Business-to-Business data sharing via data marketplace meta-platforms: Exploring governance mechanisms to enhance data sovereignty

Thomas van Velzen

Delft University of Technology

TU Delft

pwc

Business-to-Business data sharing via data marketplace meta-platforms: Exploring governance mechanisms to enhance data sovereignty

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Management of Technology**

Faculty of Technology, Policy and Management

by

Thomas Michiel van Velzen

Student number: 4347129

To be defended in public on September 15, 2022

Project duration : February - September, 2022

First supervisor	: Dr. A.M.G. Zuiderwijk - van Eijk	Section ICT
Chairman, second supervisor	: Dr. G. van de Kaa	Section ET&I
Internal advisor	: A.E Abbas MSc	Section ICT
External advisor	: B. Schouten MSc	PwC Netherlands

Acknowledgements

This master thesis project marks the end of my Management of Technology master program and hence also my time at the Delft University of Technology. During this project I have not only learned about data sovereignty and governance of digital platforms, with data marketplaces in particular, but several other skills that come with carrying out a research project as well. For example, contacting potential interviewees, conducting interviewees and analysing both existing and new data.

However, I could not have completed this project without the help of several people that have supported me along this journey. Firstly, I am very grateful to my first supervisor, Anneke Zuiderwijk-van Eijk, who did not only provide valuable feedback, but also gave helpful and practical suggestions when things were difficult or when I was struggling to take the next steps. Additionally, I would like to thank my chair and second supervisor, Geerten van de Kaa, who helped by providing valuable questions that lead to new insights and helped to improve the quality of the work. Thirdly, I am grateful to have Antragama Ewa Abbas on board during this project. His in-depth knowledge to the subject resulted in discussions that were not only helpful, but joyful as well. Furthermore, I am very grateful to the access he provided to his network when searching for interviewees.

Secondly, I would like to thank all the people of PwC Risk Assurance who were so open and welcoming during my internship. In particular, my gratitude goes to Bram Schouten and Issehak Boukhizzou. Even before the actual start of my internship, Bram contacted me to get to know each other. Furthermore, your help during the project helped to contact people within the organisation that could potentially help me with my research. Besides my project, I really enjoyed our conversations about PwC and all the answers you provided to my questions about working at PwC. A graduation internship is not only about carrying out a research project and to find out whether it might be the right spot for your future career, but is also about working alongside new people and having fun from time to time. I am very grateful to Issehak Boukhizzou for experiencing this and his help along the way. Lastly, I would like to thank PwC Netherlands for all the opportunities during my graduation internship at the Amsterdam office, not to mention the numerous social events that made me as intern really feel part of the team.

Last but not least, I would like to thank my parents, sister and friends (I am talking about you, Tim and Jeffrey!) for their continued support and occasional distractions that were very welcome from time to time.

Thomas van Velzen, Oegstgeest, September 2022

Executive summary

Modern organisations are increasingly dependent on data in their pursuit for new innovative products and services. Additionally, they process enormous amounts of data to operate on a daily basis. Furthermore, these organisations are no longer solely dependent on internal data resources, but on external sources as well. Sharing of business data is also acknowledged by governmental bodies such as the European Commission which has adopted the Data Governance Act recently. Consortia of both commercial, research and governmental organisations are currently exploring the opportunities and first implementations of data sharing at a larger scale. These efforts have lead to the first harmonised data sharing networks: data spaces and data marketplaces. However, many organisations still have the fear of losing control over their data which translates into a lack of data sovereignty. On top of that, there is currently fragmentation in the data sharing landscape which can make it more difficult for demand and supply of data to connect and can also harm trust in data sharing platforms. As a result, the idea of platforms that connect existing data sharing platform has risen: data marketplace meta-platforms. The question still remains, how do data providers stay in control over their data in this context? As a result, this research project tries to answer the question:

What governance and control mechanisms can enhance data sovereignty for business-to-business data sharing via DMMPs?

Data sovereignty: what to control?
Data providers that want to stay in control over their data while using data marketplaces should consider four areas: data ownership, data access, data usage and data storage. Whereas data ownership is about which organisation has generated the data or is the rightful owner to the data, data access is about which party can access it and under which conditions. Considering data consumers that want to create value from the data of providers, data providers need to stay in control of the actual usage of their data as well. For which purpose is the data used? Which party is acquiring the data? Is it a competitor? Ultimately, all data needs to find its storage somewhere, and data providers have to carefully consider what they prefer, depending on the specific use case. As a result, the platform has to offer flexible options regarding data storage, and for example separation between the storage of metadata and the actual datasets.

The current data sharing landscape: strategies in practice
After an analysis of the current data sharing landscape using industry literature combined with insights from expert interviews, it became clear that there are already clear strategies that are adopted by data sharing initiatives and data marketplaces to enhance data

sovereignty. First, they make sure that all participants accept a shared legal basis to establish a basic level of trust and to provide grip when disputes arise. Specifically for access and usage control, they have developed processes were data providers and data providers have a clear responsibility and where a transaction can only take place after explicit consent by data providers. Furthermore, certification processes are used to screen new participants and to maintain quality among existing ones. However, current initiatives have to balance between generic and specific agreements considering different use cases and platform participants. Some key areas are standardised across the entire platform, for example identity, whereas other aspects, such as data models, are open for participants to decide.

