experience with Data Mesh architectures seem to be a lot more critical of the concept than organizations that do have this knowledge. They quickly compare the concept to a compilation of existing technological aspects and question the impact of the innovation. One reason for this could be that these organizations do not yet have the experience because, for example, they have not yet felt the need for a Data Mesh architecture, or because they have not come into contact with it because they have found alternative solutions for the existing needs. In addition, it is more difficult for these organizations to use the DMRM, because the lack of experience can occasionally ensure that the terminology associated with the Data Mesh topic is not yet fully understood by the organization, causing the interpretation of some factors may differ. One way to solve this is, for example, by adding more and clearer follow-up questions to the description of the factor in the model.

Second, organizations that consult other organizations in a migration to a Data Mesh architecture seem to be more critical of the degree of readiness for migration. This may be because the revenue model of these organizations consists of presenting a complete plan for the migration, while some case studies have shown that a Data Mesh migration is an iterative and lengthy process, emerging and developing over a time span of multiple years. It may therefore be more difficult to come up with a complete plan of action in one go, showing complete readiness for a Data Mesh architecture.

Third, organizations that are more traditional and less data mature seem to score worse on Data Mesh Readiness. This could be explained by the fact that they, for example, do not yet feel the bottlenecks in the data platforms, or are simply not yet data-driven enough to start with a Data Mesh migration. For these organizations, it may be too early to think about Data Mesh readiness, and the priority is currently starting to increase the smaller digital efforts instead of changing the entire organizational architecture.

5.7 CHAPTER CONCLUSION

This chapter presented an overview of the demonstrations performed and provided different perspectives on the performance and the content of the DMRM. The results from the theoretical and empirical design of the DMRM were used to gain insights from the practical application environment. This chapter aimed to provide different perspectives, which have showed that there are mahy different ways in which organizations approach the migration to a Data Mesh architecture. When it comes to determining the readiness for a Data Mesh migration, the DMRM offers guidance in assessing the as-is situation of organizations, and aims to provide new and clear insights to these organizations. Because the conduct of the case study is supervised by the researcher, when assessing the readiness factors, follow-up questions were asked that make it easier for the organizations to arrive at these insights. The cross-case analysis has shown that the perceived degree of readiness among organizations depends on the context in which they are located. For example, more data-driven organizations are generally more aware of their position towards Data Mesh migration, and organizations with no experience at all with Data Mesh architectures are more critical of the concept than organizations that do have this experience. In addition, it appears that generic background knowledge about the functioning of a Data Mesh architecture is required to be able to use a large part of the model, because this is based on knowledge about the terminology within the topic. In the following chapter, the DMRM will be tested in practice in organizations under different circumstances, so that the potential differences can become more clearly visible.

6 TESTS WITH THE DMRM

This chapter presents the research phase of testing the developed DMRM and its accompanying assessment instrument. In order to determine the generalizability of the model, as well as the influence of the researcher during the assessment, the model in this chapter is tested under a group of test cases. According to Hevner (2010), the new artifact should be evaluated in a real business context. Therefore, the model is tested among three organizations, varying on background knowledge about Data Mesh and the presence of the researcher during the assessment. Testing of the DMRM was conducted under different conditions, further explained in the next section. The purpose of testing the artifact is to see whether users of the DMRM can use and interpret the model in the same way, or whether this interpretation differs under the circumstances in which the model is used. In this way the model is tested for perceived comprehensibility and perceived usability. After conducting the test cases, the participants were questioned about the perceived usability, completeness and comprehensibility of the DMRM and the assessment instrument.

6.1 TESTCASES

It is expected that the generalized conclusions from the DMRM will differ if the organizations' experience with Data Mesh architectures differ. In addition, it can be expected that the availability of the researcher during the assessment can also influence the conclusions from the model. That is why it was decided to test the DMRM under three test cases:

- First, the model is tested with an organization **without** experience with Data Mesh architectures and **without** presence of the researcher
- Second, the model is tested with an organization **with** experience with Data Mesh architectures and **without** presence of the researcher
- Third, the model is tested with an organization **without** experience with Data Mesh architectures and **with** presence of the researcher

The fourth test case in these dimensions, namely an organization with experience with Data Mesh architectures and with the presence of the researcher, is not included, because this dimension is equal to the case studies performed in this research. An overview of the testcases and their contextual differences is shown in table 6.1.

Table 6.1: Testcase matrix

	Without presence of the researcher	**With presence of the researcher**
Without Data Mesh experience	Testcase 1	Testcase 3
With Data Mesh experience	Testcase 2	Case studies

In the test cases that are taken without the presence of the researcher, a Data Mesh Readiness Assessment instrument is used. This instrument takes the form of a questionnaire, using the follow-up questions as described in chapter 4. The entire Data Mehs Readiness Assessment instrument can be found in Appendix F.

6.1.1 Testcase 1
Without Data Mesh experience — self-assessment without researcher

The first test case was done with an organization without experience with Data Mesh architectures and without the presence of the researcher. The organization within this test case is a traditional organization with a large and fairly centralized IT department with limited organizational power to make real changes, but enough power to maintain ownership over many of the systems. Each enterprise operates significant sub-businesses in the business as large domains with quite some autonomy for their own business solutions, with central governance. Within the organization, software engineering skills are necessary for data analysis within IT. The organization has more than 15,000 employees. When assessing the need for a Data Mesh architecture, the organization mainly agreed with the need for more autonomy for the domains and the need for more domain understanding. The needs that were less present in the organization were assessed sufficiently, and the assessor indicated after the assessment that the sub-questions helped to assess the as-is situation of the organization. The needs assessment also unconsciously gave direction to determining the areas from which the organization wants and can derive value. Regarding the assessment of the ability to decentralize and decouple, it was clear from the assessment that the participant tried extensively and in detail to answer the question. The first efforts to decentralize and decouple are there, but it still turns out to be a complicated process. After the test, the participant indicated that answering the questions within these topics helped a lot in recalling the current situation and possible approaches to improve it. The participant gained a clear picture of the needs for decentralization, by means of giving specific and concrete answers to the questions. The assessor stated that although they already gained some idea of what would be needed within the organization for a Data Mesh migration, assessing the preparedness of the organization towards a Data Mesh migration was helpful in creating insights. Although not all factors were present in the organization yet, the assessor indicated for some factors they would form a good approach to integrate the factor into the organization. For the factors that were not present, the assessor was able to clearly indicate why they were not yet there, even without the presence of the researcher to guide these questions. A full description of Testcase 1 can be found in Appendix G table G.1.

6.1.2 Testcase 2
With Data Mesh experience — self-assessment without researcher

The second test case was carried out with a data-driven organization with almost 3,000 employees, and with a lot of experience in the field of Data Mesh architectures. The latter is also clearly reflected in the results of the assessment. The participant was able to give short and concise answers to the questions and at the end of the assessment it also appears that many of the Data Mesh building blocks were already present in the organization. When assessing the need, it was also clear that the participant understood where the factors in the model came from, because this organization could identify with almost all needs. This, in contrast to the other test cases where the participants had no experience with Data Mesh architectures. There were also no problems in assessing the ability to change and the preparedness for Data Mesh architectures: almost all factors were present in the organization, and for the factors that were not present within the organization but were included in the assessment, the participant obtained new insights. After the self-assessment, the participant indicated that the assessment was very complete on the topics and that the assessment was easy to complete and understand, so that this self-assessment could also be performed without the presence of the researcher. A full description of Test Case 2 can be found in Appendix G table G.2.

6.1.3 Testcase 3
Without Data Mesh experience — with presence of the researcher

The third test case was conducted with an organization with no experience with Data Mesh architectures, and also with no intention of migrating to a Data Mesh architecture. The organization is, compared to the other testcases, a smaller organization with less than 1,000 employees, with a centralized data platform. An initial prognosis showed that the organization itself still processed too little data to be able to come into contact with the concept of Data Mesh architectures. This image also emerged during the assessment of the need. The organization could identify with only a few needs, but especially did not experience the needs surrounding a shortage of scalability and major bottlenecks. During the assessment, it emerged that there were no problems with shortcomings of the central data platform within this organization, and that this platform was still able to meet the requirements well. During the problem identification, it was therefore found that it was not necessary to continue the assessment to the next step in the readiness chain. What has become apparent from this test case is that it is not necessarily necessary to have a researcher present for the assessment. The follow-up questions and the factor descriptions make it clear which factors are tested and the participant only needs to formulate answers to the questions. During the assessment, the participant's suspicions were confirmed, because there was indeed no need for a migration to a Data Mesh architecture. Assessing the ability to change, despite the lack of the need for a Data Mesh architecture, did provide some insights into areas within the organization for improvement. Factors specifically targeting the readiness for a Data Mesh architecture were not relevant for the organization to answer. The results of this testcase are in line with the conclusions drawn up in section 5.6. A full description of the assessment of Testcase 3 can be found in Appendix G table G.3.

6.1.4 Testcase Outcomes

Table 6.2 presents an overview of the outcomes of the testcases, categorized by the type of testcase, the resulting readiness phase, the perceived usability by the participant of the testcase and the general outcome of the testcase.

Table 6.2: Outcomes of the testcases

Case	Readiness	Testcase Results
1	Almost ready	**Type of testcase:** self-assessment without researcher **Current readiness phase:** 3 - Preparedness for migration **Perceived usability:** the participant indicated that answering the questions within the factor dimensions helped a lot in recalling the current situation and possible approaches to improve it. **General outcome**: as the assessor was already relatively far along the organizational readiness steps, using this assessment was simple and understandable. The assessment brought the assessor new insights and improvement areas in the future.

2	Ready	**Type of testcase:** self-assessment without researcher **Current readiness phase:** 4 - Determining readiness **Perceived usability:** the participant indicated that no problems were encountered with answering the assessment questions **General outcome**: because this participant was already approaching the end of a successful Data Mesh migration, it was easy to go through the assessment in its entirety. A factor here is that the participant had already thought about many of the assessment questions himself for a long time and therefore had already prepared a large part of the answers. The participant indicated that the assessment was very complete on the topic and that it was easy to complete and understand. In addition, the participant indicated that he had gained new insights about the improvements of certain factors that had not yet been implemented in the organization.
3	Not ready	**Type of testcase:** assessment with researcher **Current readiness phase:** 1 - Assessing the need **Perceived usability:** due to the participant's lack of Data Mesh experience, the usability of the assessment was reduced compared to the other test cases. Going through the assessment was difficult for certain factors, because the participant had not yet thought about it. **General outcome**: this test case has shown that in the absence of a clear problem identification, it is not necessary to complete the assessment in its entirety. Because it was still unclear to this organization whether a Data Mesh architecture would offer the solution to their problems, the first step of the assessment could not be completed in its entirety. It is recommended to discontinue the assessment in such cases and to resume it (in an iterative manner) as soon as a clear choice can be made towards a Data Mesh migration.

6.2 CHAPTER CONCLUSION

Chapter 6 presented the results for the testing phase of this research. The model was operationalized by designing a readiness assessment questionnaire, formed through the formulation of factor-specific follow-up questions that guide the self-assessment of the model. This operationalized self-assessment can be used by organizations to determine and track progress in their process prior to their Data Mesh migration. This phase tests the gained insights from the design and development phase by means of an operationalized artifact in three different practical contexts. In conclusion, it can be stated that when using the operationalized readiness assessment, the results of the assessment do not depend on the presence of the researcher. What does influence the results of the readiness assessment is the degree of experience with Data Mesh architectures. Although organizations with little experience within the Data Mesh topic, in addition to a basic background knowledge about the concept, can conduct the readiness assessment using the follow-up questions, it strongly depends on the presence of the Data Mesh needs whether the assessment is useful for an organization or not. In the next chapter, the performance of the DMRM will be evaluated on the base of usability, comprehensibility and validity during a set of evaluation sessions.

7 EVALUATION OF THE DMRM

7.1 EVALUATION

In order to evaluate the DMRM, evaluation sessions were organized to evaluate the model and the accompanying assessment on their performance. During these evaluations, the assessment document was sent to the participant, after which the participant was asked to evaluate the model and the assessment in terms of usability, comprehensibility, and validity. The evaluations were conducted with the participants as shown in table 7.1. In order to gain an objective perspective towards the performance of the DMRM, the participants of the evaluation sessions vary among their experience with Data Mesh architectures.

Table 7.1: Evaluation participants

Participant	Operating Group	Data Mesh experience
U	IT Consulting	No
V	Financial Services	Yes
W	Retail	Yes
X	Insurance	No
Y	IT Consulting	No

7.1.1 Usability

The outcomes of the usability evaluation sessions were:

- Participant U, Participant V and Participant Y stated that the model was experienced as complete on the Data Mesh influencing factors it covers.
- Participant V indicated that the assessment itself lacks a direct follow-up section that provides clear advice based on the answers given. This would shorten the processing process, as a report is currently being compiled manually and after completion of the assessment. An improvement could be made here, for example by establishing the reporting automatically or by already indicating in the assessment what advice is based on potential answers.
- Participant W indicated that the model is very focused on how large organizations could move from a centralized data-team paradigm to a federated Data Mesh paradigm, for managing their own data in a Data Mesh architecture. The participant indicated that this makes the model less useful for his organization, which is a B2B product organizations, which would for example work with multiple "Data Mesh systems" on behalf of their customers, with data that is owned by the customers and not them.
- Participant X indicated that the model would be less useful for his organization because, due to the lack of large amounts of data to be processed, they would not meet the problem identification for a Data Mesh. Although the participant indicated that the model was useful for assessing maturity towards data drive and other digital ambitions, a large part of the model was too specific for the organization to be fully useful.

7.1.2 Comprehensibility

- All participants indicated that the assessment was comprehensible to use. Specifically, they indicated that the way in which the assessment is set up, namely with the factor descriptions and the follow-up questions, supports the purpose and comprehensibility of the assessment.

7.1.3 Validity

- Participant U stated that regarding the validity of the model, the interpretation of readiness may differ among the assessors. While one interprets the assessment as an iterative tool that can be used during the entire process prior to and at the start of a Data Mesh migration, the other indicates that readiness is in principle only the area prior to the implementation process. The statement of participant U is in line with the outcomes of the cross-case analysis with regard to the Assessing the Preparedness stage in section 5.5.3. Due to the fact that from this analysis it had turned out that assessing the degree of existence of executable preparedness components can be useful for determining Data Mesh readiness or even for guiding Data Mesh implementation, it had been decided to keep this component in the DMRM.
- Participant V and participant X did not mention any improvements regarding the validity of the model.
- Participant W indicated, in line with the comments about the usability of the model, that the model was designed for a specific business context, different from that of their organization. Although the model appeared valid for the scope of the study, improvements could be made in the expansion of the model for changing business contexts and research scopes, according to this participant.
- Participant Y indicated that instead of organization domains, the model could also refer to organization capabilities. In that case, the primary modeling would be the capability, rather than information or process orientation. It is then about the overall capability to execute a process, rather than the process or information needed for it. The participant indicated that this approach would also map neatly to domains.

7.2 CHAPTER CONCLUSION

This chapter gave an overview of the results gathered through 5 evaluation sessions on the usability, comprehensibility and validity of the DMRM. Overall, the model was perceived as complete on its dimensions addressing organization readiness towards Data Mesh migrations. The participants also have stated that the assessment tool was perceived as easy to use and comprehensible on its structure. Suggested improvements lie in extending the scope of the model to B2B assessments and extending applicable business contexts, building automated feedback mechanisms, and automatically aborting the assessment when the problem identification is not met by the organization.

8 | CONCLUSION AND RECOMMENDATIONS

This chapter presents the conclusions which can be drawn up from this research endeavor. Since this research was conducted following the DSR research methodology, the presented conclusions serve as contributions to the existing body of knowledge on Data Mesh architectures (Hevner, 2010).
The aim of this study was to develop a Data Mesh Readiness Model (DMRM) that would guide organizations in the process of determining their readiness towards migrating to a Data Mesh architecture. The results presented in previous chapters are discussed in this chapter to provide an overview and further explanation of these results. This chapter presents the main findings of this research, as well as the limitations of the study, the contributions of the study to the existing knowledge base, and recommendations for further research within the topic of Data Mesh architectures.

In order to build the required theoretical background knowledge on Data Mesh architectures, chapter 2 presents a description of the design of Data Mesh architectures based on its core principles. This design is visualized by a reference architecture of a Data Mesh architecture design at the end of chapter 2. These research activities serve as an answer to the first sub-question: *"What are the characteristics of a Data Mesh architecture?"*.

This research introduced a Data Mesh Readiness Model for assessing the factors influencing Data Mesh readiness of organizations. A comprehensive assessment model was developed that could be utilized by organizations before and during their Data Mesh migration. The DMRM is used to determine the as-is state of organizations relative to their current readiness level and their desired state of readiness, by assessing the organizational and technological requirements needed to successfully engage in a Data Mesh migration. Subsequently, the gap between the as-is state and the desired state of Data Mesh readiness can be analyzed, on the basis of which an improvement path can be formulated.

The building blocks of the first version of the model were synthesized from a literature research and exploratory interviews, aiming at the identification of organizational factors that influence Data Mesh readiness. This complete set of influencing factors, as shown in table 3.2, provides an answer to the second research-question: *"What are the factors that influence readiness for migrating to a Data Mesh architecture?"*.

In order to determine the structure of the model, the socio-technical complex nature of a Data Mesh migration has been taken into account: to determine Data Mesh readiness, both the technological and organizational dimensions of an organization have to be assessed. Due to this fact, a new kind of two-dimensional readiness model has been designed, which assesses technological factors within each step through organizational steps towards readiness. For the design of this structure, both the TOGAF Business Transformation Readiness Assessment (TOGAF, 2018), as well existing studies on organizational change were used. The combination of these existing studies also provide an answer to sub-question 3: *"Which readiness assessments with regard to IT architecture transformation are provided in literature?"*.

After mapping the identified factors onto this structure, the initial DMRM emerged. This model was further developed and refined through subject-matter expert refinement sessions and a Data Mesh expert group session. Because the path to a Data Mesh architecture for organizations is often unclear and unstructured, during the development of the model an attempt was made to simplify this path by identifying specific influencing factors with clear follow-up questions, to enable a structured approach to this assessment. Subsequently, these activities resulted in the Data Mesh Readiness Model (DMRM), as shown in figure 8.1. The combination of the aforementioned research activities add up to an answer to sub-question 4: *"How to design a model that guides organizations through assessing their readiness for migration to a Data Mesh architecture?"*.

In order to demonstrate, test and evaluate the DMRM in practice, the model has been operationalized for practical use by formulating follow-up questions for the assessment of the specific factors in the model. These factor descriptions can be found in chapter 4, and provide an answer to sub-question 5: *"How can the readiness model be operationalized for practical use?"*.

In the following research activities, the model was demonstrated in the context of seven case studies, in order to gather the functioning and insights of the DMRM in practice. Moreover, these demonstrations allowed for insights into the created knowledge base on Data Mesh readiness within an application environment. The model was then tested over a number of test cases, in order to determine the influence of the presence of the researcher and existing Data Mesh experience on the outcome of the model. Finally, the model was evaluated with five different participants, in order to determine the usability, comprehensibility and validity of the model. The results from the testing and evaluating research activities combined serve as answer to the last sub-question of this research: *"Is the designed readiness model and readiness assessment instrument applicable in practice?"*. The shared opinion of the participants on characteristics was positive. Improvements were suggested in applying the model in a wider range of contexts and organizations, automating assessment results and generic refinement opportunities.

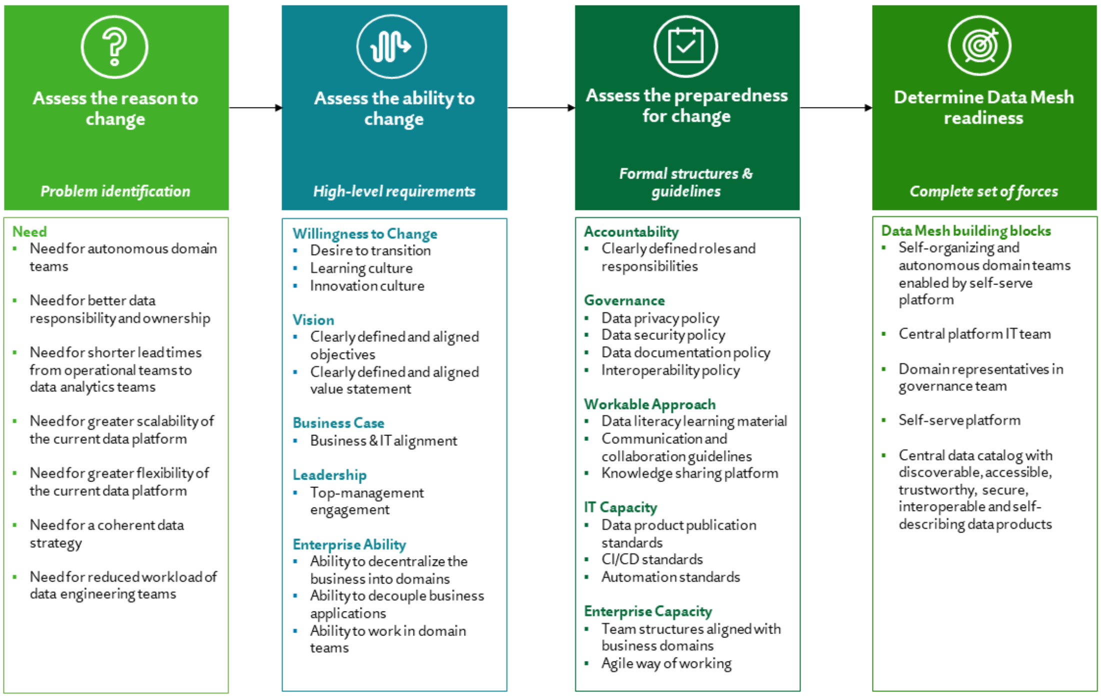

Figure 8.1: Data Mesh Readiness Model (DMRM)

8.1 CONTRIBUTIONS TO THE KNOWLEDGE BASE

This research extends the body of knowledge on the academic literature of Data Mesh architecture assessment methodologies by providing a two-dimensional readiness model that can be utilized to assess both the organizational capabilities as well as the technological implementations required for a Data Mesh migration. Because this design encompasses a new way of assessing readiness, the novel structure of the model could also be extended in the future to assess readiness towards a variety of other organizational transformations.

Since current research lacks a suitable Data Mesh readiness framework, organizations are currently unable to assess their readiness towards a Data Mesh migration. This makes it difficult for them to make a well-considered choice on when and how to transition (I. A. Machado et al., 2022b). This research addresses the current research gap by providing a Data Mesh readiness framework, which enables organizations to identify factors that influence their readiness towards Data Mesh migration and guides them through the process of assessing this readiness.

Moreover, due to the novel character of the concept of Data Mesh architectures, this research has laid an empirical foundation for the hitherto missing research on Data Mesh architectures, by acquiring empirical data among different contexts. This empirical foundation can fuel more scientific research on Data Mesh architectures in the future.

In conclusion, the DMRM provides a unique approach to assessing readiness towards Data Mesh migrations. Concurrently, this model broadens the knowledge base on readiness models that could be used in the domain of Data Mesh architectures.

8.2 PRACTICAL CONTRIBUTIONS

The main practical contribution lies in the design and development of a usable DMRM and associated assessment, which can also be used as a self-assessment by organizations apart from the presence of the researcher. The assessment thus offers new insights for customers of the assessment and can also guide them in an iterative way through the process of preparing for a Data Mesh architecture, as well as its implementation. Because the context in which the model and assessment were designed and developed differed within organizations based on degree of Data Mesh experience and progress, the practical contribution for these groups is different:

First, for organizations with a clear experience with Data Mesh organizations and an intention to embark on a Data Mesh migration, the DMRM provides a useful guideline for identifying needs and structuring the preparation process. The assessment can be used to assess the as-is situation and identify areas for improvement. With this it can give the organizations new insights and thus act as a roadmap within the trajectory towards migration to a Data Mesh architecture.

Second, for organizations without a Data Mesh experience and with no intention to start a migration, the model can help identify shortcomings of the current data architecture, and provide inspiration for possible future strategies. It is not necessary to go through the entire model, but parts of the model can be borrowed to build a custom-build future strategy in the field of data architectures. In addition, these organizations can also gain insights from the assessment in the field of improving the data maturity of the organization and a tactic for stimulating data drive.

The DMRM contributes in a more effective and organized way of preparing for a major organizational change. For example, it makes organizations consider how they can prepare their employees for the new way of working, but it also evaluates their own data maturity against new technological applications within the Data Mesh. The model also considers the implications of a Data Mesh migration, helping organizations assess their needs and determine whether a Data Mesh architecture is the right solution for them. This allows practitioners of the model in a specific sense to assess the bottlenecks and value of data architecture solutions within their own organizational context.

8.3 LIMITATIONS

In order to better understand the value of the results of the study while identifying possible directions for further research, the limitations of this study are presented below.

8.3.1 Design Science Research

Because the Design Science Research approach was chosen in this research, one of the limitations is the availability of resources for building the model. Because there was little theoretical and empirical foundation within the subject, it was decided to extend the theoretical research with exploratory interviews. A choice of participants was made based on their availability, and a different configuration of exploratory interview participants could have resulted in a different theoretical set-up of the model. Second, when selecting case studies within the model, no overall adequate set of use cases was found within which the model could be applied. Because assessing the readiness of an organization is very context dependent, a different use case within an unrepresented context could provide different results. In addition, the interviews and case studies were semi-structured, leaving room for the participants' own interpretation. As a result, the collection of the empirical data has depended on their perspectives and interpretations, and it is possible that different interviewees identified different factors with different interpretations.

8.3.2 Innovation enthusiasm

Because many of the participants of this study were found through internet search terms, word of mouth, or the Data Mesh Learning Slack community, the vast majority were found through the publications of their own articles on the topic and are thus excited about new developments and innovations. Therefore, it was harder than expected to find an objective set of participants that had a critical attitude towards Data Mesh architectures. Through the course of actions, this *innovation enthusiasm bias* is kept as small as possible by leaving a lot of room for criticism during the interviews and by adopting a critical attitude towards the topic. In addition, the participants in this study were not only selected based on different sectors, backgrounds and roles within the organization, but also based on their current experience with Data Mesh architectures. In this way, efforts have been made to find a representative est of objective attitudes towards the topic of this research.

8.3.3 Theoretical saturation

Due to the highly exploratory nature of this research, the main aim was to create a new theory within the Data Mesh topic, based on empirical insights. Because the generalization of existing theory was not feasible, a rich description and consensus of these empirical insights was sought throughout the research. The quality of this research can be supported by looking for a certain degree of theoretical saturation, which means that an extra observation, interview or focus group no longer yields extra insights with regard to the subject of the research (Glaser & Strauss, 1967). During the research, this theoretical saturation was sought by strengthening the empirical foundation with 13 exploratory interviews, 4 individual refinement sessions, one evaluation session with 26 experts, 7 case studies, 3 test cases and 5 evaluation sessions. Progressing through this empirical foundation, it turned out that the study was getting closer to a point of theoretical saturation. Due to the novelty of the subject of Data Mesh architectures, it is unrealistic to assume full theoretical saturation after this research. During the research process, attempts were made to continuously evaluate and refine new additions to the model, in order to increase the completeness of the model. However, it is very difficult to give a complete representation of reality within a new topic such as this one. Certainly because this research is largely based on empirical evidence, it is quite possible that the readiness steps and the factors in the model are not an exhaustive set of readiness factors. A larger number of participants in the study and additional iterative evaluation and refinement rounds could increase the completeness of research within the Data Mesh topic in the future.

8.4 RECOMMENDATIONS FOR FUTURE RESEARCH

First, the model in this study has been demonstrated in a number of different environments. Since the exploratory interviews and the case studies have shown that the readiness of organizations depends on contextual influences, it would be of added value to apply the model in more different contexts. Involving more different use cases could contribute to a broadened understanding of the readiness of organizations towards Data Mesh migration and how the influencing factors relate to it. For example, an application within a risk-averse and heavily government-regulated sector such as healthcare could provide interesting insights. The contexts within this research were often less regulated and more data-driven.

Second, an empirical follow-up study could focus more on recruiting more critical participants in the study. Because the topic is currently very popular among data-driven enthusiasts, it was difficult to find participants who could express themselves critically towards the topic and who also had empirical evidence for this. Follow-up research could take place in the future if there are more practical examples of Data Mesh implementations and their critics.

Third, follow-up research could focus on the factors that had little consensus during the case studies, such as establishing an appropriate governance model for Data Mesh architectures or addressing the decentralization and decoupling process. While this study has provided an overview of the as-is situation of participating organizations and offered different perspectives, follow-up research can delve deeper into these topics and explore the best ways to address these factors. For example, a Design Science Research into the design and development of a Data Mesh governance model could be of added value.

Fourth, follow-up research could focus on designing and developing an execution framework for implementing a Data Mesh architecture. While this research mainly focused on readiness and preparation prior to implementation, follow-up research can take a look at an actual implementation process of a Data Mesh architecture. Specific attention could be paid to certain implementation challenges that flow from the readiness assessment, or tackling the identified areas of improvement in a practical sense.

Finally, more research on the societal and organizational impact caused by the implementation of a Data Mesh architecture is proposed. Because of the novelty of the subject and the socio-technical context in which it takes place, it is of great value to know what the specific societal and organizational changes are that are involved in the large-scale implementation of Data Mesh architectures. This research can provide more insight for organizations to assess whether they are suitable for this architecture design and then also provide insights into the future vision of distributed architecture designs and the impact it has on society as a whole.

9 REFLECTION

9.1 REFLECTION ON THE PROCESS

First, I will reflect on the process leading me to the outcomes of this research. I believe that the choice for the Design Science Research approach was the right one, since it appears to be most suitable for an exploratory research like this one. I enjoyed researching the topic of Data Mesh architectures very much, since it gave me the opportunity to speak with a lot of experts from all around the world. The insights they've given me and the lessons I've learned from them give me an additional knowledge that I can carry with me for the upcoming years. As for the things I would do differently, if I were to redo this research, I would come to the following four points.

- First, I would take more time in the first few weeks to define the research problem and the facets around it. Out of enthusiasm I immediately started determining the necessary literature, interviews, and other matters. It might have been better to scope the problem clearly from the start so that I didn't have to spend more time on this later.
- Second, I would be more critical of the information that I would include in the research. I noticed that - especially due to the lack of literature that focuses specifically on the Data Mesh architecture - I occasionally got lost in information. I found it difficult to determine whether it was relevant or not, so I tried to include it all. In retrospect, I think it's more valuable to be critical of the information you send yourself during the investigation from the start, and it's best to make choices from the start.
- Third, I would also be more selective about the amount of empirical data I needed for my interview. Because I was completely taken in by the interviews, the interesting people, and the rich insights they gave me, in retrospect I may have involved too many people in my research. This meant that I was too often working late at night to process and analyze all this empirical data. Keeping in mind the theoretical saturation I was looking for, it may not have been necessary to speak to more people from the same quarter: however interesting and fun it was to get to know all these people.
- Fourth, I've noticed that I'm good at conducting research on my own, but maybe too well. Because of my partnership with Deloitte Consulting and the resources they had to offer me, I think in hindsight I could and should have made more use of it. It now took me a lot longer to gather all my information from outside Deloitte, when it might have been easier to occasionally look for it internally. I noticed that I was afraid that my research would become too biased if I only got information from Deloitte, but in retrospect I should have been a little less critical about this.

Finally, what I found most difficult in this research was finding a consensus among all the people involved in my research. Due to the highly exploratory nature of my research, I had to deal with a lot of empirical data on a completely new topic. This gave me many different perspectives, and it was sometimes a long search for a consensus or common thread in all these perspectives. This made it difficult to conceptualize the factors into an overarching idea, and to make a synthesis that was useful for my research. This meant that I often spent a long time looking for a conclusion, or that I had to contact my interview participants again to explain certain ideas. Again, this took a lot of time, and I hope I learned enough from this to be able to do this more efficiently in the future. Overall, I'm happy with the process of the research, and I've learned an incredible amount from it.

9.2 REFLECTION ON THE FINDINGS

For the reflection on the findings of this study, I discuss the main and most striking results of this study. I will discuss these results on the basis of the various components of the model. First, when it comes to assessing the Need for a Data Mesh architecture, it has been found that it is more difficult than expected to identify this specific need for organizations that have not yet thought of a Data Mesh solution. Many organizations are not yet data mature enough to know all the ins and outs of the overall functioning of their data architecture, which makes identifying specific bottlenecks and focus areas complicated. The main need pointing to a Data Mesh turned out to be the need to bring the analytic and operational side of the data platform together, so that there is more awareness among these people of the specific responsibilities they bear. Consequently, this would allow for more efficient and improved collaboration, which can exist with fewer dependencies. In addition, a need for improved scalability is a clear need that can be addressed by a Data Mesh architecture.

With regard to the ability to migrate to a Data Mesh, the biggest challenge for organizations is to be able to decompose their monolith in domain teams, and this is also a very important factor in determining readiness. Organizations that fail to gain readiness towards a Data Mesh architecture often fail to identify decentralized teams. There are several ways to deal with this decomposition, but it is certain that a lot of attention should be paid to it.

With regard to the preparedness for a Data Mesh architecture, the importance of a clear division of roles and responsibilities emerged. This is often underestimated by organizations - also by organizations that already have a higher readiness level - and it is therefore often a point where organizations still have to implement improvements afterwards. Overall, this research has shown that achieving Data Mesh readiness requires organizations to map out a lengthy and costly process. As a researcher, I have a critical position in this: I don't think a Data Mesh architecture is suitable for all organizations, and I am afraid that organizations will start the process without being suitable for it. I have seen that a Data Mesh architecture mainly offers a solution for large and complex organizations that are digital native, have an established Agile way of working and have a clear innovation culture incorporated. And while I can imagine that there are a lot of organizations that want to become more data-driven in the future, I also believe that for many of these organizations there are cheaper and less complex ways to address this need.

9.3 REFLECTION ON THE DMRM

Regarding the reflection on the DMRM, one of the main insights has to do with the intended iterative use of the model. While most readiness assessments are static checklists that give some sort of quantifiable result after completion, that is not the case with this model. Accordingly, during the design & development process, it turned out to be unrealistic to attach a quantitative outcome to an assessment, because the way in which the assessment is expressed depends very much on the organization in question. Where one organization faces major problems in the division of roles and tasks, the other organization will experience more problems with the central Business & IT alignment. For the first organization in question, it is therefore more relevant to pay attention to a renewed appointment of roles, while the second organization attaches more value to achieving data awareness among all teams in the organization. This insight also underlines the iterative nature of the instrument: because a Data Mesh migration takes place over several years in a constantly changing environment, it is very important to frequently assess the performance towards readiness. A Data Mesh migration is a complex and organization-wide transformation that requires structure and guidance. The DMRM seeks to provide this guidance by assessing key organizational factors and offering insights into the organization's position in relation to these factors.

Finally, I would like to add one more critical note to the DMRM. The topic of Data Mesh is currently very popular, and organizations are trying from all directions to achieve the 'perfect' Data Mesh. Although the opinions among the participants of this study differ about the extent to which an organization should implement all principles in detail, I think it is important as an organization to attain a critical attitude towards this. Perhaps you as an organization do not need all components, if a partial Data Mesh could serve as the solution for your needs. The DMRM supports organizations in this by leaving it up to the organizations themselves which factors and components of the model are relevant to them. During the assessment it will become clear which improvement areas will be discussed, and it is up to the organizations themselves to determine whether they want to tackle all these improvement areas. I think that the ideas surrounding the Data Mesh architecture, and the purpose that this new design of data architecture has, will prove valuable in the future. But for now it is important not to blindly adopt such ideas, but to position yourself as an organization within the hype and to only incorporate the principles that could be of value.

REFERENCES

Agarwal, R., & Prasad, J. (1997). The Role of Innovation Characteristics and Perceived Voluntariness in the Acceptance of Information Technologies. *Decision Sciences*, *28*(3), 557–582. Retrieved 2022-04-08, from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-5915.1997.tb01322.x (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-5915.1997.tb01322.x) doi: 10.1111/j.1540-5915.1997.tb01322.x

Al-Ammary, J., & Saleh, Z. (2021, February). Assessing the Readiness for Cloud Computing in Higher Education Institutions in the Kingdom of Bahrain: Towards an Education Cloud Computing Strategy. *International Journal of Managing Information Technology*, *13*(1), 33–54. Retrieved 2022-03-17, from https://aircconline.com/ijmit/V13N1/13121ijmit03.pdf doi: 10.5121/ijmit.2021.13103

Al-Omari, A., & Al-Omari, H. (2006, November). E-Government Readiness Assessment Model. *Journal of Computer Science*, *2*(11), 841–845. Retrieved 2022-03-08, from http://www.thescipub.com/abstract/?doi=jcssp.2006.841.845 doi: 10.3844/jcssp.2006.841.845

Atwal, H. (2020). The DataOps Factory. In H. Atwal (Ed.), *Practical DataOps: Delivering Agile Data Science at Scale* (pp. 249–266). Berkeley, CA: Apress. Retrieved 2022-03-08, from https://doi.org/10.1007/978-1-4842-5104-1_10 doi: 10.1007/978-1-4842-5104-1$_1$0

Avci Salma, C., Tekinerdogan, B., & Athanasiadis, I. N. (2017). Domain-Driven Design of Big Data Systems Based on a Reference Architecture. In *Software Architecture for Big Data and the Cloud* (pp. 49–68). Elsevier. Retrieved 2022-05-11, from https://linkinghub.elsevier.com/retrieve/pii/B9780128054673000041 doi: 10.1016/B978-0-12-805467-3.00004-1

Bahadorpoor, Z., Tajafari, M., & Sanatjoo, A. (2018, February). Implementation of 5S Methodology in Public Libraries: Readiness Assessment. *Library Philosophy and Practice*, *2018*.

Balalaie, A., Heydarnoori, A., & Jamshidi, P. (2016, May). Microservices Architecture Enables DevOps: Migration to a Cloud-Native Architecture. *IEEE Software*, *33*(3), 42–52. (Conference Name: IEEE Software) doi: 10.1109/MS.2016.64

Balasubramanian, S., Shukla, V., Sethi, J. S., Islam, N., & Saloum, R. (2021, April). A readiness assessment framework for Blockchain adoption: A healthcare case study. *Technological Forecasting and Social Change*, *165*, 120536. Retrieved 2022-03-17, from https://linkinghub.elsevier.com/retrieve/pii/S0040162520313627 doi: 10.1016/j.techfore.2020.120536

Barham, H., & Daim, T. (2020, September). The use of readiness assessment for big data projects. *Sustainable Cities and Society*, *60*, 102233. Retrieved 2022-03-15, from https://linkinghub.elsevier.com/retrieve/pii/S2210670720302201 doi: 10.1016/j.scs.2020.102233

Becker, J., Knackstedt, R., & Pöppelbuß, J. (2009, June). Developing Maturity Models for IT Management: A Procedure Model and its Application. *Business & Information Systems Engineering*, *1*(3), 213–222. Retrieved 2022-03-03, from http://link.springer.com/10.1007/s12599-009-0044-5 doi: 10.1007/s12599-009-0044-5

Blanco, P., Kotermanski, R., & Merson, P. (2007, September). *Evaluating a Service-Oriented Architecture:* (Tech. Rep.). Fort Belvoir, VA: Defense Technical Information Center. Retrieved 2022-05-04, from http://www.dtic.mil/docs/citations/ADA475130 doi: 10.21236/ADA475130

Braun, S., Bieniusa, A., & Elberzhager, F. (2021, April). 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* (pp. 1–12). Online United Kingdom: ACM. Retrieved 2022-02-21, from https://dl.acm.org/doi/10.1145/3447865.3457969 doi: 10.1145/3447865.3457969

Buono, A., & Kerber, K. (2009). *Intervention and Organizational Change: Building Organizational Change Capacity*.

Chanyagorn, P., & Kungwannarongkun, B. (2011). ICT Readiness Assessment Model for Public and Private Organizations in Developing Country. *International Journal of Information and Education Technology*, 99–106. Retrieved 2022-03-17, from http://www.ijiet.org/index.php?m=content&c=index&a=show&catid=23&id=19 doi: 10.7763/IJIET.2011.V1.17

Chen, D., Doumeingts, G., & Vernadat, F. (2008, September). Architectures for enterprise integration and interoperability: Past, present and future. *Computers in Industry*, *59*(7), 647–659. Retrieved 2022-02-17, from https://linkinghub.elsevier.com/retrieve/pii/S0166361508000365 doi: 10.1016/j.compind.2007.12.016

Cruz-Jesus, F., Pinheiro, A., & Oliveira, T. (2019, August). Understanding CRM adoption stages: empirical analysis building on the TOE framework. *Computers in Industry*, *109*, 1–13. Retrieved 2022-03-17, from https://linkinghub.elsevier.com/retrieve/pii/S0166361518305438 doi: 10.1016/j.compind.2019.03.007

Dang, D., & Pekkola, S. (2016). *ROOT CAUSES OF ENTERPRISE ARCHITECTURE PROBLEMS IN THE PUBLIC SECTOR.*

DataKitchen. (2021). *What is DataOps? | DataKitchen.* Retrieved 2022-02-16, from https://datakitchen.io/what-is-dataops/

De Bruin, T., Rosemann, M., Freeze, R., & Kaulkarni, U. (2005). Understanding the main phases of developing a maturity assessment model. In *Australasian Conference on Information Systems (ACIS)* (pp. 8–19). Australasian Chapter of the Association for Information Systems.