DMMPs: meta-platform, challenges at scale
When integrating existing data marketplaces via a data marketplace meta-platform, data sovereignty becomes an even more delicate exercise. Although their ability to offer data providers a single stop to offer data to many marketplaces, data providers are hesitant and fear the dominance of Data Marketplace Meta-Platforms (DMMPs). Furthermore, many potential data providers are currently not prepared to start sharing data at scale yet, especially regarding sensitive but valuable proprietary business data. Even when providers enter a DMMP in the future, findings from this research show that they still have fears to grant access and to lose control over their usage and storage of data. Lastly, a large DMMP where many different industry players find each other can lead to disputes at scale as well. A purely legal enforcement of data contracts could not be viable for DMMPs.

Solutions: architectural, trust and tech
However, there are many potential solutions that could address the challenges for DMMPs mentioned above. Firstly, architectural decisions could positively influence data sovereignty for providers, especially when DMMPs succeed in creating a truly decentralised structure. This does not only entails data storage, but also shared decision making and potentially even shared ownership of the platform. One of the top priorities for DMMPs should be the building of trust among potential and existing participants. This entails giving autonomy for data providers when arranging a data transaction with their interested data consumer, but also careful management of identity of DMMP users. Certifying new entrants and periodically certifying existing participants can also help to establish trust, but please note that DMMPs will also have to be able to ensure data providers that only trustworthy data marketplaces can enter the DMMP. This means and extra task regarding certification compared to single data marketplaces. Furthermore, audits by third parties of the DMMP could also be helpful to build up a reputation of trustworthiness.

As development progresses, DMMPs need to retain trust among their platform, for example by continuous monitoring of data consumers to assure data providers that their interests are protected. It could also entail data licensing to help data providers retaining data ownership, and not only at the DMMP, but also after data has found its way to data consumers via connected data marketplaces. All these solutions can only work when DMMPs are able to manage their diverse and large network of participants. In order to achieve this, many

technology-based solutions are needed to support higher level goals across the platform. For example, using Multi-party Computation (MPC) to technically enforce data usage without disclosing the data provider's dataset to the consumer directly, but also labelling and tagging of data to help data providers keeping track of what happens with their data after successful transactions.

All in all, DMMPs can offer a valuable contribution to the data economy of the future and help to reduce fragmentation where every data marketplace is using different standards and conditions, which can cause data discovery issues for data consumers and difficulties for data providers to start sharing or monetising their data. However, this potential comes with many additional challenges, which will need to be addressed by a delicate architectural design where only the necessary aspects are centralised, but everything that can be decentralised will be decentralised. This supported by fine-grained access and usage controls, scale-resistant data licensing to retain ownership and literate user communities can lead to trust and protect data sovereignty. But please keep in mind, it is trust, but verify.

Table of contents

List of Figures

List of Tables

List of acronyms

AI	Artificial Intelligence
DMMP	Data Marketplace Meta-Platform
DSC	Data Sharing Coalition
E1	Expert 1
E2	Expert 2
E3	Expert 3
E4	Expert 4
E5	Expert 5
E6	Expert 6
E7	Expert 7
E8	Expert 8
E9	Expert 9
E10	Expert 10
E11	Expert 11
EC	European Commission
EU	European Union
ERP	Enterprise Resource Planning
GDPR	General Data Protection Regulation
IDS	International Data Spaces
IDSA	International Data Spaces Association
IT	Information Technology
MPC	Multi-party Computation

PTPO Platform-to-Platform Openness

SME Small and Medium-sized Enterprise

1

Introduction

1-1 Data, oil of the 21st century?

People like to speak of data as the oil for the 21st century very frequently. Modern organisations gather and process enormous amounts of data originating from their customers, suppliers and internal processes. The vast amount of sensors that is used in a modern manufacturing process generates even more data. Additionally, administrative processes like payment and procurement are processed digitally using Information Technology (IT) resulting in data that is valuable for modern firms. At the same time, organisations are faced with disruptive digital innovations, a highly competitive market environment and customers that are increasingly demanding. As a result, organisations can consider business data as an asset that can provide insights for process optimisation and new product development (Tallon, 2013). Consequently, access to data can be a factor for gaining a competitive advantage, both directly or as an enabler for digital innovation (Teece, 2018). The modern economy that depends increasingly on digital and data-driven innovation is sometimes characterised as a data economy. It might not come as a surprise that the European Commission (EC) estimates a rise in global data volume of 530 percent between 2018 and 2025[1]. Although this comparison between oil and data might sound tempting, there are still a lot of open challenges regarding data. Several characteristics of data require additional care compared to a good like oil, especially in a trading-context.

[1] https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en#examples-of-industrial-and-commercial-data-use

1-2 The emergence of data marketplaces

This demand for business data by modern organisations has driven the development of many data marketplaces. For example, a recent study by TRUSTS resulted in over 170 cases of data marketplaces (Simon & Natalia, 2021, p. 26). Traditionally, marketplaces bring buyers and sellers together to enable the transfer of goods or services. Similarly, data marketplaces enable data sharing between data providers and data consumers by using a digital platform. Several data marketplaces have emerged, both in the United States, Europe and China (Spiekermann, 2019). These data marketplaces are not only characterised by their technical infrastructure, but also by non-technical attributes, for example pricing model, revenue model and value proposition (Spiekermann, 2019). Additionally, these data marketplaces often focus at a specific industry sector or geographic region which has led to a fragmented data marketplace-landscape (Abbas, 2021; Richter & Slowinski, 2019). For data consumers, this fragmentation can lead to difficulties to discover the desired data (Simon & Natalia, 2021, p. 47). For the data marketplaces, it can be difficult to attract enough data providers and consumers. This could hinder the development of a critical market mass and subsequent network effects (Katz & Shapiro, 1985). For example, if German car manufacturers use their own German data market to buy and sell data to part suppliers, smaller Italian car manufacturers that use their own data market do not discover the German data. These market dynamics can lead to sub-optimal results for both German and Italian companies in this example. Last but not least, a highly fragmented market of data marketplaces can lead to divergence of the platforms (e.g. many different pricing conditions, different security provisions, variance of conditions to enter marketplaces). This divergence can harm trust in data marketplaces as a whole (Simon & Natalia, 2021, p. 12).

These rise of these data marketplaces is not only driven by business dynamics. Recently, the EC has presented the European Data Strategy which, among other things, focuses on "pooling European data in key sectors, with common and inter-operable data spaces."[2] These steps show that the EC is aware of the potential and challenges of the modern data economy. Within the European Union (EU) there is a single market that allows free flow of people, products, services and capital. To realise a common European data space, it is necessary to integrate the different data marketplaces that are already existing. In addition, the EC acknowledges the risk that individual EU member states adopt individual legislation that can make it difficult to achieve a single market for data within the EU and has recently adopted the Data Governance Act as a first step to work towards a single European market for data (Baloup et al., 2021, p. 6).

In summary, data and data marketplaces can offer potential for modern organisations, but the landscape of data marketplaces is still very fragmented. The question now rises: how could this challenge of fragmentation be solved? A potential solution is the integration of existing data marketplaces using an over-arching platform: a meta-platform (Chen et al., 2022; Kretschmer et al., 2022). A meta-platform is able to communicate with existing platforms

[2] https://digital-strategy.ec.europa.eu/en/policies/strategy-data

using shared agreements and standards and has the potential to simplify the data discovery process for data consumers with lowered transaction costs and higher quality standards (Floetgen et al., 2021).

Compared to traditional firms and markets, a platform as a meta-organisation is less hierarchical than traditional firms, but less uncoordinated than a pure market (Kretschmer et al., 2022). For value creation for example, a platform is dependent on complementors instead of being able to control all resources centrally as would be the case with a traditional firm. On the other hand, platforms can use governance mechanisms to steer the network of complementors that is linked to the platform towards a desired direction. This gives a platform more control over resources and complementors than would be the case in a pure market, where price is the main mechanisms to match supply and demand. A combination of this concept of a meta-platform with data markets results in DMMPs (Abbas, 2021).

1-3 Data sovereignty: a data provider's perspective

Besides the fragmented data marketplace landscape discussed in 1-2, there is another problem that remains open in the context of business data sharing via data marketplaces: data providers still face the risk of losing control over data. There is a lack of data sovereignty, the self-determination over the organisation's own data. Organisations are hesitant to share their data for several reasons, for example because they do not want to share their data with competitors. Additionally, these data providers can be afraid to grant data access to parties of which the identity is unverified. Please note that these fears could be amplified by the fragmented data marketplace landscape discussed in section 1-2.

This lack of data sovereignty can lead to several problems, but a key issue is that these data providers do not want to provide their data for sharing and trading with others. As a result, a lot of data can remain unused or only partly used. Looking at the level of an individual organisation, this could appear to be a minor problem. However, at the broader market and EU level, a lot of value can remain locked. Although data sovereignty will be discussed in more detail in 4, the importance of understanding and realising data sovereignty for future data sharing could already become visible.

Although it can be clear what data sovereignty is from a high-level (self-determination over data), how to achieve it is highly context-specific. Considering the potential solution that DMMPs can be in the future of data sharing to harmonise the fragmented data marketplace landscape, this research will focus on that particular DMMP-context to explore governance mechanisms to enhance data sovereignty for data providers.

1-4 Problem statement

1-4-1 Scientific gap

Within the scope of the single firm, data governance is placed under the larger practice of corporate governance in the literature (e.g., Khatri & Brown, 2010). Data governance is a collection of policies and procedures that determine how data should be managed (Tallon, 2013). Data governance that sets minimum standards for data (e.g. data quality, data performance), can be seen as outcome-based control mechanisms. However, several scholars have indicated that 'traditional' data governance mechanisms are not always applicable to modern multi-sided platforms and ecosystems (and DMMPs) (Abbas et al., 2021; Koutroumpis et al., 2020; Lee et al., 2018; Otto & Jarke, 2019). Further, de Reuver et al. (2018) call for more research on the governance of these platforms.