Dehghani, Z. (2019). *How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh* [Martin Fowler]. Retrieved from https://martinfowler.com/articles/data-monolith-to-mesh.html

Dehghani, Z. (2022). *Data Mesh: Delivering Data-Driven Value at Scale*. O'Reilly Media. Retrieved 2022-02-15, from https://www.oreilly.com/library/view/data-mesh/9781492092384/ (ISBN: 9781492092391)

Dragoni, N., Giallorenzo, S., Lafuente, A. L., Mazzara, M., Montesi, F., Mustafin, R., & Safina, L. (2017, April). Microservices: yesterday, today, and tomorrow. *arXiv:1606.04036 [cs]*. Retrieved 2022-05-04, from http://arxiv.org/abs/1606.04036 (arXiv: 1606.04036)

Einav, L., & Levin, J. (2014). *The Data Revolution and Economic Analysis.* Retrieved 2022-02-15, from https://www.journals.uchicago.edu/doi/epdf/10.1086/674019 doi: 10.1086/674019

Engström, A. (2020, May). Towards Agile Data Engineering for Small Scale Teams. , 78.

Enyo-one Musa, A. (2021). *Data Management Arcitectures - Monolithic Data Architectures vs Distributed Data Mesh.* Retrieved from https://towardsdatascience.com/data-management-architectures-monolithic-data-architectures-and-distributed-data-mesh-63743794966c

Ereth, J. (2018). DataOps – Towards a Definition. , 9.

Ferreira, C., Andrade, P., & Almeida, F. (2020, August). How to Improve the Validity and Reliability of a Case Study Approach. *Journal of Interdisciplinary Studies in Education*, *9*, 273–284. doi: 10.32674/jise.v9i2.2026

Fleury, J. H. (2021). *Data Mesh is a Solution for Functional Organizational Challenges* [Crystalloids]. Retrieved from https://www.crystalloids.com/news/data-mesh-solution-to-functional-challenges

Flyvbjerg, B. (2006, April). Five Misunderstandings About Case-Study Research. *Qualitative Inquiry*, *12*(2), 219–245. Retrieved 2022-04-20, from http://journals.sagepub.com/doi/10.1177/1077800405284363 doi: 10.1177/1077800405284363

Fraser, P., Moultrie, J., & Gregory, M. (2002, August). The use of maturity models/grids as a tool in assessing product development capability. In *IEEE International Engineering Management Conference* (Vol. 1, pp. 244–249 vol.1). doi: 10.1109/IEMC.2002.1038431

Furia, B. (2021, March). *Data Mesh – Rethinking Enterprise Data Architecture.* Retrieved 2022-02-15, from https://www.cuelogic.com/blog/data-mesh

Gangwar, H., Date, H., & Ramaswamy, R. (2015). Understanding determinants of cloud computing adoption using an integrated TAM-TOE model. *Emerald Insight*. Retrieved 2022-03-17, from https://www.emerald.com/insight/content/doi/

10.1108/JEIM-08-2013-0065/full/html?casa_token=zYJ1WkwQdycAAAAA:bW2-N-7pWafaV3GSgGEHAZ8uEb2LHdTxxuODbIuqpUkaNOxubAOvTN4cI40wY2DPe1gq4kxIfIfqK_Kn-6hd5zKQXDAFB2WeClGqMnOl9WmHaHFSkxk

Genovese, S. (2021). *Data Mesh: the newest paradigm shift for a distributed architecture in the data world and its application* (Doctoral dissertation, Politecnico di Torino, Torino). Retrieved from https://webthesis.biblio.polito.it/20415/

Giebler, C., Gröger, C., Hoos, E., Eichler, R., Schwarz, H., & Mitschang, B. (n.d.). The Data Lake Architecture Framework: A Foundation for Building a Comprehensive Data Lake Architecture. , 21.

Glaser, B. G., & Strauss, A. L. (1967). Grounded theory: Strategies for qualitative research. *Chicago, IL: Aldine Publishing Company*.

Goetz, M. (2022, March). *Exposing The Data Mesh Blind Side.* Retrieved 2022-05-05, from https://www.forrester.com/blogs/exposing-the-data-mesh-blind-side/

Goniwada, S. R. (2022). *Cloud Native Architecture and Design: A Handbook for Modern Day Architecture and Design with Enterprise-Grade Examples*. Berkeley, CA: Apress. Retrieved 2022-03-07, from https://link.springer.com/10.1007/978-1-4842-7226-8 doi: 10.1007/978-1-4842-7226-8

Gouigoux, J.-P., Tamzalit, D., & Noppen, J. (2021). Microservice Maturity of Organizations: Towards an Assessment Framework. In S. Cherfi, A. Perini, & S. Nurcan (Eds.), *Research Challenges in Information Science* (Vol. 415, pp. 523–540). Cham: Springer International Publishing. Retrieved 2022-03-18, from https://link.springer.com/10.1007/978-3-030-75018-3_34 (Series Title: Lecture Notes in Business Information Processing) doi: 10.1007/978-3-030-75018-3$_3$4

Griggs, S. (1987). Analysing qualitative data. *Journal of the Market Research Society*, *29*(1), 15–34. (Place: United Kingdom Publisher: Market Research Society)

Hagiu, A., & Wright, J. (2020). When Data Creates Competitive Advantage... *S T R AT E GY*, 9.

Harrison, T. (2014). Rethinking Readiness. *Strategic Studies Quarterly*, *8*(3), 38–68. Retrieved 2022-04-12, from https://www.jstor.org/stable/26270619 (Publisher: Air University Press)

Hazel, T. (n.d.). *Data lake vs. data mesh: Which one is right for you?* Retrieved 2022-03-11, from https://www.chaossearch.io/blog/data-lake-vs-data-mesh

Henry, A., & Ridene, Y. (2020). Assessing your microservice migration. *Microservices*, 73–107.

Hevner, A. R. (2010). A Three Cycle View of Design Science Research. , 6.

Hokkanen, S. (2021). *Utilization of Data Mesh Framework as a Part of Organization's Data Management* (Unpublished doctoral dissertation).

Hussein, S., Mahrin, M., Maarop, N., & Abu Bakar, N. A. (2020, February). Development and Validation of Enterprise Architecture (EA) Readiness Assessment Model. *International Journal on Advanced Science, Engineering and Information Technology*, *10*, 157. doi: 10.18517/ijaseit.10.1.9007

Hyperight. (2021, January). *Is Data Mesh right for your organisation?* Retrieved 2022-02-16, from https://hyperight.com/is-data-mesh-right-for-your-organisation/

Jahani, B., Reza Seyyed Javadein, S., & Abedi Jafari, H. (2010, January). Measurement of enterprise architecture readiness within organizations. *Business Strategy Series*, *11*(3), 177–191. Retrieved 2022-03-17, from https://doi.org/10.1108/17515631011043840 (Publisher: Emerald Group Publishing Limited) doi: 10.1108/17515631011043840

Johannesson, P., & Perjons, E. (2021). *An Introduction to Design Science*. Cham: Springer International Publishing. Retrieved 2022-02-15, from https://link.springer.com/10.1007/978-3-030-78132-3 doi: 10.1007/978-3-030-78132-3

Joshi, D., Pratik, S., & Podila, M. (2021). *Data Governance in Data Mesh Infrastructures: the Saxo Bank Case Study* (Tech. Rep.). EasyChair.

Kalske, M., Mäkitalo, N., & Mikkonen, T. (2018). Challenges When Moving from Monolith to Microservice Architecture. In I. Garrigós & M. Wimmer (Eds.), *Current Trends in Web Engineering* (Vol. 10544, pp. 32–47). Cham: Springer International Publishing. Retrieved 2022-02-16, from http://link.springer.com/10.1007/978-3-319-74433-9_3 (Series Title: Lecture Notes in Computer Science) doi: 10.1007/978-3-319-74433-9$_3$

Khazaei, H., Barna, C., Beigi-Mohammadi, N., & Litoiu, M. (2016, December). Efficiency Analysis of Provisioning Microservices. In *2016 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)* (pp. 261–268). (ISSN: 2330-2186) doi: 10.1109/CloudCom.2016.0051

Kitchenham, B., & Charters, S. (2007, January). Guidelines for performing Systematic Literature Reviews in Software Engineering. , *2*.

Klein, H. K., & Myers, M. D. (1999). A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems. *MIS Quarterly*, *23*(1), 67–93. Retrieved 2022-06-06, from https://www.jstor.org/stable/249410 (Publisher: Management Information Systems Research Center, University of Minnesota) doi: 10.2307/249410

Kotorchevikj, I. (2021). *What Is Data Mesh? And Should You Mesh It Up Too? | by Ivana Kotorchevikj | Towards Data Science.* Retrieved 2022-02-16, from https://towardsdatascience.com/what-is-data-mesh-and-should-you-mesh-it-up-too-364b28fe2ae9

Krivic, P., Skocir, P., Kusek, M., & Jezic, G. (2018). *Microservices as Agents in IoT Systems.* (Pages: 31) doi: 10.1007/978-3-319-59394-4$_3$

Kulkarni, R. (2019). *Big Data Goes Big.* Retrieved 2022-02-21, from https://www.forbes.com/sites/rkulkarni/2019/02/07/big-data-goes-big/ (Section: Fintech)

Lawal Moshood, O., Ileladewa Adeoye, A., & Lawal Habibu, O. (2020). Microservice Architecture: A Better Option over Monolithic Architecture. *Technology (ICONSEET)*, *5*(19), 148–156.

Lee, Y., Kozar, K. A., & Larsen, K. R. (2003). The Technology Acceptance Model: Past, Present, and Future. *Communications of the Association for Information Systems*, *12*. Retrieved 2022-03-17, from https://aisel.aisnet.org/cais/vol12/iss1/50 doi: 10.17705/1CAIS.01250

Lnenicka, M., & Komarkova, J. (2019, June). Developing a government enterprise architecture framework to support the requirements of big and open linked data with the use of cloud computing. *International Journal of Information Management*, *46*, 124–141. Retrieved 2022-02-17, from https://www.sciencedirect.com/science/article/pii/S0268401218305942 doi: 10.1016/j.ijinfomgt.2018.12.003

Losey, R. (2004). *Enterprise Data Architecture: Who Needs it?* Retrieved from http://www.information-management.com/issues/20040101/7910-1.html?zkPrintable=true

Loukiala, A., Joutsenlahti, J.-P., Raatikainen, M., Mikkonen, T., & Lehtonen, T. (2021). Migrating from a Centralized Data Warehouse to a Decentralized Data Platform Architecture. In L. Ardito, A. Jedlitschka, M. Morisio, & M. Torchiano (Eds.), *Product-Focused Software Process Improvement* (Vol. 13126, pp. 36–48). Cham: Springer International Publishing. Retrieved 2022-03-18, from https://link.springer.com/10.1007/978-3-030-91452-3_3 (Series Title: Lecture Notes in Computer Science) doi: 10.1007/978-3-030-91452-3$_3$

Machado, I., Costa, C., & Santos, M. Y. (2021, August). Data-Driven Information Systems: The Data Mesh Paradigm Shift. *International Conference on Information Systems Development (ISD)*. Retrieved from https://aisel.aisnet.org/isd2014/proceedings2021/currenttopics/9

Machado, I. A., Costa, C., & Santos, M. Y. (2022a). Data Mesh: Concepts and Principles of a Paradigm Shift in Data Architectures. *Procedia Computer Science*, *196*, 263–271. Retrieved 2022-02-21, from https://linkinghub.elsevier.com/retrieve/pii/S1877050921022365 doi: 10.1016/j.procs.2021.12.013

Machado, I. A., Costa, C., & Santos, M. Y. (2022b). Data Mesh: Concepts and Principles of a Paradigm Shift in Data Architectures. *Procedia Computer Science*, *196*, 263–271. Retrieved 2022-02-15, from https://linkinghub.elsevier.com/retrieve/pii/S1877050921022365 doi: 10.1016/j.procs.2021.12.013

Marr, B. (2018). *How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read.* Retrieved 2022-02-21, from https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/ (Section: Enterprise & Cloud)

Mehmandarov, R., Waaler, A., Cameron, D., Fjellheim, R., & Pettersen, T. B. (2021, December). A Semantic Approach to Identifier Management in Engineering Systems. In *2021 IEEE International Conference on Big Data (Big Data)* (pp. 4613–4616). doi: 10.1109/BigData52589.2021.9671515

Mirarab, A., Fard, N. G., & Kenari, A. R. R. (2014). Soa Readiness Assessment, a New Method. , *4*(8), 10.

Mora, M., Gelman, O., Steenkamp, A. L., & Raisinghani, M. (Eds.). (2012). *Research Methodologies, Innovations and Philosophies in Software Systems Engineering and Information Systems:*. IGI Global. Retrieved 2022-03-03, from http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-4666-0179-6 doi: 10.4018/978-1-4666-0179-6

Moses, B. (2020). *What is a Data Mesh — and How Not to Mesh it Up | by Barr Moses | Towards Data Science.* Retrieved 2022-02-21, from https://towardsdatascience.com/what-is-a-data-mesh-and-how-not-to-mesh-it-up-210710bb41e0

Munappy, A. R., Bosch, J., & Olsson, H. H. (2020). Data Pipeline Management in Practice: Challenges and Opportunities. In M. Morisio, M. Torchiano, & A. Jedlitschka (Eds.), *Product-Focused Software Process Improvement* (Vol. 12562, pp. 168–184). Cham: Springer International Publishing. Retrieved 2022-03-08, from https://link.springer.com/10.1007/978-3-030-64148-1_11 (Series Title: Lecture Notes in Computer Science) doi: 10.1007/978-3-030-64148-1$_1$1

Munappy, A. R., Mattos, D. I., Bosch, J., Olsson, H. H., & Dakkak, A. (2020, June). From Ad-Hoc Data Analytics to DataOps. In *Proceedings of the International Conference on Software and System Processes* (pp. 165–174). Seoul Republic of Korea: ACM. Retrieved 2022-03-08, from https://dl.acm.org/doi/10.1145/3379177.3388909 doi: 10.1145/3379177.3388909

Nadareishvili, I., Mitra, R., McLarty, M., & Amundsen, M. (2016). *Microservice Architecture: Aligning Principles, Practices, and Culture.* "O'Reilly Media, Inc.". (Google-Books-ID: Ev2wDAAAQBAJ)

Nair, G. S., & Riege, A. M. (1995). Using convergent interviewing to develop the research problem of a postgraduate thesis. In *Proceedings of Marketing Education and Researchers International Conference, Gold Coast.*

Namiot, D., & Sneps-Sneppe, M. (2014, August). On Micro-services Architecture. *International Journal of Open Information Technologies*, *2*(9), 24–27. Retrieved 2022-02-15, from http://injoit.org/index.php/j1/article/view/139 (Number: 9)

Newman, P., & Kotonya, G. (2015, June). A resource-aware framework for resource-constrained service-oriented systems. *Future Generation Computer Systems*, *47*, 161–175. Retrieved 2022-02-21, from https://www.sciencedirect.com/science/article/pii/S0167739X14001800 doi: 10.1016/j.future.2014.09.010

Niemi, E., & Ylimäki, T. (2007). Enterprise Architecture Evaluation Components. , 12.

Okorie Awa, H., Ukoha, O., & Emecheta, B. (2012). Integrating TAM and TOE Frameworks and Expanding their Characteristic Constructs for E-Commerce Adoption by SMEs. In (pp. 571–588). Retrieved 2022-04-08, from https://www.informingscience.org/Publications/1676 doi: 10.28945/1676

Oreščanin, D., & Hlupić, T. (2021, September). Data Lakehouse - a Novel Step in Analytics Architecture. In *2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO)* (pp. 1242–1246). (ISSN: 2623-8764) doi: 10.23919/MIPRO52101.2021.9597091

Ortiz, G., Boubeta-Puig, J., Criado, J., Corral-Plaza, D., Garcia-de Prado, A., Medina-Bulo, I., & Iribarne, L. (2022, April). A microservice architecture for real-time IoT data processing: A reusable Web of things approach for smart ports. *Computer Standards & Interfaces*, *81*, 103604. Retrieved 2022-02-15, from https://www.sciencedirect.com/science/article/pii/S0920548921000994 doi: 10.1016/j.csi.2021.103604

Pellettiere, V. (2006). Organization Self-Assessment to Determine the Readiness and Risk for a Planned Change. *Organization Development Journal*, *24*(4), 38.

Ponce, F., Márquez, G., & Astudillo, H. (2019, November). Migrating from monolithic architecture to microservices: A Rapid Review. In *2019 38th International Conference of the Chilean Computer Science Society (SCCC)* (pp. 1–7). (ISSN: 1522-4902) doi: 10.1109/SCCC49216.2019.8966423

Priebe, T., Neumaier, S., & Markus, S. (2021, December). Finding Your Way Through the Jungle of Big Data Architectures. In *2021 IEEE International Conference on Big Data (Big Data)* (pp. 5994–5996). doi: 10.1109/BigData52589.2021.9671862

Pruijt, L., Slot, R., Plessius, H., Bos, R., & Brinkkemper, S. (2012). *The Enterprise Architecture Realization Scorecard: A Result Oriented Assessment Instrument* (Vol. 131). (Journal Abbreviation: Lecture Notes in Business Information Processing Publication Title: Lecture Notes in Business Information Processing) doi: 10.1007/978-3-642-34163-2$_1$8

Ray, P., & Pal, P. (2020, November). Extending the SEMAT Kernel for the Practice of Designing and Implementing Microservice-Based Applications using Domain Driven Design. In *2020 IEEE 32nd Conference on Software Engineering Education and Training (CSEE T)* (pp. 1–4). (ISSN: 2377-570X) doi: 10.1109/CSEET49119.2020.9206200

Riege, A. M. (2003, January). Validity and reliability tests in case study research: a literature review with "hands-on" applications for each research phase. *Qualitative Market Research: An International Journal*, *6*(2), 75–86. Retrieved 2022-06-06, from https://doi.org/10.1108/13522750310470055 (Publisher: MCB UP Ltd) doi: 10.1108/13522750310470055

Rigol, X. G. (2021, July). *Data as a product vs data products. What are the differences?* Retrieved 2022-02-18, from https://towardsdatascience.com/data-as-a-product-vs-data-products-what-are-the-differences-b43ddbb0f123

Rodriguez, M., de Araújo, L. J. P., & Mazzara, M. (2020, December). Good practices for the adoption of DataOps in the software industry. *Journal of Physics: Conference Series*, *1694*(1), 012032. Retrieved 2022-03-08, from https://iopscience.iop.org/article/10.1088/1742-6596/1694/1/012032 doi: 10.1088/1742-6596/1694/1/012032

Rosas, J., & Camarinha-Matos, L. M. (2009, September). An approach to assess collaboration readiness. *International Journal of Production Research*, *47*(17), 4711–4735. Retrieved 2022-04-08, from https://www.tandfonline.com/doi/full/10.1080/00207540902847298 doi: 10.1080/00207540902847298

Saddad, E., El-Bastawissy, A., M., H., & Hazman, M. (2020). Lake Data Warehouse Architecture for Big Data Solutions. *International Journal of Advanced Computer Science and Applications*, *11*(8). Retrieved 2022-02-15, from http://thesai.org/Publications/ViewPaper?Volume=11&Issue=8&Code=IJACSA&SerialNo=54 doi: 10.14569/IJACSA.2020.0110854

Sahoo, P. R. (2019, January). DataOps in Manufacturing and Utilities Industries. *International Journal of Applied Information Systems*, *12*, 6.

Saransig, A., & Tapia Leon, F. (2019, January). Performance Analysis of Monolithic and Micro Service Architectures – Containers Technology: Proceedings of the 7th International Conference on Software Process Improvement (CIMPS 2018). In (pp. 270–279). doi: 10.1007/978-3-030-01171-0$_2$5

Saurabh, S. (2021). *Data Mesh: Design, Benefits, Hype, and Reality | LinkedIn.* Retrieved 2022-05-05, from https://www.linkedin.com/pulse/data-mesh-design-benefits-hype-reality-saket-saurabh/

Schultze, M. (2020). *Data Mesh in Practice: A Zalando Case Study.* DataBricks Summit 2020.