Part of why there is still a call for research on governance of digital platforms could be that although existing research has studied mechanisms that platform owners use to stay in control over app developers (e.g. Ghazawneh & Henfridsson, 2013; Goldbach et al., 2018), the impact of platform openness in the diffusion phase of the platform (e.g. Ondrus et al., 2015), the goal of governance by the platform owner is often gaining market share or optimising revenues. However, there is still a lack of research on digital platform governance of governmental or non-profit platforms where goals are different (e.g. protecting privacy) (Mukhopadhyay & Bouwman, 2019, p. 346).

By researching governance mechanisms to achieve a goal that is not directly related to finance, in a relatively novel context of DMMPs, this research tries answer calls for research regarding governance both for different goals (i.e. non-financial) in new contexts (i.e. DMMPs).

1-4-2 Research objective

Although proper data governance has been identified to overcome the risk of losing control over data (e.g., Lee et al., 2018) and DMMPs could be useful to reduce data marketplace fragmentation, governance mechanisms to enhance data sovereignty on DMMPs remain largely unexplored (Abbas, 2021). As a result, **this study aims to explore governance mechanisms to enhance data sovereignty for business-to-business data sharing via DMMPs.**

1-4-3 Research questions

To be able to achieve the research objective, several research questions will have to be answered by the research project. The main research question is supported by four sub-questions. The main research question is:

RQ: What governance mechanisms can enhance data sovereignty for business-to-business data sharing via DMMPs?

This research question addresses the concept of data sovereignty in the context of DMMPs. To be able to create an actionable research project, this question is decomposed in four sub-questions.

Sub-RQ1: What are the antecedents of data sovereignty in the context of a data marketplace?

Before the sub-questions regarding the DMMP-context can be addressed, it is necessary to get an understanding of data sovereignty and governance in the existing situation of data marketplaces. A review of academic literature will be used to identify the antecedents of data sovereignty in the context of a data marketplace. Additionally, this review will provide examples preliminary problems that can arise for the data marketplace context.

Sub-RQ2: What are the current strategies to enhance data sovereignty in data marketplaces and data sharing initiatives?

Before exploring the context of DMMPs, an analysis will be conducted of the current state of business-to-business data sharing. This analysis will derive strategies that are currently used or under development to enhance data sovereignty by analysing literature from practice (white papers, online resources, industry reports). Furthermore, expert interviews will be conducted to add expert perspectives from practice.

Sub-RQ3: How do the challenges of SQ2 translate into a DMMP-context and what additional data sovereignty related challenges can arise?

Next, the DMMP-context will be explored using expert interviews. The goal of this sub-question is to explore how the DMMP-context compares to single data marketplaces and other data sharing initiatives and what this means regarding data sovereignty challenges.

Sub-RQ4: Which potential solutions can address the data sovereignty related challenges identified for the DMMP-context?

To find solutions to address the challenges that were identified by Sub-RQ3, the experts from the interviews will be asked to give their view on potential solutions. The end goal is to use the output of this sub-question to develop a portfolio of governance mechanisms that can enhance data sovereignty for data providers in the context of DMMPs.

1-5 Research relevance

1-5-1 Managerial relevance

From a practitioner perspective, this research can help to get a better understanding of the data sovereignty considerations of data providers, specifically in the context of DMMPs. This understanding can be beneficial for the development of future and current data marketplace initiatives, both at the level of the larger data sharing network and at the level of the individual organisation. At the data marketplace-level, the data sovereignty related solutions can help to further refine the specific platform, especially to prepare for a consolidation in the data marketplace landscape. For example, specific data sharing agreements that data providers and data consumers need to share business data in a helpful manner. At the level of individual organisations, the findings from this research can help to prepare organisaton's data sharing strategies or to improve existing activities in data sharing. Considering that there is currently attention from both academia and governmental bodies (e.g. the EC) to address data marketplace fragmentation, the findings can help to prepare for the future. Lastly, this research can help both individual organisations and operators of data marketplaces and future DMMPs to get to a better collaboration, as the findings of these research can possibly lead to a better understanding of the interplay between the platform and data providers. Hopefully, this leads to a more effective sharing of business data that supports future digital innovation and contributes to solve current business and societal challenges.

1-5-2 Academic relevance

This research is also relevant from an academic perspective. Currently, there is already a stream of literature on digital platforms. Literature regarding data marketplaces is part of this stream and is also emerging. However, research on DMMPs and reducing data marketplace fragmentation is growing. By taking DMMPs as the research context to explore governance mechanisms to enhance data sovereignty, this research brings a novel contribution to the digital platform literature. At a broader level, this research also contributes to the stream of data governance literature, as it helps to better understand data governance at a cross-organisational and cross-industry context.

1-6 Research scope

This research project scopes on a few specific topics in addition to data sovereignty in the DMMP-context. In addition, the scope is specifically on business-to-business data sharing. The first parties that are exploring the usage of marketplaces for data sharing are primarily organisations, both commercial companies and governmental organisations. Although it could be possible that individuals use DMMPs in the future, considering the current state of data

sharing, the scope towards business-to-business was considered to be appropriate. Using this scope leads hopefully to insightful results for the parties that are probably the early adopters. The focus on business-to-business data sharing impacts both the type of experts needed for the interviews and the interview protocol, which is further discussed in chapter 3. Secondly, this research focuses specifically on the European geographical area. The EC has created the Data Strategy which aims to use data within the EU in a value-creating manner according to European values. As a result, using the European region as a geographical scope, can contribute to these initiatives. This impacts the type of interviewees as well, as ideally the majority of the interviewees works or lives in this area.

1-7 Thesis structure

After this chapter, existing research related to the (related) topics of data marketplaces and digital platform governance is discussed in chapter 2. The methodology that was used during this research project will be elaborated in chapter 3. Next, the findings from the literature review to identify the antecedents of data sovereignty will be discussed in chapter 4. Furthermore, the findings from the analysis of the current state of data sharing (either data spaces or data marketplaces) by analysing industry literature and expert interviews will be discussed in chapter 5. Chapter 6 will introduce the empirical findings from the expert interviews related to the DMMP-context, both related to data sovereignty challenges and potential solutions. Furthermore, the discussion and conclusion is provided in chapter 7. This chapter will close with the limitations of this research project and recommendations for future research.

2

Background

This chapter will start by providing a background based on existing research regarding digital platforms in general, with data marketplaces in particular. Section 2-1 starts with an introduction to digital platforms and data marketplaces, and will then introduce the concept of Platform-to-Platform Openness (PTPO) to arrive at meta-platforms and DMMPs · As a result, it will provide the background of the research context of this research project: DMMPs. Next, section 2-2 discusses governance, both for a traditional context (i.e. individual organisations), and governance in a digital platform context. One of the goals of governance can be to control certain people, processes or resources. As a result, control mechanisms are part of the larger topic of governance. However, as all control mechanisms are part of the larger group of governance mechanisms, in this report, both of these will be described as governance mechanisms to improve readability.

2-1 Data marketplace meta-platforms: a platform of platforms

2-1-1 Digital platforms

Whereas pure markets have the goal to couple buyers and sellers of goods and services, digital platforms often try to deliver additional services (Yoo et al., 2010). Platforms do not only try to match supply and demand, but try to realise value from the large pool of participants. Partially, they realise this by being highly flexible and relying on a modular architecture (Cusumano & Gawer, 2002; Jacobides et al., 2018; Smedlund & Faghankhani, 2015). This enables these platforms to host a wide array of complementors, often supported by pricing mechanisms to stay in control of these complementors (Boudreau & Hagiu, 2008; Cennamo, 2018). These complementors add value to the platform network by providing additional

services or products. An example could be eBay, that is the focal platform, but relies for payments on additional services by PayPal. Equally, PayPal uses services by credit card companies (e.g. Mastercard, Visa and American Express) to further process payments by its users (Teece, 2018).

The intense use of digital technology enables platforms and complementors to develop and change products and services more quickly. Additionally, transaction costs can be lowered by using digital technology and leveraging complementors (Smedlund & Faghankhani, 2015). The capability to quickly reconfigure also results in a greater level of resilience, as was seen after the COVID19-pandemic hit (Floetgen et al., 2021). This research by Floetgen et al. (2021) revealed that platforms did not only achieve resilience by diversification of their offerings, business model adaptation and service optimisation, but also that platforms fulfilled the role of meta-platform. This concept, where a platform becomes a platform-of-platforms is further elaborated on in section 2-1-4.

A stream of literature deals with the coordination of these platforms. This is an important issue, as these platforms can neither be considered as typical single organisations, nor as pure markets (Kretschmer et al., 2022). Both the platform, its complementors and end-users are interdependent (Boudreau & Hagiu, 2008; Iansiti & Levien, 2004). A supermarket chain that uses payment terminals (PoS) by company A, that make use of a digital platform by company B, is dependent on capable security software providers C. These security software providers will only be willing to contribute to the PoS if they feel that potential revenues can be captured. Accordingly, these revenues depend both on the revenue sharing model of the platform and the size of the group of PoS users in this example. This example illustrates an additional characteristic of digital platforms: two-sided network effects (Parker & Alstyne, 2005). The concept of two-sided network effects (Eisenmann et al., 2006) originates from the seminal work on network effects by Katz and Shapiro in which they investigate the relation between the utility that a user perceives and the group of other users of this particular product. For many digital platforms, participants (both end users and complementors) are more likely to be attracted to a platform with many other participants versus a platform with less users.

2-1-2 Data marketplaces

In this research project, data marketplaces are considered as a special instance of a digital platform. Data as a commodity (Schomm et al., 2013) will be further discussed in section 4. The practice of data sharing is not new. For example, studies on research data sharing by academics or household credit data sharing by businesses has been done for a long time (Koutroumpis et al., 2020). However, data marketplaces handle data differently, and instead of just providing a place to share data, they provide a platform for data as a trade good (Koutroumpis et al., 2020; Schomm et al., 2013). This means that data marketplaces need a business model to monetise these goods and services using these data. Additionally, they need to manage their complementors and users to achieve optimal value creation (Richter & Slowinski, 2019; Spiekermann, 2019; van de Ven et al., 2021).