Selmadji, A., Seriai, A.-D., Bouziane, H. L., Oumarou Mahamane, R., Zaragoza, P., & Dony, C. (2020, March). From Monolithic Architecture Style to Microservice one Based on a Semi-Automatic Approach. In *2020 IEEE International Conference on Software Architecture (ICSA)* (pp. 157–168). doi: 10.1109/ICSA47634.2020.00023

Serra, J. (2021, July). *Data Mesh: Centralized ownership vs decentralized ownership | James Serra's Blog.* Retrieved 2022-03-08, from https://www.jamesserra.com/archive/2021/07/data-mesh-centralized-ownership-vs-decentralized-ownership/

Smith, I. (2005, January). Achieving readiness for organisational change. *Library Management*, *26*(6/7), 408–412. Retrieved 2022-04-12, from https://doi.org/10.1108/01435120510623764 (Publisher: Emerald Group Publishing Limited) doi: 10.1108/01435120510623764

Solano, M. A., & Jernigan, G. (2012, July). Enterprise data architecture principles for High-Level Multi-Int fusion: A pragmatic guide for implementing a heterogeneous data exploitation framework. In *2012 15th International Conference on Information Fusion* (pp. 867–874).

Stalla-Bourdillon, S., Wintour, A., & Carmichael, L. (2019, December). *Building trust through data foundations: a call for a data governance model to support trustworthy data sharing* [Monograph]. Retrieved 2022-02-15, from https://eprints.soton.ac.uk/443715/ (Num Pages: 31 Publisher: University of Southampton) doi: 10.5258/SOTON/WSI-WP002

Stoianova, O., Lezina, T. A., & Ivanova, V. V. (2020, July). The framework for assessing company's digital transformation readiness. - ., *36*(2), 243–265. Retrieved 2022-04-14, from https://economicsjournal.spbu.ru (Number: 2) doi: 10.21638/spbu05.2020.204

TOGAF. (2018). *Business Transformation Readiness Assessment.* Retrieved 2022-03-09, from https://pubs.opengroup.org/architecture/togaf9-doc/m/chap26.html

Tornatzky, L. G., Fleischer, M., & Chakrabarti, A. K. (1990). *processes of technological innovation* (Tech. Rep.). Lexington Books. Retrieved 2022-03-17, from https://scholar.google.com/scholar_lookup?title=processes+of+technological+innovation&author=Tornatzky%2C+Louis+G.&publication_year=1990 (ISBN: 9780669203486)

Vakola, M. (2013, March). Multilevel Readiness to Organizational Change: A Conceptual Approach. *Journal of Change Management*, *13*(1), 96–109. Retrieved 2022-04-12, from http://www.tandfonline.com/doi/abs/10.1080/14697017.2013.768436 doi: 10.1080/14697017.2013.768436

van der Raadt, B., Bonnet, M., Schouten, S., & van Vliet, H. (2010, October). The relation between EA effectiveness and stakeholder satisfaction. *Journal of Systems and Software*, *83*(10), 1954–1969. Retrieved 2022-04-14, from https://linkinghub.elsevier.com/retrieve/pii/S0164121210001482 doi: 10.1016/j.jss.2010.05.076

Vasconcelos, A., Sousa, P., & Tribolet, J. (2007, January). Information System Architecture Metrics: an Enterprise Engineering Evaluation Approach. *Electronic Journal of Information Systems Evaluation*, *10*(1), pp91-122–pp91-122. Retrieved 2022-04-14, from https://academic-publishing.org/index.php/ejise/article/view/317 (Number: 1)

Velitchkov, I. (2008). Integration of IT strategy and enterprise architecture models. In *Proceedings of the 9th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing - CompSysTech '08* (p. V.7). Gabrovo, Bulgaria: ACM Press. Retrieved 2022-04-14, from http://portal.acm.org/citation.cfm?doid=1500879.1500955 doi: 10.1145/1500879.1500955

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425–478. Retrieved 2022-03-03, from https://www.jstor.org/stable/30036540 (Publisher: Management Information Systems Research Center, University of Minnesota) doi: 10.2307/30036540

Vernadat, F. B. (2003). Open Systems Architectures: From monolithic approaches to service- based architectures. , 6.

Waseem, M., Liang, P., Shahin, M., Di Salle, A., & Márquez, G. (2021, December). Design, monitoring, and testing of microservices systems: The practitioners' perspective. *Journal of Systems and Software*, *182*, 111061. Retrieved 2022-02-22, from https://linkinghub.elsevier.com/retrieve/pii/S0164121221001588 doi: 10.1016/j.jss.2021.111061

Waterman Jr, R., Peters, T. J., & Phillips, J. (1980). Structure is not organization. *Business Horizons*, *23*(3), 14–26.

Weisman, R. (2004). The Canadian Government and Business Transformation. , 33.

Wen, K., & Chen, Y. (2010). E-business value creation in Small and Medium Enterprises: a US study using the TOE framework. *International Journal of Electronic Business*, *8*(1), 80–100. Retrieved 2022-03-17, from https://www.inderscienceonline.com/doi/abs/10.1504/IJEB.2010.030717

Yin, R. K. (1994). Discovering the Future of the Case Study. Method in Evaluation Research. Retrieved 2022-06-06, from https://journals.sagepub.com/doi/abs/10.1177/109821409401500309?journalCode=ajeb

Yin, R. K. (2017). *Case Study Research and Applications* (6th ed.). COSMOS Corporation.

Zdun, U., Navarro, E., & Leymann, F. (2017). Ensuring and Assessing Architecture Conformance to Microservice Decomposition Patterns. In M. Maximilien, A. Vallecillo, J. Wang, & M. Oriol (Eds.), *Service-Oriented Computing* (Vol. 10601, pp. 411–429). Cham: Springer International Publishing. Retrieved 2022-02-15, from http://link.springer.com/10.1007/978-3-319-69035-3_29 (Series Title: Lecture Notes in Computer Science) doi: 10.1007/978-3-319-69035-3$_2$9

Zimmermann, A., Schmidt, R., Sandkuhl, K., Jugel, D., Bogner, J., & Möhring, M. (2018, October). Evolution of Enterprise Architecture for Digital Transformation. In *2018 IEEE 22nd International Enterprise Distributed Object Computing Workshop (EDOCW)* (pp. 87–96). (ISSN: 2325-6605) doi: 10.1109/EDOCW.2018.00023

A RECOMMENDATIONS FOR ORGANIZATIONS

This appendix contains the recommendations based on the results of the assessment. The recommendations are based on the best practices and insights of all participants of this research.

A.1 RECOMMENDATIONS ON THE ASSESSMENT RESULTS

Recommendations regarding Willingness to Change
Organizations that score low on the Willingness to Change dimension can improve themselves by starting with a clear understanding of the concept of Data Mesh and the value it brings to the entire organization. This understanding is supported by a supportive culture in the field of learning and innovation. This culture can be fostered by, for example, setting up extensive R&D teams, facilitating data literacy workshops in the organization, making data awareness training accessible to everyone and upskilling the domain teams so that they can create and maintain data products in the future. Ensure within the organization an appetite for organizational change, and leverage the right people who have the right influence throughout this process.

Recommendations regarding Vision
In the absence of a clear vision towards a Data Mesh architecture, it is could be useful to start with an appropriate data strategy for the organization. Within this strategy, tactical objectives and strategic objectives can be formulated. In addition, it is recommended to map out where redundant work is currently being performed. Organizations that want to become data-driven ultimately want to base their decision-making on facts and insights from the data. A marketplace of data could be established within the organization, where people can find and use data for these insights. In addition, it can be useful to measure the demand for value creation from different perspectives within the organization. Not only the business side of the organization should be involved, but also the operational teams. These teams should attain a focus on the consumers of their data, and the value they deliver to these consumers.

Recommendations regarding Business Case
In the absence of a clear Business & IT alignment, it is useful to reflect on when mapping the organization around each other's operations and values of the data. It is not recommended to look in a customer supply relationship between business and IT, but to change the perception towards an expectation-based relationship. What do the teams expect from each other, and how do they support each other? In this way, the operational teams can gain a better insight into what they contribute to business-wise. The alignment can also be found in improved collaboration, so that common interest in data can arise through improved data awareness. Programs as formulated in the objectives support this awareness. These programs can be initiated by the business users who want to become more actively involved in the processing of data.

Recommendations regarding Top-Management Engagement
In the absence of top-management engagement, an improved value statement of the Data Mesh migration for these executives can be looked into. Top-management support is recommended for an organization-wide migration, because they can translate the ambition across the entire organization. Additionally, their motivations and directions should be aligned with the focus areas of a Data Mesh

architecture. The focus areas of a Data Mesh lie in developing an architecture that is easy to evolve and access, and to better accommodate large amounts of data. Where there is a need for improved management of the complexity of the domains, a Data Mesh architecture can contribute to improving the understanding and trustworthiness of the data. Workshops or written value statements that showcase or describe these contributions can help create support among top management.

Recommendations regarding Enterprise Ability

With deficiencies in the enterprise ability to migrate, it is recommended to realize that decentralizing and decoupling current monolithic structures is a lengthy and iterative process. Recommendations within this process are, for example, to start identifying the domains, and to consider how these domains can be enabled to work autonomously. Not all domains have the capability to work autonomously, so it is recommended to start slowly with the domains that are the most mature and eager to be autonomous. These domains can serve as an example and pilot study. Outlining the business domains on a map can help with the decentralization process, for example based on existing domains that already own business artifacts or based on each vice president within the organization. The ownership boundaries are formulated so that the domains do not step on each other's toes.

In line with this, it is useful to realize that not every organization is able to decouple the central platforms. For these organizations, a logical split in the domains can be chosen, whereby each part of the central application belongs to one domain. Perform a translation which allows to decouple the operational plane from the analytical plan in this application, so that an abstraction layer is created between business application and analytical application.

Finally, the ability to work in domain teams can be increased by looking closely at the demand or data engineering skills in each domain. Not all domains have a constant workload over time, so at the beginning of the migration it is possible to work with, for example, a center of excellence of a pool of people who are able to jump into domain problems. It is recommended to retain a certain amount of domain knowledge within the data engineers. In addition, a migration program could be established that redesigns the entire organization based on domains and processes around the domains. This program is carried across the organization as a whole and ensures that the most important concepts - such as the definition of a data product, how you publish, how you manage data - are clear and understandable to the people who become part of a domain.

Recommendations regarding Clearly defined Roles and Responsibilities

Firstly, if there is a shortcoming within this dimension, it is recommended to realize that not everything has to be perfectly defined, but that there is a clear ownership alignment among the domains and the organization as a whole. A data product manager in any domain should have the information what the data product means, and know and understand the data product. This improves trust in data products between the different domains. Certain roles that could be divided in this context are, for example, those of a data owner in terms of business insights, a data custodian in terms of IT responsibilities and a data steward as data product manager and to bridge the gap between both teams.

Recommendations regarding Governance

It has become clear from this study that few recommendations can be made in the field of setting up the governance model correctly. Recommendations were to focus on making it easy to share and access data within the Data Mesh architecture, within the regulatory boundaries. Automation and tools help with this, but are not necessary. In addition, a central governance body is needed that can enforce the organization-wide standards and rules across the domains. Within the domains, governance must ensure that the data is correctly defined and placed within the correct business context. Possibilities for realizing democratized governance are, for example, with a governance forum in which the domains can provide input and be part of the decision-making process. Formal governance can be made available on this forum so that certain policies are discoverable. From the technology perspective, components can be built into the platform that can apply certain automated rules and policies.

Recommendations regarding Workable Approach

If there is a shortcoming within the workable approach dimension, the needs of the cross-domain collaboration could be looked at with regard to communication and collaboration. This can be done through data exchange, recurring meetings or collaboration agreements. With many inter-domain dependencies, data contracts or service level agreements can be drawn up to improve reliability. It is also possible to look at a place in the organization, physical or digital, where this inner- and cross-domain collaboration can take place.
With regard to data literacy, it depends on the choice of technology within the organization. With more complex technology choices, it is more important to have a greater degree of data literacy. You can also opt for more modern tools that do not require data engineers. Data literacy can be improved through a knowledge sharing platform, data literacy workshops or data awareness programs.

Recommendations regarding IT Capacity

In case of shortcomings in IT capacity, it is recommended to prioritize. Where continuous integration and automation standards have not been applied, it does not have the highest priority to develop them. Continuous deployment is more important when it comes to managing multiple data products. In addition, it is possible to look at developing approaches that counteract breaking use cases for downstream users, in case of changes to existing datasets. In addition, it is also up to the domains themselves to determine with which maturity in terms of CI/CD and automation they want to work on their inner-domain processes. Important to start with are data product publication standards. If they don't exist yet, they should be set up to structure this process. These standards contain, for example, APIs for the structure of data, data contracts for the trust in the data, a description of the data so that consumers can use the data, or a technology standard with which data products can be built.

Recommendations regarding Enterprise Capacity

In case of shortcomings in the Enterprise Capacity dimension, it is recommended to start by recording the new team structures. A certain alignment with the business domains is suggested, because in this way business domain knowledge can be preserved within the Data Mesh domains. For example, this domain knowledge can come in useful when feeling the responsibility about the domains, and it can be easier to feel this responsibility if it already existed for a certain part of the business. In addition, this way of structuring ensures that the Data Mesh domains are not placed too far from the business.

B | SYSTEMATIC LITERATURE REVIEW

This appendix describes the method used to find answers to the first four sub-questions. The method that is used in this thesis follows the iterative cycles of the stages define, *search*, and *select*. During the first stage, the scope of the research is defined and the inclusion and exclusion criteria of the scientific sources are formulated. Moreover, the search engines that are used were investigated and the search queries are stated. After that, the search is conducted. During the final stage, the sources that can be used for this study were selected. In order to keep track of the sources, reference manager Zotero^{TM} was used.

B.1 INCLUSION AND EXCLUSION CRITERIA

The inclusion and exclusion criteria of this search are specified to obtain relevant literature to answer the research question and decrease the bias in the search process (Kitchenham & Charters, 2007). Grey literature has also been incorporated in this search since Data Mesh is a very new concept and there is not much scientific literature written on the topic yet. However, since it is a very popular concept, a lot of tech forums contain expert blog-posts on the topic. According to Kitchenham and Charters (2007), grey literature has to be incorporated in a well-performed systematic literature review in order to increase its value. The inclusion and exclusion criteria are presented in table B.1.

Table B.1: Inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
English based literature	Non-English based literature
Scientific and grey literature	Duplicate literature
Literature on distributed data architectures	Literature 2000
Literature on domain-oriented data architectures	Literature that focuses on software development
Literature on assessment frameworks and models	Literature that is unrelated to the research questions

B.2 OVERVIEW OF THE SELECTED LITERATURE

Table B.2: Overview Selected Literature

Title	Reference	Main theme
Finding your Way through the Jungle of Big Data Architectures	Priebe \ Markus (2021)	Big Data
Data Mesh: the Newest Paradigm Shift for a Distributed Architecture in the Data World and its application	Genovese (2021)	Data Mesh
The Data Lake Architecture Framework: a Foundation for building a Comprehensive Data Lake Architecture	Giebler et al. (2020)	Data Mesh
Utilization of Data Mesh Framework as a part of Organization's Data Management	Hokkanen (2021)	Data Mesh
Data Governance in Data Mesh Infrastructures: the Saxo Bank Case Study	Joshi et al. (2021)	Data Mesh
Data Mesh: Concepts and Principles of a Paradigm Shift in Data Architectures	Machado et al. (2021)	Data Mesh
DataOps in Manufacturing and Utilities Industries	Sahoo (2019)	DataOps
Migrating from a Centralized Data Warehouse to a Decentralized Data Platform Architecture	Loukiala et al. (2021)	Distributed Architecture
Data Lakehouse - a Novel Step in Analytics Architectures	Orescanin \ Hlupic (2021)	Distributed Architecture
Evaluating a Service-Oriented Architecture	Blanco et al. (2007)	Distributed Architecture
Microservices: Yesterday, Today, and Tomorrow	Dragoni et al. (2017)	Distributed Architecture
Open Systems Architectures: from Monolithic Approaches to Service-Based Architectures	Vernadat (2003)	Distributed Architecture
Challenges when moving from Monolith to Microservice Architecture	Kalske et al. (2018)	Distributed Architecture
Domain-Driven Design of Big Data Systems based on a Reference Architecture	Avci Salma et al. (2017)	Domain-Driven Design
Cloud Native Architecture and Design: a Handbook for Modern Day Architecture and Design	Goniwada (2022)	Enterprise Architecture
The Relation between EA Effectiveness and Stakeholder Satisfaction	Van der Raadt et al. (2010)	Enterprise Architecture
Assessing your Microservice Migration	Henry \ Ridene (2020)	Microservices
Microservices Architecture enables DevOps: Migration to a Cloud-Native Architecture	Balalaie et al. (2016)	Microservices
The Data Revolution and Economic Analysis	Einav \ Levin (2014)	Microservices
Development and Validation of Enterprise Architecture Readiness Assessment Model	Hussein et al. (2020)	Readiness Assessment
ICT Readiness Assessment Model for Public and Private Organizations in Developing Countries	Chanyagorn (2011)	Readiness Assessment
The Use of Readiness Assessment Framework for Blockchain Adoption: A Healthcare Case Study	Barham \ Daim (2020)	Readiness Assessment
Soa Readiness Assessment, a New Method	Mirarab et al. (2014)	Readiness Assessment
Implementation of 5S Methodology in Public Libraries: Readiness Assessment	Bahadorpoor et al. (2018)	Readiness Assessment

C FACTOR IDENTIFICATION

This appendix gives an overview of the different studies that were analyzed in order to extract dimensions and factors from literature that are of influence on an organization's readiness towards adopting new technologies.

C.1 DIMENSIONS AND FACTOR EXPLORATION

C.1.1 Barham and Daim (2020)

- **Context**: Big Data Projects
- **Dimensions**: People, Technology, Legal, Organization
- **Factors**: Data scientists, Technological skills, Public acceptance, Analytical skills, Data integration, Data availability, Technology solutions, External sources of data, Data ownership, Data security, legislation's adaptability, Management support, Data governance, Clarity of objectives

C.1.2 Balasubramanian et al. (2021)

- **Context**: Blockchain adoption
- **Dimensions**: Stakeholders, Readiness
- **Factors**:Governments, Business entities, Blockchain providers, Customers, Motivational readiness, Engagement readiness, Technology readiness, Structural readiness

C.1.3 Chanyagorn and Kungwannarongkun (2011)

- **Context**: ICT Readiness
- **Dimensions**: Environment, People, Process, Technology
- **Factors**: Vision, Culture, Management, Resources, Governance, Stakeholder support, Motivation, Repository

C.1.4 Mirarab et al. (2014)

- **Context**: SOA Readiness
- **Dimensions**: Integration, Technology, Support, Governance, Security, Standards
- **Factors**: Application integration, Dynamic architecture, Layered architecture, Existing SOA Capabilities, Changing culture, Business view of IT, Perception of SOA, Self-organized teams, Automation, Policies for SOA, Governance policies, Enterprise security, Access security, Communications, Technology standards, SOA Governance model, Development technique

C.1.5 Al-Ammary and Saleh (2021)

- **Context**: Cloud Computing
- **Dimensions**: Technology, Organization, Environment

- **Factors**: Privacy, Security, Compatibility, Vendor lock-in, Management support, Organization size, Governance, Organization readiness, Competitive pressure

C.1.6 Bahadorpoor et al. (2018)

- **Context**: ...
- **Dimensions**: Organization, Management
- **Factors**: Financial power, Required facilities, Cultural flexibility, Changeability, Organizational flexibility, Staff training, Risk-taking spirit, Management support

C.1.7 Hamid and Mansor (2016)

- **Context**: Software projects
- **Dimensions**: Strategy, Culture, Process, Technology, Management, People
- **Factors**: Finance, Time, Resource, Approval, Leadership, Business and IT alignment, Commitment, Compliance, Requirements, Security, Support, Architecture, Risk, Knowledge, Skills, Experience, Roles and responsibilities

C.2 FACTOR IDENTIFICATION

Table C.1: Factor identification from literature

Category	Factor	Source
Culture competences	Desire to transition	S17, S18
	Learning culture	S7, S19, S21
	Top-management engagement	S17, S18, S19, S21
	Data literacy standards	S1, S4, S11, S12, S20
Responsibilities and ownership	Data ownership appointment	S3, S4, S7, S19
	Clearly defined roles and responsibilities	S3, S4, S6
Strategy	Clearly defined objectives	S17, S18, S19
	Bottleneck identification	S12
	Understanding of business domains	S3, S4, S5, S6, S9, S15
	Ability to decouple business applications	S3, S6, S12
Governance	Business IT alignment	S6, S7, S8, S16, S20
	Collaboration guidelines	S3, S21
	Data privacy policy	S17, S18, S19
	Interoperability policy	S1, S7, S8, S15, S16, S21
Enabling technologies	Accessible data products	S2, S5, S6, S14, S15, S19
	Use of intelligent tooling	S5, S10, S20
	Discoverable data products	S5, S6, S9, S10, S14
Central integration	Central data catalog	S2, S5, S14, S15
	Data quality management	S5, S10, S14, S20
	CI/CD standards	S1, S5, S6, S7, S8, S15, S16, S20, S21
	Automation standards	S3, S4, S5, S6, S9, S13, S15, S20

Table C.2: Sources numbering

Citation	Source number
Genovese (2021)	S1
Giebler (2021)	S2
Goniwada (2021)	S3
Hokkanen (2021)	S4
Joshi (2021)	S5
Loukiala (2021)	S6
Machado (2021)	S7
Orescanin Hlupic (2021)	S9
Priebe Markus (2021)	S10
Hyperright (2021)	S12
Kotochevikj (2021)	S13
Salma (2017)	S14
Henry Ridene (2020)	S15
Hussein (2020)	S17
Chanyagorn (2011)	S18
Barham Daim (2020)	S19
Mirarab et al. (2014)	S20
Bahadorpoor et al. (2018)	S21

D EXPLORATORY INTERVIEWS

D.1 INTERVIEW SELECTION

Due to a cooperation with Deloitte Consulting and its role in evaluating the artifact, interviewees were sought primarily outside of Deloitte. In order to find experts on such a new topic, Tech Fora and Data Mesh LinkedIn communities served as the prior base of finding experts. Due to the snowballing effect, a Slack Channel 'Data Mesh Learning Community' with over 5,000 members was incorporated to find experts to interview. Moreover, several Data Mesh initiatives such as the 'Data Mesh Podcast' and 'Data Mesh Knowledge Exchange Platform' were used to get in contact with experts from the field. The selection criteria for incorporating interviewees within this research was based on providing IT/data services, current way of working, Data Mesh implementation intention, and availability. Eventually, fourteen interviewees have been involved in the research. An overview of the interviewees and their specific roles can be found in table D.1.