Data marketplaces typically offer different types of services: trading facility of raw data and additional data-based services (Schomm et al., 2013; Stahl et al., 2016). Data marketplaces depend on data providers for the supply of raw data and data consumers for making revenues by selling the data. Additionally, complementors play an important role for data markets as they can provide additional components, for example services to process raw data into business intelligence services (Abbas et al., 2021; Huang et al., 2021). An example could be that banks supply (anonymised) client data to the data market which is further processed by a data market complementors to create risk profiles that can be used by mortgage lenders. In this example, the data marketplace (platform), banks (data supplier), data processing providers (complementors) and mortgage lenders (data consumers) work together to create value using raw data.

2-1-3 Platform-to-Platform openness

The previous sections have tried to elaborate on digital platforms and data markets. Before the concept of a Data Market Meta-Platform (DMMP) can be discussed, it is important to introduce PTPO first. Whereas a lot of research has dealt with the openness of a platform towards users and complementors (e.g., Ondrus et al., 2015; Schreieck et al., 2018), recent research focuses on the openness *between* platforms (Mosterd et al., 2021).

The choices that platform operators make can have a major impact on how the platform, its complementors and the relations between platforms evolve (Cusumano & Gawer, 2002; Kretschmer et al., 2022; Tiwana et al., 2010). For example, collective agreement on interoperability standards can attract more complementors to a platform (Mineraud et al., 2016). Similarly, when different platforms agree on key issues (e.g. revenue sharing, technical standards), networks of platforms can emerge. Several scholars have identified this phenomenon, often marked as platform ecosystems (Jacobides et al., 2018; Schreieck et al., 2016). Using the analogy with biological ecosystems (Iansiti & Levien, 2004), in a platform ecosystem the health of the system as a whole depends on the decisions and actions by the individual organisations in the ecosystem. This already indicates that platforms and networks of platforms are complex systems. Micro-decisions can have a macro-impact, and vice versa.

The decision to make a platform more open or closed is also an important governance decision (Mukhopadhyay & Bouwman, 2019). Opening up a platform can lead to attraction of more complementors and easier collaboration between platforms. However, it can also posit strategic threats and potential security issues. A platform that has very little requirements for new complementors can attract players of very low quality that might harm the platform. This issue was already known in the pre-digital era, for example when Atari created an overly open platform and harmed itself with low-quality complementors (Boudreau & Hagiu, 2008). To resume back to openness of digital platforms, de Reuver et al. (2018) argue that additional openness aspects of technologies behind digital platforms (e.g. APIs and SDKs) play an important role as well.

PTPO can be achieved by different approaches. Two platforms can open up to each other by sharing standards (e.g. data format) and by closely connecting their APIs for example (Floetgen et al., 2021; Richter & Slowinski, 2019). This increases inter-operabability between the platforms and when several platforms perform this action, a web of closely collaborating platforms emerges. Although different data marketplaces might now be able to communicate and transfer data with each other, data consumers and data providers still need to find their way in the web of platforms. The problem of fragmentation that was discussed in section 1-1 might still exist. More specifically, the issue of data discovery remains unsolved (Abbas, 2021). A platform that does not facilitate easy "search/discovery" lacks for one of its most important activities (Koutroumpis et al., 2017, p. 25). Additionally, this structure might still be vulnerable to the disappearance of individual platforms as this results in a gap in the network.

2-1-4 Meta platforms

Although PTPO is a requirement for platforms to create platform ecosystems, the main interest of this research is on a particular kind of platforms. Whereas platform ecosystems can have very different players (e.g. several platforms of similar size, a few large platforms with smaller competitors, etc), this research focuses on a system where one dedicated platform connects the other platforms. This over-arching platform is the meta-platform.

Using a meta-platform has several advantages over direct openness between platforms. Firstly, it can provide a stable basis for other platforms to connect to. From an infrastructural perspective it provides a set of technical components that is used within the network of other platforms. This use of a shared infrastructure can lead to better inter-operability between platforms. Additionally, the meta-platform can set quality and safety standards and act as an orchestrator (Mukhopadhyay & Bouwman, 2019).

On the other hand, a meta-platform can have negative consequences. Potential platforms that could link to the meta-platform can feel threatened by an over-arching player. This issue where larger players have to carefully balance their power over smaller players is also discussed in the literature on pre-digital platforms (e.g., Boudreau & Hagiu, 2008; Cusumano & Gawer, 2002; Eisenmann et al., 2006; Iansiti & Levien, 2004). Furthermore, although the platform could deliberately try to lock users in the ecosystem, these users can fear potential lock-in.

2-1-5 Data marketplace meta-platforms

Before section 4 will discuss data sovereignty, this section concludes with the discussion on DMMPs. Data markets are a particular instance of digital platforms. The concept of PTPO links several platforms (and thus data marketplaces) together. This can be either by direct inter-operability or by using a meta-platform. This meta-platform in the context of data

marketplaces is the data marketplace meta-platform (DMMP) (Abbas, 2021). The DMMP offers data consumers a centralised portal for data acquisition. Additionally, just as with regular data marketplaces, the DMMP can offer additional services by combining data from the several single data marketplaces.

For example, a DMMP could combine data from a telecommunications data marketplace with data from an automotive data marketplace to provide public bodies with information about the coverage of the telecommunications network for the communication between self driving cars. Another example could be that a large food and beverages producer wants to know how the local taste for soft drinks in Italy differs from Poland. Buying data from an Italian and Polish retail data marketplace could provide these insights. The DMMP enables the food and beverage producer to consume data at one platform, instead of having to search for national data marketplaces.

2-2 Governance and control

This section will start with a brief discussion of organisational control theory before focusing on digital platform governance, which includes control mechanisms. This section 2-2 provides the theoretical background of control which will later be used to analyse mechanisms from current practice in chapter 5 and to identify proposed mechanisms for DMMPs in chapter 6.

2-2-1 Control mechanisms as part of organisational governance

Organisational control theory originated from organisational design research and focuses on how organisations distribute and manage work and workers to achieve organisational objectives (Ouchi, 1979). According to this same author, control theory is about ”the mechanisms through which an organization can be managed so that it moves towards its objectives” (Ouchi, 1979, p. 883). Furthermore, Eisenhardt (1985) notes that the most basic form of control is between an agent and a principal (i.e. a worker and its superior). As organisations are often complex, a portfolio of control mechanisms is often used in a single organisation. Furthermore, these control mechanisms can be divided in formal and informal. Formal control mechanisms can be sub-divided in input control, process control and output control. For example, a high-ranking university that is only admitting students with an average grade of a 9.5 is using an input control mechanisms. If this university is using a strict course-program with intensive supervision by its instructors, it uses process control as well. Lastly, by deciding that students can only pass when they successfully complete a final exam, there is output control as well. Another example could be that a restaurant hires all people that apply and has no internal policies for its workers but only pays its workers by the tip that guests pay. In this second example, as the restaurant trusts that the tip-incentive will indirectly make sure that workers will treat guests friendly, the restaurant is relying on output control only. However, the ideal mix of control mechanisms is highly dependent on the organisational

structure, the organisational goals, the type of workers and the type of work that has to be done. Furthermore, the type of control mechanisms that can be used depend on the level of information that an organisation can acquire about the performance of its workers (Ouchi, 1979). According to Eisenhardt (1985, p. 147), as tasks become more complex and diverse, its measurability decreases which requires additional measures such as additional layers of management or budgeting systems.

Furthermore, organisations can adopt informal control mechanisms. In contrast to formal control mechanisms, informal control mechanisms do not rely on formal procedures and measurements, but focus on the internal culture within a group of people. According to Ouchi (1979, pp. 836-837), informal control is based on a "clan"-culture with shared values and beliefs. For example, all players in a professional soccer team want to win the competition, and are likely to train frequently and to eat healthy.

However, traditional organisational control theory translates not automatically to the complex environment of digital platforms. For example, it is often a challenge for platform owners to balance retaining control and encouraging innovation at the same time (Tiwana et al., 2010, p.679). Furthermore, platform owners rely on other participants for certain resources (e.g. development of complementary applications), whereas traditional organisations used to manage much of their resources internally. Governance of digital platforms will be discussed next in section 2-2-2.

2-2-2 Governance in a digital platform context

As already mentioned in the previous section 2-2, a digital platform ecosystem can differ substantially from a traditional single organisation. Research on digital platform governance and control has very frequently investigated the role of the platform owner and its relation with related parties, primarily complementors such as third party application developers (e.g. Eaton et al., 2015; Ghazawneh & Henfridsson, 2013; Goldbach et al., 2018). According to Tiwana et al. (2010, p. 679), a challenge for platform owners is to balance governance mechanisms to reach the right level of control, without destroying innovation on the platform, which they label as the *"Goldilocks Governance Problem"*. Furthermore, Tiwana et al. (2010) argues that the governance design of a digital platform includes three dimensions: 1) decision rights partitioning, 2) control, 3) proprietary versus shared ownership. This project will mainly focus on the second dimension, but results from chapter 5 and chapter 6 will show that the other two dimensions can not be neglected for data sovereignty as well. However, Mukhopadhyay and Bouwman (2019, p. 336) provide critical views on this three-dimensional view on digital platform governance and control: "This definition completely overlooks the important role governance plays in ecosystem-centric value co-creation as well as value appropriation among ecosystem participants." Furthermore, these scholars argue that "platform governance needs to address the paradox of control and autonomy to have the right kind of control and autonomy in the ecosystems" (Mukhopadhyay & Bouwman, 2019, p. 340).

Additionally, a different perspective on digital platform governance exists, the boundary resource-model by Ghazawneh and Henfridsson (2013). This theory considers the platform owner as a potential provider of resources for third party developers to secure the platform provider's goals and standards. More specifically, these scholars identify two drivers of the design of boundary resources: 1) resourcing, to enhance scope and diversity of the platform, 2) securing, to increase control over the platform by the platform owner (Ghazawneh & Henfridsson, 2013, p. 185). For example, Apple can provide software development kits to third party developers to help them develop new iPhone applications, but can also restrict these developers by forcing them to use those resources and accept the accompanying conditions (Ghazawneh & Henfridsson, 2013).