Table D.1: Interviewees for refining theoretical model

Interviewee	Role	Operating Group	Years of Experience
A	Data Mesh Researcher	Engineering Research	5 years
B	Data Engineering Manager	Financial Services	6 years
C	Platform Product Manager	Public Services	10 years
D	Principal Scientist	Information Technology	20 years
E	Senior Manager	Financial Services	15 years
F	Chief Technology Officer	Information Technology	10 years
G	Senior Data Consultant	Information Technology	10 years
H	Managing Partner	Information Technology	7 years
I	Data Specialist	Public Services	8 years
J	Global Senior Director	Consumer Goods	5 years
K	Senior Customer Engineer	Information Technology	15 years
L	Principal Engineer	Consumer Goods	6 years
M	Principal Data Consultant	Information Technology	7 years

D.2 INTERVIEW PROTOCOL

This appendix describes the process of conducting the model refinement interviews. It describes the interview questions in detail, it provides the used interview protocol which is used during each interview and it provides transcriptions of relevant parts of the interview.

D.3 INTERVIEW QUESTIONS

1. Demographic information of the participant

(a) Can you tell me about your work: what is your role in the organization?
(b) How does your role relate to Data Mesh architectures?

2. Discussing the experience cases of the expert

(a) Have you ever (partly) implemented a Data Mesh architecture within an organization?
(b) Why did you choose a Data Mesh architecture as a solution? What was the current situation like?
(c) What do you think are requirements of an organization for migrating to a Data Mesh architecture?

3. Discussing the identified capabilities from the theoretical model

(a) The found readiness dimensions are: Culture & Competences, Roles & Responsibilities, Strategy, Governance, Enabling Technologies and Central Integration. Do you observe these dimensions in your point of view of Data Mesh architectures? Can you relate to any of these dimensions?
(b) Are these dimensions complete? Would you think some factors need to be added?
(c) In your opinion, which dimensions or factors are most important?

4. Assessing readiness of organizations for migrating to a Data Mesh architecture

(a) In your opinion, what are the biggest hurdles or limitations for organizations wishing to migrate to a Data Mesh architecture?
(b) In your opinion, what kind of organizations are suited for a Data Mesh architecture?
(c) Why and to what extend could a Data Mesh architecture be beneficial for an organization?

D.4 PRIMARY DATA

The research data is retrieved by conducting interviews with experts from different operating groups, in order to get a broader view on the topic within various sectors. Empirical data is retrieved with using qualitative research, and the theoretical model of this research is through these interviews further developed towards a theoretically and empirically-based Data Mesh readiness assessment model. The theoretical model gave a clear overview of the generic readiness assessment model, and the practical data enriches this model with a larger focus on the best practices from the practical world.
Each interview took approximately an hour and was done via online meetings. All interviews were assessed in English since this made the transcription process of the interviews easier. This transcription document was used for qualitative analysis of the data. Summarized conclusions and quotes of the interviewees are used in the results section of this research.

D.5 SECONDARY DATA

The conducted interviews are seen as the primary data for this research. Besides this data, there is also secondary data retrieved. Most of the times, it was recommended by the interviewees to look into some specific external data. This external data is seen as secondary data in this research. The secondary data is used to complement the cases with extra insights from earlier researches, projects, papers or books.

D.6 DRAWBACKS

The expert interviews have other drawbacks that are necessary to mention. First, due to geographical and timely restrictions, all interviews were conducted online. This can be seen as a drawback, since it creates a distance between the researcher and the interviewee which might result in some bias towards the topic. Secondly, due to time limitations and the availability of people, it was not always possible to ask all questions of the interview protocol to the interviewee. Moreover, some opinions were interesting enough to dive deeper into, however, this was not possible due to these time restrictions. Next to that, it seemed quite hard for the interviewees to rate the factors as included in the theoretical model by means of importance, due to the novelty of the topic. Most interviewees only have suggestive or minor experience with implementing a data mesh architecture, and found it therefor hard to make bold statements on the factors within the theoretical model. Although an attempt was made within this research to find interviewees that have already implemented a complete data mesh architecture in practice, this was not possible since there weren't any. This drawback was assessed by stating that the interviewees should indicate whether they think it would be important for them in their organization, or whether they would focus on the aspect. Most of the interviewees doubted on their answers but could make accurate estimates on whether the factor could be important when potentially implementing a data mesh architecture.

E CASE STUDY PROTOCOL

The case study protocol aims to keep the researcher targeted on the topic of the case study. In addition, the protocol increases the reliability of the case study and is intended to guide the researcher in consistently carrying out the data collection. The protocol consists of the following sections:
1. Overview of the case study
2. Data collection procedures
3. Protocol questions
4. Tentative outline for the case study report

The mission of the case study is to apply and evaluate the created artifact in a real-life business environment.

E.1 OVERVIEW OF THE CASE STUDY

The participants of the case studies are all organizations for which a data mesh architecture could be beneficial and which have an intention for migrating to a distributed architecture. The interviews will be done following a semi-structured interview format based on the designed model. This interview format can be send to the interviewee if he/she prefers that. The idea of a semi-structured interview is that there are questions for the main line of the interview, but depending on the answers the interviewee gives, other questions can also be asked. Also, the interview format can be updated based on previous interviews, for example if a question or the structure should be formulated differently.

E.2 DATA COLLECTION PROCEDURES

Data will be collected in four steps. First, three introductory questions will be asked to determine the organization's perspective towards Data Mesh migration. Moreover, the participant is asked to assess its own readiness for change, without the use of the Data Mesh Readiness Model. Second, the Data Mesh Readiness Model is applied in practice and the assessment is done together with the participant and the researcher. Third, the Data Mesh Readiness Model is evaluated with the participant with respect of usability, validity and comprehensibility. Lastly, the participant is given the opportunity to state additional remarks.

E.3 PROTOCOL QUESTIONS

Preparation

(a) Do you mind if I record this interview?
(b) Introducing myself: Who am I? Who is Deloitte and how are they involved? What kind of information am I looking for? Why am I speaking to the participant specifically?
(c) Ask participant to introduce himself/herself. What is your position? How are you involved with Data

Mesh architectures?
(d) Introduction of research topic and presentation of Data Mesh Readiness Model.

1. Open questions

(a) Would it be desirable for your organization to migrate to a Data Mesh architecture?
(b) What would be the main barriers to migrate to a Data Mesh architecture?
(c) How would you assess your own organization's ability to change to a Data Mesh architecture?

2. Assess the organization's Data Mesh readiness, by:

(a) Indicating to what extent the factors within the model are already in place within the organization. For questioning the factors, the factor descriptions from Appendix **??** are used;
(b) Indicating the degree of importance of these aspects.

3. Evaluate the model, by:

(a) Asking whether the structure of the model is logical and comprehensive;
(b) Stating whether the readiness factors as provided by the model are a complete spectrum of Data Mesh readiness, or whether there are any factors missing.

4. Enabling the participant to state any other comments and/or recommendations.

5. Follow-up

(a) Thank you
(b) Confirming e-mail with transcript of interview
(c) Follow-up email with assessment results

E.4 TENTATIVE OUTLINE FOR THE CASE STUDY REPORT

After the assessment session, a small report will be generated by the researcher and shared with the case study participants by mail. This report will contain:
(a) Description of the Data Mesh Readiness Model used in the case study;
(b) Summary of the results/feedback/comments as observed by the researcher during the case study.

F DATA MESH READINESS ASSESSMENT INSTRUMENT

In this appendix, the assessment instrument and the example results report are given.

F.1 DATA MESH READINESS ASSESSMENT INSTRUMENT

On the following pages, the instrument is shown with which an organization's readiness towards a Data Mesh architecture can be assessed.

DATA MESH READINESS ASSESSMENT

This readiness assessment guides organization through the several stages determining the readiness for a Data Mesh implementation in a stepwise manner.

Assess the reason to change

Problem identification

Need
- Need for autonomous domain teams
- Need for better data responsibility and ownership
- Need for shorter lead times from operational teams to data analytics teams
- Need for greater scalability of the current data platform
- Need for greater flexibility of the current data platform
- Need for a coherent data strategy
- Need for reduced workload of data engineering teams

Assess the ability to change

High-level requirements

Willingness to Change
- Desire to transition
- Learning culture
- Innovation culture

Vision
- Clearly defined and aligned objectives
- Clearly defined and aligned value statement

Business Case
- Business & IT alignment

Leadership
- Top-management engagement

Enterprise Ability
- Ability to decentralize the business into domains
- Ability to decouple business applications
- Ability to work in domain teams

Assess the preparedness for change

Formal structures & guidelines

Accountability
- Clearly defined roles and responsibilities

Governance
- Data privacy policy
- Data security policy
- Data documentation policy
- Interoperability policy

Workable Approach
- Data literacy learning material
- Communication and collaboration guidelines
- Knowledge sharing platform

IT Capacity
- Data product publication standards
- CI/CD standards
- Automation standards

Enterprise Capacity
- Team structures aligned with business domains
- Agile way of working

Determine Data Mesh readiness

Complete set of forces

Data Mesh building blocks
- Self-organizing and autonomous domain teams enabled by self-serve platform
- Central platform IT team
- Domain representatives in governance team
- Self-serve platform
- Central data catalog with discoverable, accessible, trustworthy, secure, interoperable and self-describing data products

1: ASSESS THE REASON TO CHANGE

symptoms on an inefficient socio-technical organization around data

NEED

Factor	Description
Need for autonomous domain teams	Indication whether there is a need for domain teams to be able to do their own data analysis. • Is there a centralization of data, capabilities or resources, that inhibit getting the full potential out of the organization's data? • Can they be independent in how they address their own domain analytics? • Are the teams experiencing organizational and technical dependencies? • Is the accountability for the data close to where the data is produced?
Need for better data responsibility and ownership appointment	Indicate whether there is a need for greater knowledge over the data that is worked with, in terms of the fact that the people that work with the data are the people that know the data. • Do the data engineers know the value and quality of their own data sets? • Are the domain experts involved in delivering their data to the data analysts? • Do the data teams understand what they contribute to in a business context?
Need for shorter lead times	Indication whether there is a need for shorter waiting times for the analytics team to have their requests to the engineering team processed.
Need for greater scalability of the current data platform	Indication whether there is a need for the platform to process bigger amounts of data through the data platform in the near future. • What is the possibility and ease of adding more resources to the current data platform? • What is the possibility and ease of growing the resource utilization of the current data platform?
Need for greater flexibility of data platform	Indication whether there is a need to ease making changes on the existing IT system/data platform. • Is the current data platform build on premise or is it easily accessible by others? • Is the current data platform changeable, or is it too complex to be changeable? • Is there siloed information or siloed data in this platform?
Need for a coherent data strategy	Indication of the existence of a coherent data strategy. • Is there a coordinated, common data strategy? • Are the data engineers and data analysts aware of each other's needs and purposes? • Is data an important asset of the organization? • Is decision-making informed on data?
Need for understanding of the domain complexity	Indication whether there is a need for the lessening of the current domain complexity, in terms of the existence of comprehensibility of inner domain processes and understandability of the end to end processes. • Do the business domains understand what they contribute to? • How many data sets exist in the domains, do they understand this data? • Do the domains implement their processes independently and in a siloed manner?
Need for reduced workload of data engineering teams	Indicate whether there is a need for healthier working conditions for the data engineering teams. • Is the data engineering team causing a bottleneck in the data processes? • Are the data teams satisfied with their workload (i.t.o stress levels)? • Are the data engineering teams making their deadlines (i.t.o. quality of work)?

2: ASSESS THE ABILITY TO CHANGE

the required organizational factors that need to be in place before moving onto the next stage

WILLINGNESS TO CHANGE

Desire to transition	Indicate whether there exists a commonly shared desire to transition to a Data Mesh within the organization as a whole and among the individuals within the organization. • Are they willing to take the efforts needed to establish this transition? • Are the existing teams eager to be autonomous? • Is there a willingness to participate in the Data Mesh migration?
Learning culture	Indicate whether there exists a culture that supports learning. • Are the people motivated to learn new procedures, such as working Agile? • Is there management involvement in establishing this culture? • Are there resources available to support this culture? • Do the organization's mission and vision statements support this culture?
Innovation culture	Indicate whether there exists a culture that supports innovation. • Do the people have interest in working with data more efficiently? • Are the people open for adapting to new technologies such as DevOps and DataOps? • Is the organization as a whole data-driven? • Do the organization mission and value statements drive this culture? • Are there resources available to support innovations? • Are there organizational processes or facilities implemented that facilitate this culture? • Are people experiencing long decision paths when implementing innovative projects?

VISION

Clearly defined objectives	Indicate whether the objectives for migration to Data Mesh are clearly defined and aligned with the people in the organization. • Why does the organization wants to migrate to a Data Mesh? • What are the needs that are addressed by the migration to a Data Mesh? • Does there exist common awareness about these objectives? • Are the people within the organization engaged with these objectives? • Are there performance measures in place for measuring the progressions?
Clearly defined value statement	Indicate whether the value of a Data Mesh to the organization is clearly stated and aligned with the people in the organization. • What is the value that it will bring to the organization? • Are the people within the organization aware of the value that a Data Mesh will bring to the organization? • Is there an idea on the returns on investment? • Do there exist reflections on past projects that brought value?

BUSINESS CASE

Business & IT alignment	Indicate whether the data operations and processes are aligned with the business needs. • What are the most important business requirements, and are they in line with the data strategy? • Is the data team aware of their contributions to the business performance? • Do the business users wish to be involved in the curation and processing of data from source applications to the reports that they use? • Has the organization benchmarked itself with respect to similar organizations? • Is there an understanding about the impact that IT has on business processes? • Is there an understanding about the amount/existence of organization strategic goals and requirements supported by IT strategic goals? • Are Business and IT aligned with prognosed future growth? • Is there an understanding of the digital capabilities needed to support the organization's business strategy? • Will the Data Mesh architecture improve the organization's current Business & IT alignment?

LEADERSHIP

Top-management engagement	Indicate to what extend the top-management of the organization is engaged with the migration to a Data Mesh. • Does the migration align with their interests? • Do the business leaders understand the Data Mesh? • Do they understand the need for a Data Mesh? • Have they allocated time and resources to the development of the Data Mesh migration?

ENTERPRISE ABILITY

Ability to decentralize the business into domains	Indicate to what extend the organization is able to decentralize its current central architecture (in terms of processes and organizational design) into different domains. • Does the organization understand the domain driven design principles? • Can the organization map its business into different domains? • Is there an idea of the domain boundaries? • Are the boundaries distinct and explicit? • Is there a reference design on the perceived decentralization of the organization? • Is there an idea on the preferred degree of decentralization? • Is the organization able to shift (a part of) the centralized decision-making into decentralized decision-making (e.g. across multiple teams)? • Is the organization able to shift (a part of) the centralized roles, functions and tasks into decentralized roles and responsibilities (e.g. across multiple teams)?
Ability to decouple business applications	Indicate to what extend the existing business applications (in terms of data processing) can be decoupled into the domain-oriented teams. • Can the organization appoint the existing tables and pipelines to domain teams? • Can the existing monolith be split up into smaller systems, following a domain-driven approach? • Is the organization very much application focused, so that they don't break well into domains? • Are the business processes very much centralized in ERPs? • Is the organization able to establish a decoupled application architecture that allows each component of the application to perform its tasks independently in the domain teams? • Can the components of the business application remain autonomous of each other? • For organizations using pre-built applications: can these applications independently run across the Data Mesh (e.g. across multiple autonomous teams)?
Ability to work in domain teams	Indicate to what extend the domain teams understand the new domain-oriented way of working. • Do the data analysts, that were primarily focused on getting insights from the data, understand how to maintain data pipelines? • Do the domain teams understand their extra responsibilities? • Do they understand the concept of a Data Mesh? • Do they understand the data-as-a-product thinking? • Do they understand data products? • Do they understand how to work together?

3: ASSESS THE PREPAREDNESS FOR CHANGE

the structures that need to be in place before implementing a Data Mesh architecture - the priorities are indicated in the factor boxes

ACCOUNTABILITY

Clearly defined roles and responsibilities **[priority]**	Indicate to what extend the roles and responsibilities are correctly described. • Will all the domains have a data product manager? • Will the domains have data engineers, and will they be given space to spend more time understanding the data? • Do there exist data ownership guidelines? • Does there exist no claimed ownership over the same applications? • Are the roles and responsibilities of the team members in line with their personal skillsets?

GOVERNANCE

Data privacy policy **[priority]**	Indicate whether there exists a data privacy policy that ensures the processed data to be GDPR compliant.
Interoperability policy **[priority]**	Indicate whether there exists an interoperability policy, that states how to enable interoperability between domains and their domain related data-products. • Does this policy allow for the fast consumption of data products?
Data security policy **[priority]**	Indicate whether there exists a data security policy.
Data documentation policy **[nice-to-have]**	Indicate whether there exists a data documentation policy, that states how to document what a data product means and how to define a domain.

WORKABLE APPROACH

Data literacy learning material **[priority]**	Indicate whether there exists learning material that enhances the data literacy among the people in the organization, in order for them to work autonomously on data products without the need for data engineers. • Does it state what tools to use for their data products and how do these tools work? • Does it state how to understand the data? • Does it state how to publish the data products? • Does it state how to make data-based decisions? • Does it support reading, understanding, analyzing, managing and acting on data products? • Does it create overall data awareness among the organization?
Communication and collaboration guidelines **[nice-to-have]**	Indicate whether there exist guidelines that enable and ease communication and collaboration between the various domain teams. • Does it state how to enable cross-domain collaboration through data exchange? • Does it state how to collaborate on a data product?

WORKABLE APPROACH

Knowledge sharing platform **[nice-to-have]**	Indicate whether there will be knowledge-sharing facilities in place that enable the education, upskilling and sharing of knowledge within the organization. • Is the platform accessible and interactive? • Does the platform support Data Mesh engagement? • Does it include information, guidelines, policies, and instructions on Data Mesh architectures?

IT CAPACITY

Data product publication standards **[priority]**	Indicate whether there exist standards that state how and when to publish data products. • What is the standard publication format? • What needs to be in a data product? • Are there technology standards with which people should build a data product?
CI/CD standards **[nice-to-have]**	Indicate whether there exist CI/CD standards that enable continuous integration and continuous delivery. • Do these standards support the efficiency of the performance of the Data Mesh? • Are these standards frequently measured in terms of cycle time, change failure rate and deployment frequency?
Automation standards **[nice-to-have]**	Indicate whether there exist automation standards for standard software procedures like data quality checking, pipeline creation, etc. • Is the organization able to automate software and data processes? • Do these automations improve the efficiency of the performance of the Data Mesh?

ENTERPRISE CAPACITY

Team structures aligned with business domains **[priority]**	Indicate whether the domain teams will be business aligned, in terms of that they have responsibility for a certain part of the business. • Will the domain teams not be too far away from the business cases? • Does the organizational structure support a Data Mesh way-of-working?
Agile way of working **[nice-to-have]**	Indicate whether the organization supports and encourages an Agile way of working throughout the organization.

4: DETERMINE DATA MESH READINESS

the executable building blocks of an entire Data Mesh that can be implemented after accomplishing Data Mesh readiness in the former stages

DATA MESH BUILDING BLOCKS

Self-organizing domain teams enabled by self-serve platform	Indicate whether the organization has established self-organizing domain teams, that are able to cross-collaborate with other domain teams and that are enabled to autonomously work on their own data products via the self-serve platform.
Central platform IT team	Indicate whether the organization has established a central platform IT team that bears responsibility over performing and maintaining the self-serve platform.
Domain representatives in governance team	Indicate whether the organization has established a central governance team consisting of domain representatives, that administrates the policies and standards required for the performance of a Data Mesh.
Self-serve platform	Indicate whether the organization has developed a self-serve platform, maintained by the central platform IT team, that enables the domain teams to work autonomously on their data products and collaborate with other domain teams.
Central data catalog	Indicate whether the organization has developed a central data catalog which enables the sharing, discovering and accessing of trustworthy, secure and self-describing data products throughout the organization.

G TESTCASES

G.1 TESTCASE 1

Table G.1: Full description of testcase 1

Factor	Interpretive
Need for greater autonomy of domain teams	Mostly, teams have a want and need to be more autonomous, but need to follow corporate guardrails; with compliance usually lower than expected.
Need for more upstream data responsibility and ownership	Too often, data engineering teams are experts at tooling and technology, rather than domain experts, leading to slow iterations and lack of domain understanding.
Need for shorter lead times	I would rather point towards the need for joint exploration and iteration, to accelerate learning.
Need for greater scalability of the data platform	Platforms become huge and unwieldly in data lake concepts, where all data is there, but governance and life cycle management lacking.
Need for greater flexibility of the data platform	The change cycle is getting faster overall, but often this is more due to implicit relaxed expectations and low LCM with immature agile approaches, than discipline of execution. Only the very best and mature teams manage to invest in continuous improvement to avoid technical debt, such teams maintain good interfaces, proper canonical/master/reference data and malleability.
Need for a data strategy	Some domains have patterns established and shared between them, many are building similar platforms.
Need for understanding of the domain complexity	Domain teams have better understanding on what they are accomplishing, compared to centralized teams. They are often also more aware of, and have better understanding of how to protect their data.
Need for reduced cognitive load of data engineering teams	Depending on the maturity of the teams, the teams and team members get managed more or less well.
Desire to transition	There would be a curiosity about the shift, but not yet managed to put digital capabilities as part of the business itself, it demands a huge mind-shift in traditional enterprises.
Learning culture	Agile adoption and the idea of becoming a learning enterprise is common, but essentially mostly in theory.
Innovation culture	There is a common interest in working more with data but the agile approach is in general fairly immature with the business often requesting "data projects", with a significant big-design-up-front approach, rather than taking a stepwise value driven approach, where platforms and capabilities are built gradually from a learning and hypothesis confirmation perspective.

Clearly defined objectives	The data driven approaches are still very exploratory, to some extent driven more to evaluate Data Mesh as a solution to current problem, than a strategy for making it easier to address opportunities faster, and with more ease and agility. I my opinion, iterating and validating/refuting hypotheses with shorter turnaround thanks to an agile approach and a data as a product mindset.
Clearly defined value statement	See above.
Business and IT alignment	Organisations having adopted data stewards as a concept have had better foundations to build upon to get data quality up and a more data driven approach. Depending on the data literacy, the involvement is different. Finance and procurement departments are driving good data and master data approaches where business users are closely involved in both curation, processing data and driving RPA approaches for some of the data handling. They still have a lot to do to become self-sufficient, but with a lot of involvement and driving the needs for a local data analytics team and a domain-orientated tooling and implementation.
Top-management engagement	Some do, but mostly rather see the needs for driving digital efforts as part of the business activities with a shift from traditional siloing of it/data skills to belong to IT departments or similar. I think a full Data Mesh approach remains with business digital champions and drivers, rather than as part of top management as imperatives for digital acceleration.
Ability to decentralize the business into domains	Efforts to decentralize are there, but conceptualizing integration needs, ensuring interoperability and providing clear data interfaces, format governance and other efforts to provide "just enough architecture" remains hard. In general, I believe few organisations invest enough in and ensure to provide the necessary guidance to achieve domains a way to express their interfaces securely and well, without compromising autonomy and agily. Data/integration contracts that are not centrally driven needs more support to be built well. Often there is a conceptual idea of the domains, but without enough specificity as to what specifically belongs where. Capability/component business modelling has seen success in some mature enterprises to govern what is in what domains.

Ability to decouple business applications	The enterprise has significant investments in ERPs which become a very centralized approach to keeping tabs of large parts of them, with vendor lock-in and limited data interoperability as a result. What often happens is that all data is pipelined out of the ERP system into various other locations, that lack necessary oversight and life cycle management. The overall reality is that the enterprise is mostly application focused, and lack tools to work directly with data driven approaches in general, where multiple steps of an overall process is handled in various application suites (and integrated very well in such suites). Businesses in general don't operation from an API exposure point of view with clear contracts that can be maintained over time. Often, it's the specific application API that is exposed, which is hard to evolve, but maintaining such canonical APIs are also quite a lot of effort as the domain expertise is most often of emergent nature.
Ability to work in domain teams	Below has limitations of not yet having seen an overall and broadly well-executed data mesh strategy, it is mostly by domain adopting what they can, and not the full enterprise adoption (yet). Mature teams, with enough size and business impact, and that have disciplined and skilled leaders in the digital realm manage to make the transition locally to use the concepts of data mash and have readiness to adapt and adopt. Mature teams face immature, inadequate or non-existing governance structures where data mesh beyond the individual team has significant barriers; where corporate architects need to build trust, step back and support the teams getting better and increase autonomy and not be a barrier. Immature and teams with significant technical debt from previous paradigms have a hard time making the shift, as they are seldom coupled well as a natural part of the business receivers of the data analysis they perform. Here it takes business leadership to drive the changes needed and ensure to appoint both vanguards and enough resource to enable the shift.
Clearly defined roles and responsibilities	Mature teams do, and have clear ownership and roles, whilst still struggling to fully operationalize and market as actual data product, due to enterprise challenges of scaling the approach (having business leaders behind that understand the paradigm shift). Others, not so much data product thinking.
Data privacy policy	Well-developed data privacy policies and guidelines.
Interoperability policy	No, or very limited in general. Previously had adopted OAGIS as canonical model for data interoperability, but found the challenge/burden of complying and maintain clarity as to how to use it very challenging (it was a custom-driven integration/SOA effort)
Data security policy	Yes. But most often focused on confidentiality, not integrity, availability, non-repudiation and only to some extent on data privacy (but if so, separately). Getting data security policies right remains a finicky endeavor; especially in making it useful for fluid decision making on how to proceed.
Data documentation policy	Have not seen any.
Data literacy guidelines	No, but I really like the approach! This would be very helpful!

Communication and collaboration guidelines	No, but equally a great approach to recommend developing!
Data product publication standards	No. Have not seen any. Pieces of it, such as data formats exists more readily, but structuring the data product itself would be great to see more of.
CI/CD standards	Standardization is mostly tooling based and usually in the software domains, dedicated platform operations teams deal with aligning and creating the pipelines. I have not seen standards in the space, but would welcome more standardized approaches as it would help interoperability and reduce the barriers to entry for new teams.
Automation standards	Unsure. I believe mostly vendor driven.
Agile way of working	Yes, everybody wants it, but few are ready to really shift out from and remove the traditional project-orientated way of working. I believe projects and output orientation are the most important barriers to adopt agile. It's much easier for manager to escape responsibilities with projects, compared to leading agile where progress and leadership has to be part (the old managers have little room, and need to start producing results).
Team structures aligned with business domains	Some are, but often service a large business, and not fine-grained enough (as they have to do much of the platform and governance work themselves)
Self-organizing domain teams	No. The organization has not reach this maturity. They are making the first steps in each area.
Central platform IT team	Platform operations is being explored, but not yet put in place at scale.
Domain representatives in governance team	No, but I fully agree on the need.
Self-serve platform	No, not at scale. But some scale up existing successes of specific domains, where autonomous teams have had success building up the necessary structures of a self-serve platform consumed by subteams.
Central data catalog	API catalogues exist where API Management is in place, but taking the leap to data catalogues would be a (in my mind) natural next step, and be done in conjunction probably! What is important is to define the domains and let them define interfaces and products they take responsibility for, and the life cycles of them (and that takes an enterprise catalogue to publish in).

G.2 TESTCASE 2

Table G.2: Full description of testcase 2

Factor	**Interpretive**
Need for greater autonomy of domain teams	Yes, domain teams have expressed the desire to be more autonomous in their data and insight activities. They can be independent, but they lack resources, tools and training.
Need for more upstream data responsibility and ownership	People know the data up to an extent, but we've found that when you deep dive, some concepts are harder to understand for Data Engineers that are seated in a central team.

Need for shorter lead times	Yes, definitely. Analysts would want to be autonomous also in answering questions.
Need for greater scalability of the data platform	Data platform infrastructure is able to scale quite well.
Need for greater flexibility of the data platform	Well documented, easily changeable.
Need for a data strategy	There is a recently developed strategy but it takes time for it to settle.
Need for understanding of the domain complexity	Domains are getting more complex, so less people can cognitively understand them, therefore you need to be focused or specialized in one of the domains.
Need for reduced cognitive load of data engineering teams	Yes, definitely a need.
Desire to transition	Yes
Learning culture	Yes
Innovation culture	Yes
Clearly defined objectives	Yes, more autonomy in domain teams and distributed ownership of datasets.
Clearly defined value statement	Have not done a ROI analysis, because it is very hard to measure the value we will get. But top of mind there is value perceived and no other way of acting around the challenges the organization is facing.
Business and IT alignment	Yes, although this has taken months.
Top-management engagement	Yes, they support it although some times it is difficult for them to have concrete actions to advance on this topic.
Ability to decentralize the business into domains	Yes, it is possible. Domain concept has already been used in the past.
Ability to decouple business applications	They already are for the operational side, only need to do that on the analytics side because until now it was centralized.
Ability to work in domain teams	Domain teams need better tooling to be autonomous and do their job well. They understand the concepts but are worried on the increase of responsibilities they will get without the increase in capacity in the teams.
Clearly defined roles and responsibilities	The role does not exist per se, but the senior analyst is doing this type of job. Not all the domains have data engineers. Data ownership guidelines exist.
Data privacy policy	Yes
Interoperability policy	Yes. The data as a product principle was implemented even when the team was a central one.
Data security policy	Yes.
Data documentation policy	Yes.
Data literacy guidelines	It is not very structured at the moment.
Communication and collaboration guidelines	Yes, they exist.

Data product publication standards	Yes.
CI/CD standards	Yes.
Automation standards	Yes, for pipeline creation.
Agile way of working	Yes.
Team structures aligned with business domains	Yes.
Self-organizing domain teams	Yes, we are on this journey.
Central platform IT team	Yes.
Domain representatives in governance team	Yes.
Self-serve platform	Working on it.
Central data catalog	Working on it.

G.3 TESTCASE 3

Table G.3: Full description of testcase 3

Factor	Interpretive
Need for greater autonomy of domain teams	Many different tools are used and local solutions implemented throughout the organization. Analysts use their own resources, not being sure of the quality of the data they are using.
Need for more upstream data responsibility and ownership	There is little sense of responsibility over the data, at least not in terms of data lineage. If I change anything here, it might change for a more downstream domain too.
Need for shorter lead times	We are not yet processing so much data that there are actually bottlenecks in the data platform. This would therefore not be a need with which we can identify very much.
Need for greater scalability of the data platform	In terms of scalability, we can still easily scale technologically, but not necessarily in terms of resources. The central solution can still handle the amount of data well, now and in the future.
Need for greater flexibility of the data platform	In terms of flexibility of the data platform, we are dealing with various data engineers who all build their own solutions. Nobody dares to come up with those personal solutions, because we are afraid that we can no longer count on the results.
Need for a data strategy	Some domains have patterns established and shared between them, many are building similar platforms.
Need for understanding of the domain complexity	The domains are not yet so complex that there is a lack of understanding.
Need for reduced cognitive load of data engineering teams	See above.
Desire to transition	There is mainly a desire to be more data driven and to create more data awareness. This awareness is often still lacking about the organization as a whole.
Learning culture	We are rolling out a program in which we try to improve this data awareness. This is partly supported by the company culture to innovate.

Innovation culture	See above.
Clearly defined objectives	There are clear objectives, but not necessarily towards a Data Mesh. It is more of a higher level, namely objectives to better handle the data we have and to process more data in the future.
Clearly defined value statement	It is clear within the organization what the value of data would bring to the organization.
Business and IT alignment	The business and IT are not always well aligned. Especially in the field of data awareness, much alignment is still missing. The individual pillars in the organization all use their own solutions, and we want to use at least the same solutions in terms of technology. In addition, there is a lack of efficiency and effectiveness that is necessary to be able to accelerate.
Top-management engagement	The top executives are pushing the change because they understand its value to the business.
Ability to decentralize the business into domains	This doesn't apply to us as we don't have the need for a decentralized setup yet.
Ability to decouple business applications	This doesn't apply to us as we don't have the need for a decentralized setup yet.
Ability to work in domain teams	This doesn't apply to us as we don't have the need for a decentralized setup yet.
Clearly defined roles and responsibilities	This could also be improved apart from a Data Mesh.
Data privacy policy	Yes.
Interoperability policy	No, but would be beneficial for the organization.
Data security policy	Yes.
Data documentation policy	No, but it needs to be worked on. Better documentation could help create an overview of the data operations we are currently making.
Data literacy guidelines	No, but would be very helpful.
Communication and collaboration guidelines	No.
Data product publication standards	No.
CI/CD standards	No, but it is being worked on. There is still a lot to be done in the field of automation and continuity..
Automation standards	Vendor driven.
Agile way of working	Yes.
Team structures aligned with business domains	Does not apply.
Self-organizing domain teams	Does not apply.
Central platform IT team	Does not apply.
Domain representatives in governance team	Does not apply.
Self-serve platform	Does not apply.
Central data catalog	Does not apply.

H | SCIENTIFIC ARTICLE

The Assessment of Organization Readiness for guiding Data Mesh Migrations

W.M. de Boer
Faculty of Technology, Policy and Management
Delft University of Technology
w.m.deboer@student.tudelft.nl

Abstract – **Part of the preparatory phase of organizations for an implementation of a Data Mesh architecture is assessing the readiness for accepting, using and operationalizing this new architecture design. The method as proposed in this paper assesses parts of the organizations, subdivided into dimensions and organizational steps towards readiness, for their degree of readiness in preparation for a Data Mesh migration trajectory. Because the concept of Data Mesh architectures is still very new and little scientifically based, this research offers a structure and overview to organizations that are considering migrating to a Data Mesh architecture. In addition, a Data Mesh migration covers the entire organizational scope, making the implementation of such an architecture design a large-scale and long-term project that must be approached accurately and systematically. A readiness assessment as presented in this study can provide guidance during this process. In this paper, a theoretical overview is first provided on Data Mesh architectures and their characteristics. Then, the design and development of the readiness model is explained, and presented as a method to facilitate Data Mesh implementation and preparation by assessing the readiness for Data Mesh implementation of an organization on various organizational factors in a step-wise manner. The main feature of the model in this study is its novel two-dimensional structure, which allows assessing Data Mesh readiness on both organizational as well as technological dimensions.**

Keywords – **Data Mesh, distributed architecture, decentralized architecture, domain-driven design**

I. INTRODUCTION

Currently, our society is in the middle of a digital revolution. This results in a bigger amount of data to be processed by organizations: research of IBM has shown that of all the world data, more than 90% has been collected in the last five years (Marr, 2018). Along with these ever-increasing amounts of data collected an processed, the concept of “Big Data” has emerged: a term that refers to an availability of data on an enormous scale, in real-time, for many different applications (Einav & Levin, 2014). Moreover, 95% of businesses cite the need to manage the large and unstructured amount of data as a problem for their business (Kulkarni, 2019). As a way to solve this problem and to address the value it can have for their performance, organizations have spent a lot of time and money in recent years developing various data and intelligence platforms that enable them to store, manage, and process this data efficiently (Dehghani, 2019). These platforms have been developed to accommodate an evolution in volume, variety, velocity, and veracity of big data, and have over the recent years fuelled an increased interest in adequate enterprise data architectures and technologies to deal with this massive increase in data volumes and types (Saddad, El-Bastawissy, M., & Hazman, 2020). For a long time, these enterprise data architectures were built as a single pipeline consisting of a single data processing unit, with the data generators at the beginning of the pipeline and the data operators at the end of it (Selmadji et al., 2020). This way of designing a data architecture is referred to as a monolithic data architecture design. However, due to the increasing amount of data to be processed, the ever-growing amount of data analysis techniques, and the amount of functionalities it has to address, these monolithic architectures have shown their limitations in different areas such as maintenance and debugging, scalability, maintaining quality data, and ownership and responsibility issues (Khazaei, Barna, Beigi-Mohammadi, & Litoiu, 2016; Khazaei et al., 2016; Lawal Moshood, Ileladewa Adeoye, & Lawal Habibu, 2020; Saransig & Tapia Leon, 2019; Selmadji et al., 2020). Consequently, these limitations result in an overload in the data teams in response to the growing needs of the organization (Dehghani, 2022).

A novel approach for addressing this problem has emerged as the Data Mesh architecture: a distributed domain-driven architecture design that splits the organization and associated data platforms into domains, each responsible for their own data processing. This new data architecture design promises to be more scalable, as adding new data-processing business units can be done by adding a domain to the existing entity. However, it is still unclear to many organizations whether this proposed solution can also offer a solution for them. And if the organizations have determined that the migration to a Data Mesh architecture can be beneficial for their performance, it is unclear how they should approach this migration. Implementing a Data Mesh architecture promises to deliver significant benefits over other approaches, but the process itself is not as simple as implementing a new tool or feature. Moreover, in order to actually deliver these significant benefits, it is important to include all relevant processes, structures, and regulations in the migration to a Data Mesh architecture. This is a very drastic and complex process for organizations that needs to be thought through very carefully before it can be carried out (Furia, 2021; Schultze, 2020). The process is in the first place hampered by the lack of a clear understanding of the specific steps that organizations need to take in order to migrate to a Data Mesh architecture. In addition, it is still unclear what the

This footnote will be used only by the Editor and Associate Editors. The edition in this area is not permitted to the authors. This footnote must not be removed while editing the manuscript.

organizational preconditions are for a successful preparation for a Data Mesh architecture. Examples of this are, for example, the presence of certain business capabilities or institutional artifacts. A framework that can guide this process can provide new insights for the positions of organizations in relation to this new data architecture design. To address this aforementioned focus point, the following research objective has been formulated:

"To design a framework that enables organizations to assess their readiness for migrating to a Data Mesh architecture."

This paper aims to identify the critical influencing factors that influence an organization's readiness towards a Data Mesh migration. In order to identify these factors, a study has been set up that develops a Data Mesh Readiness Model. In order to enhance its theoretical and empirical foundation, the model is applied in practice by means of case interviews and test cases. This application provides a practical representation of the way of using the readiness model, and serves as a practical evaluation. This paper is structured as follows. In the following section, a Data Mesh architecture and its characteristics are elaborated on. Next, the research approach for the design and development of the Data Mesh Readiness Model is presented. The paper is concluded with a general recommendation on potential results of the model, followed by a general conclusion on the research and directions for future research.

II. DATA MESH ARCHITECTURE DESIGN

A Data Mesh is a domain-driven distributed architecture design, that is all about "recognizing and identifying data domains in an organization, as well as constructing an architecture based on numerous components that make up these domains" (Hokkanen, 2021). The main objective of a Data Mesh is to eliminate the challenges of data availability and accessibility at scale (Dehghani, 2022). The domain-driven design characteristic of a Data Mesh architecture tackles issues when different business units are working on large projects, speaking different business languages and formulating different requirements for that project, as is happening in monolithic architectures (Braun, Bieniusa, & Elberzhager, 2021). The bounded context in a domain-driven design defines clear boundaries for every business domain in the organization, ensuring that domain experts and data engineers within that domain are working closely together (Braun et al., 2021). The architecture is thus consisting of domain-driven components that operate interoperable and independently of each other (Enyo-one Musa, 2021). It is essentially an organizational change of current enterprise data architectures, rather than a technological transformation, in which the main focus lies on distributed data domain teams consisting of domain-specific data engineers and data owners (Fleury, 2021). This distributed Data Mesh architecture is built upon earlier alternatives to monolithic architecture and promises a more empowered, scalable, agile architecture comprising of multi-function teams and a domain-driven business structure, aimed to address the organizational challenges that were neglected by earlier enterprise architecture designs (Enyo-one Musa, 2021; Ray & Pal, 2020). Figure 1 gives a simplified representation of a Data Mesh architecture design.

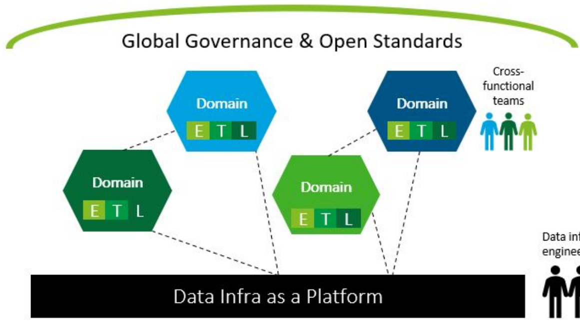

Fig. 1. Data Mesh Architecture Design

A Data Mesh is built upon four core principles, which are a combination of existing ideas on Distributed Architectures and Domain-Driven Design (DDD) (Dehghani, 2022). These principles are (1) Domain-Oriented Decentralised Data Ownership and Architecture, (2) Data as a Product, (3) Self-Serve Data Infrastructure as a Platform, and (4) Federated Computational Governance.

A. Domain-Oriented Decentralised Ownership

The first principle deals with decoupling and decentralizing the traditional monolithic data platform into a distributed variant. This decentralization is domain-oriented: that is, the decentralized components of the former large whole are selected on the domain they carry within the organization (Goniwada, 2022). The motivation behind this idea is about placing the ownership of analytical data for business domains close to the data itself, so that the processing, managing and analysing of the data can be done independently by the domain teams itself (Dehghani, 2022). The domains host and serve their datasets in an easily consumable way, while being close to the point of destination of the data (Serra, 2021).

B. Data as a Product

The second principle deals with the way data is viewed within the organization. Where data in the earlier data architectures was mostly considered to be a huge data repository, in a data mesh data is considered as a product that is ready-to-use for analytical purposes. Since the data in a Data Mesh becomes a prior business asset, the *Product Thinking* approach is applied to the data in a Data Mesh (I. A. Machado, Costa, & Santos, 2022). This approach sees data as a product, and this product needs to have a set of certain characteristics that maintain the quality of the data and the efficiency of the Mesh (Dehghani, 2022). These characteristics are in scientific literature formulated as the DATSIS principles: the data must be Discoverable, Addressable, Trustworthy, Self-Describing, Interoperable and Secure (I. Machado, Costa, & Santos, 2021).

C. Self-Serve Data Infrastructure as a Platform

In order to enable the domain's cross-functional teams to share their data, a self-serve platform must be built into the Data Mesh (Dehghani, 2022). This platform is built and maintained by the central IT organization, is domain agnostic and must enable users to surface data lineage across the Mesh. In addition, this platform enables users to control the full life cycle of individual data products, as well as to manage a reliable mesh of interconnected data products (Dehghani, 2019).

D. Federated Computational Governance

The last principle deals with the construction of a federated and global governance among the Data Mesh. The organizational structure of this federated governance and global standards is challenging, but very important for the functioning of the Data Mesh. The governance needs to set global rules on local data practices and decisions, and consists of principles underpinning the scope of the global governance. Moreover, the global governance is lead by a team that consists of members of the group from the domain, the self-serve platform and global compliance stewards that both have global and local incentives. Moreover, the global decision standards contain platform capability automating the decisions and computationally validating it continuously across all data products and domains (Dehghani, 2022).

III. DATA MESH BENEFITS AND LIMITATIONS

In order to maintain a critical view of this new innovation, the advantages and disadvantages of a Data Mesh architecture have been investigated. These advantages and disadvantages are considered during the entire study, in order to stimulate an objective outcome of the study. The advantages of a Data Mesh architecture can primarily be found in improving the accessibility of the data, due to the fact that the data products from the domains are published in a discoverable way on an easily accessible platform. In addition, the domain-oriented decentralization ensures that both the operations side of the data platform as well of the analytics side work together in domain teams, which can have a positive effect on the lead times between requests from these teams. Finally, a Data Mesh architecture advocates a clearer and more logical appointment of ownership over data and its products in the organization, which benefits the responsibility over and the quality of the data (Saurabh, 2021).

The limitations of a Data Mesh architecture lie primarily in a greater need for data specialists within organizations, because each domain is responsible for its own data processing. Solutions for this can be sought in hiring more data specialists, or upskilling the current people in the organization. In addition, within a Data Mesh architecture there will always be a need for a central body in the organization that monitors the governance and interoperability of the cross-domain collaboration. The boundaries between the dependencies of the central team and the independencies of the domains themselves can become vague (Saurabh, 2021). Finally, it is very important that there is a proper alignment of the interfaces of the different business domains, because otherwise there is a risk of unintegrated data silos in the domains. These silos can cause many copies of data to be made, which can be problematic in terms of data latency and data quality. In conclusion, a Data Mesh architecture should primarily be seen as a socio-technical approach for solving contemporary problems with the accessibility, management and analysis of data within large and complex organizations. The organizational challenge primarily consists of changing the way of thinking, namely about data as products, and the way of working, namely with a high degree of self-service (Goetz, 2022). And although this is more of an organizational issue than a technological issue, technology is needed that can make this possible and prevent the organizational chaos in a distributed architecture. The aim of this technology should be to support and enable the various principles of the Data Mesh, while complying with a well-thought-out and partially federalized governance model that enables the Data Mesh and structures and organizes its decentralized nature.

IV. RESEARCH APPROACH

The aim of this research is to design a model that enables organizations to assess their readiness for migrating to a Data Mesh architecture. In order to be able to grasp Data Mesh readiness within organizations, a theoretical model has been developed that contains factors that influence this readiness. In order to give a clear overview on influencing factors on Data Mesh readiness within all facets of the organization, the conceptual framework capturing Data Mesh readiness should contain a number of dimensions that address the organizational facets and a number of factors that correspond to these dimensions (Fraser, Moultrie, & Gregory, 2002). The identification of the components of the model can be achieved by carrying out extensive literature research here (de Bruin & Rosemann, 2005). This literature review resulted in a list of components that influence organizational readiness in relation to a new technology, but are not yet specifically aimed at a Data Mesh architecture. Due to the novelty of the subject and the lack of scientific literature on the subject, the first list of influencing factors has been supplemented with identified factors from thirteen exploratory interviews with Data Mesh experts. These experts identified new influencing factors on Data Mesh readiness as well as evaluated factors extracted from the literature. Together, this resulted in the compilation of a list of influencing factors that will form the basis of the Data Mesh Readiness Model.

Since Data Mesh readiness emerges from both the technological dimension of the organization, as well as from the organizational side, the Data Mesh Readiness Model has been designed upon a combination of the structures of technology adaption models and organizational change models. The two-dimensional design ensures that both the organizational side as well as the technological implications can be taken into account when assessing Data Mesh readiness. The model measures the readiness in various organizational steps on the horizontal axis, whereby a factor within the associated factor dimension can be assessed within each step on the vertical axis. The factor dimensions are retrieved from TOGAF's Business Transformation Readiness Model (TOGAF, 2018). Figure 2 presents the Data Mesh Readiness Model.

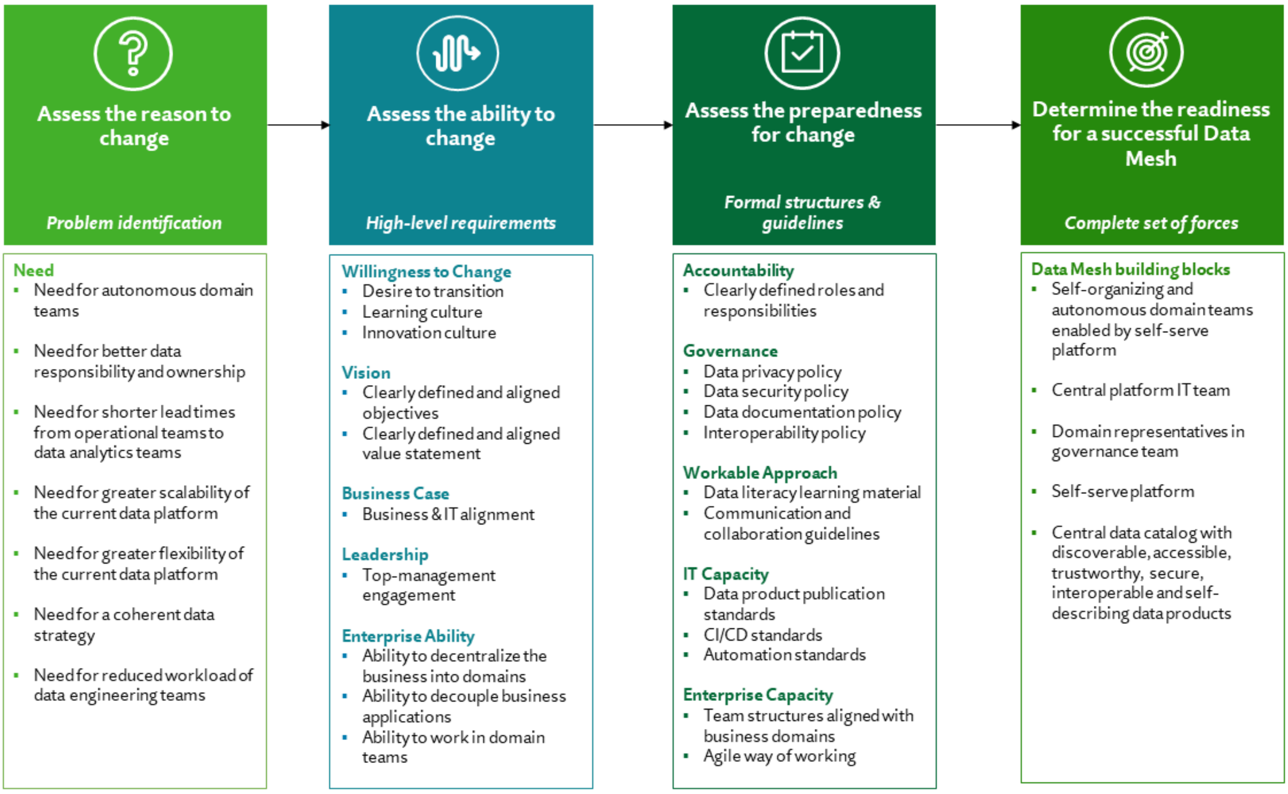

Fig. 2. Data Mesh Readiness Model

V. FACTORS AFFECTING DATA MESH READINESS

Based on the literature found in relation to IT architecture readiness, as well as the empirical research conducted with Data Mesh subject-matter experts, a large number of factors influencing Data Mesh readiness were found. Using the dimensions of the TOGAF Business Transformation Readiness Model (TOGAF, 2018), as well as the theoretically and empirically identified factors, the following readiness steps, factor dimensions and factors comprise the Data Mesh Readiness Model.

A. Assessing the reason to change

1) Need:

- **Need for autonomous domain teams**: Indication whether there is a need for domain teams to be able to do their own data analysis. Can they be independent in how they address their own domain analytics? Are the teams experiencing organizational and technical dependencies? Is the accountability of the data close to where the data is produced?
- **Need for better data responsibility and ownership appointment**: Indicate whether there is a need for greater knowledge over the data that is worked with, in terms of the fact that the people that work with the data are the people that know the data. Do the data engineers know the value and quality of their own data sets? Are the domain experts involved in delivering their data to the data analysts? Do the data teams understand what they contribute to in a business context?
- **Need for shorter lead times**: Indication whether there is a need for shorter waiting times for the analytics team to have their requests to the engineering team processed.
- **Need for greater scalability of the current data platform**: Indication whether there is a need for the current data platform to process bigger amounts of data through the data platform in the near future. Moreover, indication of the possibility and ease of adding business units to the data platform? What is the possibility and ease of growing the resource utilization of the current data platform?
- **Need for greater flexibility of the current data platform**: Indication whether there is a need to ease making changes on the existing IT system/data platform. Is the platform build on premise or is it easily accessible by others? Is the platform changeable, or is it too complex to be changeable? Is there siloed information or siloed data in the platform?
- **Need for a coherent data strategy**: Indication of the existence of a coherent data strategy. Are the data engineers rebuilding the same platforms? Are the data engineers and data analysts aware of each other's needs and purposes? Is there a coordinated, common data strategy? Is data an important asset of the organization? Is decision-making informed on data?
- **Need for understanding of the domain complexity**: Indication whether there is a need for the lessening of the current domain complexity, in terms of the existence of comprehensibility of inner domain processes and understandability of the end to end processes. Do the

domains understand what they contribute to? How many data sets exist in the domains? Do the domains implement their processes independently and in a siloed manner?

- **Need for reduced workload of data engineering teams**: Indicate whether there is a need for healthier working conditions for the data engineering teams. Is the team always full? Is the data engineering team causing a bottleneck in the data processes? Are the data engineering teams satisfied with their workload (i.t.o. stress levels)? Are the data teams making their deadlines (i.t.o. quality of work)?

B. Assessing the ability to change

1) Willingness to Change:

- **Desire to transition**: Indicate whether there exists a commonly shared desire to transition to a Data Mesh within the organization as a whole and among the individuals within the organization. Are they willing to take the efforts needed to establish this transition? Are the existing teams eager to be autonomous? Is there a willingness to participate in the Data Mesh migration?
- **Learning culture**: Indicate whether there exists a culture that supports learning. Are the people motivated to learn new procedures, such as working Agile? Is there management involvement in establishing this culture? Are there resources available to support this culture? Do the organization's mission and vision statements support this culture?
- **Innovation culture**: Indicate whether there exists a culture that supports innovation. Do the people have interest in working with data more efficiently? Are the people open for adapting to new technologies such as DevOps and DataOps? Is the organization as a whole data-driven? Do the organization mission and vision statements drive this culture? Are there resources available to support innovations? Are there organizational processes or facilities implemented that facilitate this culture? Are people experiencing long decision paths when implementing innovative projects?

2) Vision:

- **Clearly defined objectives**: Indicate whether the objectives for migration to Data Mesh are clearly defined and aligned with the people in the organization. Why does the organization wants to migrate to a Data Mesh? What are the needs that are addressed by the migration to a Data Mesh? Does there exist common awareness about these objectives? Are there performance measures in place for measuring the progressions?
- **Clearly defined value statement**: Indicate whether the value of a Data Mesh to the organization is clearly stated and aligned with the people in the organization. What is the value that it will bring to the organization? Is there an idea on the returns on investment? Do there exist reflections on past projects that brought value?

3) Business Case:

- **Business & IT alignment**: Indicate whether the data operations and processes are aligned with the business needs. What are the most important business requirements, and are they in line with the data strategy? Is the data team aware of their contributions to the business performance? Do the business users wish to be involved in the curation and processing of data from source applications to the reports that they use? Has the organization benchmarked itself with respect to similar organizations? Is there an understanding aout the impact that IT has on business processes? Is there an understanding about the amount/existence of organization strategic goals and requirements supported by IT strategic goals? Are Business and IT aligned with prognosed future growth? Is there an undertsanding of the digital capabilities needed to support the organization's business strategy? Will the Data Mesh architecture improve the organization's current Business and IT alignment?

4) Leadership:

- **Top-management engagement**: Indicate to what extend the top-management of the organization is engaged with the migration to a Data Mesh. Does the migration align with their interests? Do the business leaders understand the Data Mesh? Do they understand the need for a Data Mesh? Have they allocated time and resources to the development of the Data Mesh migration?

5) Enterprise Ability:

- **Ability to decentralize the business into domains**: Indicate to what extend the organization is able to decentralize its central architecture (in terms of processes and organizational design) into different domains. Does the organization understand the domain driven design principles? Can the organization map its business into different domains? Is there an idea on the domain boundaries? Are the boundaries distinct and explicit? Is there a reference design on the perceived decentralization of the organization? Is there an idea on the preferred degree of decentralization? Is the organization able to shift (a part of) the centralized decision-making into decentralized decision-making (e.g. across multiple teams?) Is the organization able to shift (a part of) the centralized roles, functions and tasks into decentralized roles and responsibilities?
- **Ability to decouple business applications**: Indicate to what extend the existing business applications (in terms of data processing) can be decoupled into the domain-oriented teams. Can the organization appoint the existing tables and pipelines to domain teams? Can the existing monolith be split up into smaller systems? Is the organization very much application focused, so that they don't break well into domains? Are the business processes very much centralized in ERPs? Is the organization able to establish a decoupled application architecture that allows each component of the application to perform its tasks independently in the domain teams? Can the components of the business application remain autonomous of each other? For organizations using pre-built applications: can these

applications independently run across the Data Mesh (e.g. across multiple autonomous teams)?

- **Ability to work in domain teams**: Indicate to what extend the domain teams understand the new domain-oriented way of working. Do the data analysts, that were primarily focused on getting insights from the data, understand how to maintain data pipelines? Do the domain teams understand their extra responsibilities? Do they understand the concept of a Data Mesh? Do they understand the data-as-a-product thinking? Do they understand data products? Do they understand how to work together?

C. Assessing the preparedness for change

1) Accountability:

- **Clearly defined roles and responsibilities**: Indicate to what extend the roles and responsibilities are correctly described. Will all the domains have a data product manager? Will the domains have data engineers, and are they given space to spend more time understanding the data? Do there exist data ownership guidelines? Does there exist no claimed ownership over the same applications? Are the roles and responsibilities of the team members in line with their personal skillsets?

2) Governance:

- **Data privacy policy**: Indicate whether there exists a data privacy policy.
- **Data security policy**: Indicate whether there exists a data security policy.
- **Data documentation policy**: Indicate whether there exists a data documentation policy, that states how to document what a data product means and how to define a domain.
- **Interoperability policy**: Indicate whether there exists an interoperability policy, that states how to enable interoperability between domains and their domain related data-products. Does this policy allow for the fast consumption of data products?

3) Workable Approach:

- **Data literacy learning material**: Indicate whether there exists learning material that enhances the data literacy among the people in the organization, in order for them to work autonomously on data products without the need for data engineers. What tools do they use for their data products and how do these tools work? How to understand the data? How to publish the data products? Does it state how to make data-based decisions? Does it support understanding, learning, analyzing and managing data products? Does it create overall data awareness among the organization?
- **Communication and collaboration guidelines**: Indicate whether there exist guidelines that enable and ease communication and collaboration between the various domain teams. How to enable cross-domain collaboration through data exchange? How do you collaborate on a data product, how do you work together to build a data product?
- **Knowledge sharing platform**: Indicate whether there will be knowledge-sharing facilities in place that enable the education, upskilling and sharing of knowledge within the organization. Is the platform accessible and interactive? Does the platform support Data Mesh engagement? Does it include information, guidelines, policies, and instructions on Data Mesh architectures?

4) IT Capacity:

- **CI/CD standards**: Indicate whether there exist CI/CD standards that enable continuous integration and continuous delivery. Do these standards support the efficiency of the performance of the Data Mesh? Are these standards frequently measured in terms of cycle time, change failure rate and deployment frequency?
- **Automation standards**: Indicate whether there exist automation standards for standard software procedures like data quality checking, pipeline creation, etc. Is the organization able to automate software and data processes? Do these automations improve the efficiency of the performance of the Data Mesh?
- **Data product publication standards**: Indicate whether there exist standards that state how and when to publish data products. What is the standard publication format? What needs to be in a data product? Are there technology standard with which people should build a data product?

5) Enterprise Capacity:

- **Agile way of working**: Indicate whether the organization supports and encourages an Agile way of working throughout the organization.
- **Team structures aligned with business domains**: Indicate whether the domain teams are business aligned, int terms of that they have responsibility for a certain part of the business. Will the domain teams not be too far away from the business cases? Does the organizational structure support a Data Mesh way-of-working?

D. Determining the readiness for a Data Mesh

1) Data Mesh building blocks:

- **Self-organizing domain teams enabled by self-serve platform**: Indicate whether the organization has established self-organizing domain teams, that are able to cross-collaborate with other domain teams and that are enabled to autonomously work on their own data products via the self-serve platform.
- **Central platform IT team**: Indicate whether the organization has established a central platform IT team that bears responsibility over performing and maintaining the self-serve platform.
- **Domain representatives in governance team**: Indicate whether the organization has established a central governance team consisting of domain representatives, that administrates the policies and standards required for the performance of a Data Mesh.
- **Self-serve platform**: Indicate whether the organization has developed a self-serve platform, maintained by the central platform IT team, that enables the domain teams to work autonomously on their data products and collaborate with other domain teams.

- **Central data catalog**: Indicate whether the organization has developed a central data catalog which enables the sharing, discovering and accessing of trustworthy, secure and self-describing data products throughout the organization.

VI. THE DATA MESH READINESS MODEL

There is a need for a study that investigates the factors that influence readiness for a Data Mesh migration and the necessary capabilities needed within organizations to start this migration. The Data Mesh Readiness Model addresses this need and enables organizations to test whether they are ready to migrate to a Data Mesh architecture, in a multi-dimensional and step wise manner. In order to elaborate on the structure of the model and the model itself, the Data Mesh Readiness Model is elaborated on in this section.

A. Description of the Structure of the Model

The model is structured onto two dimensions: on the horizontal axis the model is build up in several **organizational steps** towards Data Mesh readiness, and on the vertical axis the model consists of several **factor dimensions** influencing the particular readiness steps.

Due to the complex nature of a Data Mesh architecture in both organizational and technological areas, the design of the model provides structure in the long-term migration process of organizations towards Data Mesh readiness.

In addition, because it became apparent during the formulation of the knowledge gap that it is not always clear to organizations whether a Data Mesh addresses their respective needs, the first step of the model consists of problem identification that must be completed before the process is started. If this problem identification shows that a Data Mesh architecture does not sufficiently address their needs, it is not necessary for the organization to resume the next assessment.

The final step of the model covers the overall technological and organizational readiness for a Data Mesh architecture, based on the structural principles of the Data Mesh concept. A remark that should be made here is that it is not necessary for all organizations to implement every specific principle in detail: going through the model will show which parts of the organization need more elaboration than others.

B. Description of the Use of the Model

The model can be used by organizations to self-assess their Data Mesh readiness, and to identify improvement opportunities based on the results of the assessment. It therefore allows for organizational learning, aiming to facilitate faster decision-making regarding the presented improvement-areas of the participating organization. When using the model, it could be beneficial for organizations to choose a group of diverse participants with different roles within the organization. Due to the broad scoping of the model, ranging from strategical and organization-wide dimensions to operational and technology implementation dimensions, readiness for a Data Mesh architecture could be assessed within all these different layers of the organization.

Due to the process and lengthy nature of a Data Mesh migration, it is recommended to use the assessment as a guideline throughout the entire process. Since it is unrealistic to assume that a Data Mesh migration can be completed within one iteration step, the step-by-step assessment can be used to assess the progression of the different steps. It is also possible to use the assessment in later readiness steps to assess earlier steps, to improve or distort implementations that have already been done. In other words, it is recommended to take the assessment repeatedly during the Data Mesh migration process. In this way, the as-is state of the organization and the aspired to-be state can be assessed frequently and in a structured manner, so that the ambition levels that the organization is seeking to achieve can be determined time and again. In this way, after each assessment, the gap between the as-is state and the desired state of the organization can be analyzed, on the basis of which an improvement path can be formulated.

The model should primarily be seen as an organizational guideline for structuring and organizing a multi-dimensional Data Mesh migration, rather than as a quantifiable assessment tool. The reason for this is twofold. First, a Data Mesh migration often depends on the specific organization in which it is performed, and it is therefore difficult to give a standardized quantitative rating to the degree of factors present. Secondly, it is beyond the aim of the assessment to present a one-off rating: after all, this number does not provide sufficient insight into the independent improvement areas and could lead to the risk of overseeing specific organizational challenges.

VII. RECOMMENDATIONS FOR ORGANIZATIONS

During the demonstration and test phase in this research, the model was tested in practice among organizations from different sectors. From the empirical data obtained from these applications, recommendations can be generated that are based on the consensus about the best practices for Data Mesh migration of these organizations. These recommendations are presented below.

A. Recommendations regarding Willingness to Change

Organizations that score low on the Willingness to Change dimension can improve themselves by starting with a clear understanding of the concept of Data Mesh and the value it brings to the entire organization. This understanding is supported by a supportive culture in the field of learning and innovation. This culture can be fostered by, for example, setting up extensive R&D teams, facilitating data literacy workshops in the organization, making data awareness training accessible to everyone and upskilling the domain teams so that they can create and maintain data products in the future. Ensure within the organization an appetite for organizational change, and leverage the right people who have the right influence throughout this process.

B. Recommendations regarding Vision

In the absence of a clear vision towards a Data Mesh architecture, it is could be useful to start with an appropriate data strategy for the organization. Within this strategy, tactical objectives and strategic objectives can be formulated. In addition, it is recommended to map out where redundant work is currently being performed. Organizations that want to become data-driven ultimately want to base their decision-making on facts and insights from the data. A marketplace of data could be established within the organization, where people can find and use data for these insights. In addition, it can be useful to measure the demand for value creation from different perspectives within the organization. Not only the business side of the organization should be involved, but also the operational teams. These teams should attain a focus on the consumers of their data, and the value they deliver to these consumers.

C. Recommendations regarding Business Case

In the absence of a clear Business & IT alignment, it is useful to reflect on when mapping the organization around each other's operations and values of the data. It is not recommended to look in a customer supply relationship between business and IT, but to change the perception towards an expectation-based relationship. What do the teams expect from each other, and how do they support each other? In this way, the operational teams can gain a better insight into what they contribute to business-wise. The alignment can also be found in improved collaboration, so that common interest in data can arise through improved data awareness. Programs as formulated in the objectives support this awareness. These programs can be initiated by the business users who want to become more actively involved in the processing of data.

D. Recommendations regarding Top-Management Engagement

In the absence of top-management engagement, an improved value statement of the Data Mesh migration for these executives can be looked into. Top-management support is recommended for an organization-wide migration, because they can translate the ambition across the entire organization. Additionally, their motivations and directions should be aligned with the focus areas of a Data Mesh architecture. The focus areas of a Data Mesh lie in developing an architecture that is easy to evolve and access, and to better accommodate large amounts of data. Where there is a need for improved management of the complexity of the domains, a Data Mesh architecture can contribute to improving the understanding and trustworthiness of the data. Workshops or written value statements that showcase or describe these contributions can help create support among top management.

E. Recommendations regarding Enterprise Ability

With deficiencies in the enterprise ability to migrate, it is recommended to realize that decentralizing and decoupling current monolithic structures is a lengthy and iterative process. Recommendations within this process are, for example, to start identifying the domains, and to consider how these domains can be enabled to work autonomously. Not all domains have the capability to work autonomously, so it is recommended to start slowly with the domains that are the most mature and eager to be autonomous. These domains can serve as an example and pilot study. Outlining the business domains on a map can help with the decentralization process, for example based on existing domains that already own business artifacts or based on each vice president within the organization. The ownership boundaries are formulated so that the domains do not step on each other's toes.

In line with this, it is useful to realize that not every organization is able to decouple the central platforms. For these organizations, a logical split in the domains can be chosen, whereby each part of the central application belongs to one domain. Perform a translation which allows to decouple the operational plane from the analytical plan in this application, so that an abstraction layer is created between business application and analytical application.

Finally, the ability to work in domain teams can be increased by looking closely at the demand or data engineering skills in each domain. Not all domains have a constant workload over time, so at the beginning of the migration it is possible to work with, for example, a center of excellence of a pool of people who are able to jump into domain problems. It is recommended to retain a certain amount of domain knowledge within the data engineers. In addition, a migration program could be established that redesigns the entire organization based on domains and processes around the domains. This program is carried across the organization as a whole and ensures that the most important concepts - such as the definition of a data product, how you publish, how you manage data - are clear and understandable to the people who become part of a domain.

F. Recommendations regarding Clearly defined Roles and Responsibilities

Firstly, if there is a shortcoming within this dimension, it is recommended to realize that not everything has to be perfectly defined, but that there is a clear ownership alignment among the domains and the organization as a whole. A data product manager in any domain should have the information what the data product means, and know and understand the data product. This improves trust in data products between the different domains. Certain roles that could be divided in this context are, for example, those of a data owner in terms of business insights, a data custodian in terms of IT responsibilities and a data steward as data product manager and to bridge the gap between both teams.

G. Recommendations regarding Governance

It has become clear from this study that few recommendations can be made in the field of setting up the governance model correctly. Recommendations were to focus on making it easy to share and access data within the Data Mesh architecture, within the regulatory boundaries. Automation and tools help with this, but are not necessary. In addition, a central governance body is needed that can enforce the organization-wide standards and rules across the domains. Within the domains, governance must ensure that the data is correctly defined and placed within the correct business context. Possibilities for realizing democratized governance are, for example, with a governance forum in which the domains can provide input and be part of the decision-making process. Formal governance can be made available on this forum so that certain policies are discoverable. From the technology perspective, components can be built into the platform that can apply certain automated rules and policies.

H. Recommendations regarding Workable Approach

If there is a shortcoming within the workable approach dimension, the needs of the cross-domain collaboration could be looked at with regard to communication and collaboration. This can be done through data exchange, recurring meetings or collaboration agreements. With many inter-domain dependencies, data contracts or service level agreements can be drawn up to improve reliability. It is also possible to look at a place in the organization, physical or digital, where this inner- and cross-domain collaboration can take place.

With regard to data literacy, it depends on the choice of technology within the organization. With more complex technology choices, it is more important to have a greater degree of data literacy. You can also opt for more modern tools that do not require data engineers. Data literacy can be improved through a knowledge sharing platform, data literacy workshops or data awareness programs.

I. Recommendations regarding IT Capacity

In case of shortcomings in IT capacity, it is recommended to prioritize. Where continuous integration and automation standards have not been applied, it does not have the highest priority to develop them. Continuous deployment is more important when it comes to managing multiple data products. In addition, it is possible to look at developing approaches that counteract breaking use cases for downstream users, in case of changes to existing datasets. In addition, it is also up to the domains themselves to determine with which maturity in terms of CI/CD and automation they want to work on their inner-domain processes. Important to start with are data product publication standards. If they don't exist yet, they should be set up to structure this process. These standards contain, for example, APIs for the structure of data, data contracts for the trust in the data, a description of the data so that consumers can use the data, or a technology standard with which data products can be built.

J. Recommendations regarding Enterprise Capacity

In case of shortcomings in the Enterprise Capacity dimension, it is recommended to start by recording the new team structures. A certain alignment with the business domains is suggested, because in this way business domain knowledge can be preserved within the Data Mesh domains. For example, this domain knowledge can come in useful when feeling the responsibility about the domains, and it can be easier to feel this responsibility if it already existed for a certain part of the business. In addition, this way of structuring ensures that the Data Mesh domains are not placed too far from the business.

VIII. CONCLUSIONS

When determining readiness for a Data Mesh migration, the Data Mesh Readiness Model offers guidance in assessing the as-is situation of organizations, and aims to mobilize decision-makers to start their migration by providing insights to what improvement areas to focus on. These insights are based on the perspective that a Data Mesh architecture mainly revolves around the organizational structure, architectural decisions for moving data, and the supporting governance for it. The choice of enabling technology is rather an executive decision than an organizational priority.

The two-dimensional design of the Data Mesh Readiness Model is based on organizational steps towards readiness, measured on a scale of various technological dimensions. Using this two-dimensional scope, the entire socio-technical environment with respect to Data Mesh readiness can be included during the assessment of the model. The dimensions and associated readiness influencing factors are based on empirical data from exploratory interviews, model refinement sessions, refinement evaluation sessions, case studies, test cases and evaluation sessions. By demonstrating and testing the model in practice, insight has been created about a set of practical recommendations for organizations experiencing shortcomings in one of the model dimensions. These recommendations are based on the empirical data from the participants within this research and their approaches towards Data Mesh readiness.

IX. FUTURE RESEARCH

First, the model in this study has been demonstrated in a number of different environments. Since the exploratory interviews and the case studies have shown that the readiness of organizations depends on contextual influences, it would be of added value to apply the model in more different contexts. Involving more different use cases could contribute to a broadened understanding of the readiness of organizations towards Data Mesh migration and how the influencing factors relate to it. For example, an application within a risk-averse and heavily government-regulated sector such as healthcare could provide interesting insights. The contexts within this research were often less regulated and more data-driven.

Second, an empirical follow-up study could focus more on recruiting more critical participants in the study. Because the topic is currently very popular among data-driven enthusiasts, it was difficult to find participants who could express themselves critically towards the topic and who also had empirical evidence for this. Follow-up research could take place in the future if there are more practical examples of Data Mesh implementations and their critics.

Third, follow-up research could focus on the factors that had little consensus during the case studies, such as establishing an appropriate governance model for Data Mesh architectures or addressing the decentralization and decoupling process. While this study has provided an overview of the as-is situation of participating organizations and offered different perspectives, follow-up research can delve deeper into these topics and explore the best ways to address these factors. For example, a Design Science Research into the design and development of a Data Mesh governance model could be of added value.

Fourth, follow-up research could focus on designing and developing an execution framework for implementing a Data Mesh architecture. While this research mainly focused on readiness and preparation prior to implementation, follow-up research can take a look at an actual implementation process of a Data Mesh architecture. Specific attention could be paid to certain implementation challenges that flow from the readiness assessment, or tackling the identified areas of improvement in a practical sense.

Finally, more research on the societal and organizational impact caused by the implementation of a Data Mesh architecture is proposed. Because of the novelty of the subject and the socio-technical context in which it takes place, it is of great value to know what the specific societal and organizational changes are that are involved in the large-scale implementation of Data Mesh architectures. This research can provide more insight for organizations to assess whether they are suitable for this architecture design and then also provide insights into the future vision of distributed architecture designs and the impact it has on society as a whole.

REFERENCES

Braun, S., Bieniusa, A., & Elberzhager, F. (2021, April). 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* (pp. 1–12). Online United Kingdom: ACM. Retrieved 2022-02-21, from https://dl.acm.org/doi/10.1145/3447865.3457969 doi: 10.1145/3447865.3457969

de Bruin, T., & Rosemann, M. (2005). Understanding the Main Phases of Developing a Maturity Assessment Model. , 11.

Dehghani, Z. (2019). *How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh* [Martin Fowler]. Retrieved from https://martinfowler.com/articles/data-monolith-to-mesh.html

Dehghani, Z. (2022). *Data Mesh: Delivering Data-Driven Value at Scale*. O'Reilly Media. Retrieved 2022-02-15, from https://www.oreilly.com/library/view/data-mesh/9781492092384/ (ISBN: 9781492092391)

Einav, L., & Levin, J. (2014). *The Data Revolution and Economic Analysis.* Retrieved 2022-02-15, from https://www.journals.uchicago.edu/doi/epdf/10.1086/674019 doi: 10.1086/674019

Enyo-one Musa, A. (2021). *Data Management Arcitectures - Monolithic Data Architectures vs Distributed Data Mesh.* Retrieved from https://towardsdatascience.com/data-management-architectures-monolithic-data-architectures-and-distributed-data-mesh-63743794966c

Fleury, J. H. (2021). *Data Mesh is a Solution for Functional Organizational Challenges* [Crystalloids]. Retrieved from https://www.crystalloids.com/news/data-mesh-solution-to-functional-challenges

Fraser, P., Moultrie, J., & Gregory, M. (2002, August). The use of maturity models/grids as a tool in assessing product development capability. In *IEEE International Engineering Management Conference* (Vol. 1, pp. 244–249 vol.1). doi: 10.1109/IEMC.2002.1038431

Furia, B. (2021, March). *Data Mesh – Rethinking Enterprise Data Architecture.* Retrieved 2022-02-15, from https://www.cuelogic.com/blog/data-mesh

Goetz, M. (2022, March). *Exposing The Data Mesh Blind Side.* Retrieved 2022-05-05, from https://www.forrester.com/blogs/exposing-the-data-mesh-blind-side/

Goniwada, S. R. (2022). *Cloud Native Architecture and Design: A Handbook for Modern Day Architecture and Design with Enterprise-Grade Examples*. Berkeley, CA: Apress. Retrieved 2022-03-07, from https://link.springer.com/10.1007/978-1-4842-7226-8 doi: 10.1007/978-1-4842-7226-8

Hokkanen, S. (2021). *Utilization of Data Mesh Framework as a Part of Organization's Data Management* (Unpublished doctoral dissertation).

Khazaei, H., Barna, C., Beigi-Mohammadi, N., & Litoiu, M. (2016, December). Efficiency Analysis of Provisioning Microservices. In *2016 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)* (pp. 261–268). (ISSN: 2330-2186) doi: 10.1109/CloudCom.2016.0051

Kulkarni, R. (2019). *Big Data Goes Big.* Retrieved 2022-02-21, from https://www.forbes.com/sites/rkulkarni/2019/02/07/big-data-goes-big/ (Section: Fintech)

Lawal Moshood, O., Ileladewa Adeoye, A., & Lawal Habibu, O. (2020). Microservice Architecture: A Better Option over Monolithic Architecture. *Technology (ICONSEET)*, *5*(19), 148–156.

Machado, I., Costa, C., & Santos, M. Y. (2021, August). Data-Driven Information Systems: The Data Mesh Paradigm Shift. *International Conference on Information Systems Development (ISD).* Retrieved from https://aisel.aisnet.org/isd2014/proceedings2021/currenttopics/9

Machado, I. A., Costa, C., & Santos, M. Y. (2022). Data Mesh: Concepts and Principles of a Paradigm Shift in Data Architectures. *Procedia Computer Science*, *196*, 263–271. Retrieved 2022-02-15, from https://linkinghub.elsevier.com/retrieve/pii/S1877050921022365 doi: 10.1016/j.procs.2021.12.013

Marr, B. (2018). *How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read.* Retrieved 2022-02-21, from https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/ (Section: Enterprise & Cloud)

Ray, P., & Pal, P. (2020, November). Extending the SEMAT Kernel for the Practice of Designing and Implementing Microservice-Based Applications using Domain Driven Design. In *2020 IEEE 32nd Conference on Software Engineering Education and Training (CSEE T)* (pp. 1–4). (ISSN: 2377-570X) doi: 10.1109/CSEET49119.2020.9206200

Saddad, E., El-Bastawissy, A., M., H., & Hazman, M. (2020). Lake Data Warehouse Architecture for Big Data Solutions. *International Journal of Advanced Computer Science and Applications*, *11*(8). Retrieved 2022-02-15, from http://thesai.org/Publications/ViewPaper?Volume=

11&Issue=8&Code=IJACSA&SerialNo=54 doi: 10.14569/IJACSA.2020.0110854

Saransig, A., & Tapia Leon, F. (2019, January). Performance Analysis of Monolithic and Micro Service Architectures – Containers Technology: Proceedings of the 7th International Conference on Software Process Improvement (CIMPS 2018). In (pp. 270–279). doi: 10.1007/978-3-030-01171-0_25

Saurabh, S. (2021). *Data Mesh: Design, Benefits, Hype, and Reality | LinkedIn.* Retrieved 2022-05-05, from https://www.linkedin.com/pulse/data-mesh-design-benefits-hype-reality-saket-saurabh/

Schultze, M. (2020). *Data Mesh in Practice: A Zalando Case Study.* DataBricks Summit 2020.

Selmadji, A., Seriai, A.-D., Bouziane, H. L., Oumarou Mahamane, R., Zaragoza, P., & Dony, C. (2020, March). From Monolithic Architecture Style to Microservice one Based on a Semi-Automatic Approach. In *2020 IEEE International Conference on Software Architecture (ICSA)* (pp. 157–168). doi: 10.1109/ICSA47634.2020.00023

Serra, J. (2021, July). *Data Mesh: Centralized ownership vs decentralized ownership | James Serra's Blog.* Retrieved 2022-03-08, from https://www.jamesserra.com/archive/2021/07/data-mesh-centralized-ownership-vs-decentralized-ownership/

TOGAF. (2018). *Business Transformation Readiness Assessment.* Retrieved 2022-03-09, from https://pubs.opengroup.org/architecture/togaf9-doc/m/chap26.html