3

Methodology

This chapter explains the research process and methodologies that were used in more detail. It starts with section 3-1 that explains the type of research that was conducted and why an exploratory approach was considered appropriate. Secondly, section 3-2 provides an overview of how the methods that were used related to the sub-questions and main research questions. Additionally, this section links the sub-questions and main research question to the different sections of the report as well. Next, section 3-3 elaborates on the main data collection methods: a review of academic literature, a brief analysis of the current state using data space and data marketplace industry sources (e.g. industry literature, online sources and white papers), and expert interviews. Furthermore, section 3-3-3 describes the research process that was used for the expert interviews including the preparation and pre-testing of the interview protocol. This chapter closes with section 3-4 that describes the process that was used to analyse the expert interviews.

3-1 The choice for an exploratory approach

Although there is academic literature on digital platforms and data sovereignty, including an emerging stream on data marketplaces, the amount of literature on DMMPs is still relatively scarce, especially regarding papers focusing on data sovereignty in a DMMP-context. Additionally, although initiatives from industry and academia that study the integration and standardisation of data marketplaces are emerging, currently DMMPs are not used by organisations on a large scale. As a result, DMMPs can be considered as a new phenomenon. The novelty of DMMPs, and especially of data sovereignty in that DMMP-context, make that this research addresses an exploratory research question (Sekaran & Bougie, 2016, p. 43).

Although DMMPs are still novel with less literature compared to 'regular' digital platforms,

this did not imply that there was no literature that could be used during the research project. Section 3-2 presents the research design that was used to address the sub-questions and main research question eventually. This visual presentation of the research flow links the different data collection methods to the research (sub-)questions, and also includes how both academic and industry literature were used. Both the use of academic and industry literature, and subsequently expert interviews will be further discussed in the next sections of this chapter.

3-2 Research design

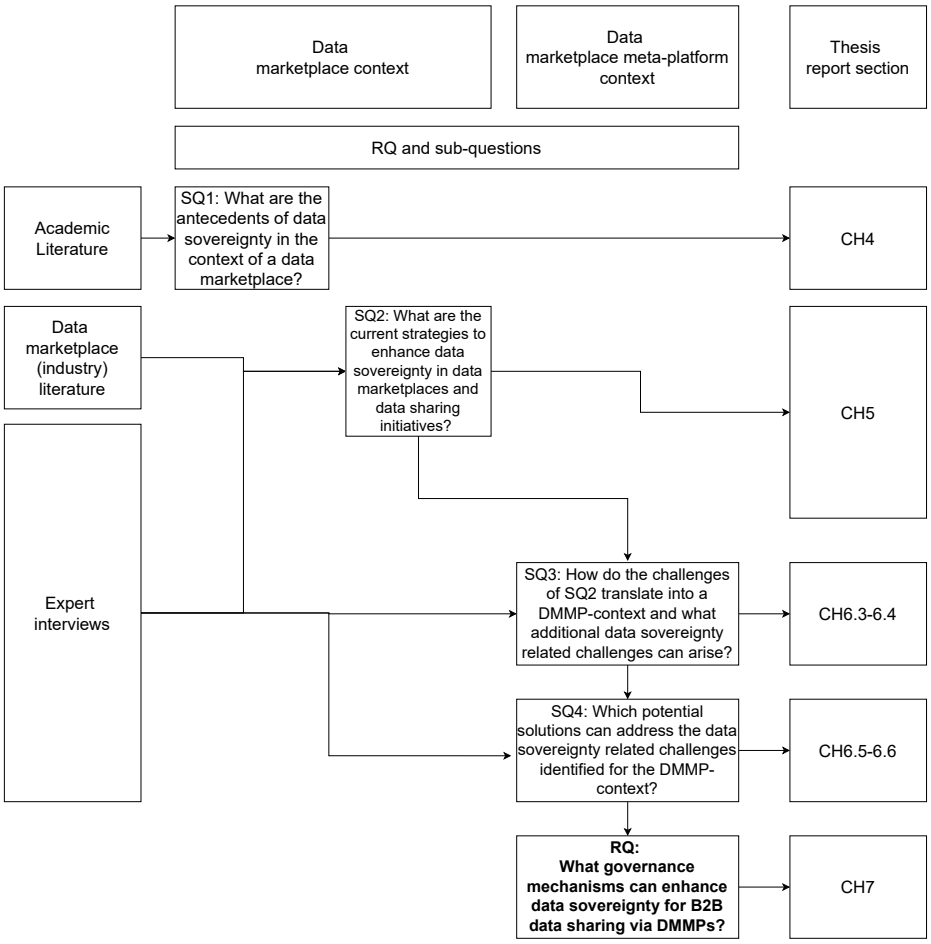


Figure 3-1: Research design flow

3-3 Data collection methods

This section elaborates on the data collection methods that were used for the research project. Section 3-3-1 describes the process to review the academic literature. The findings from this step were used to get an understanding of governance of digital platforms, with data marketplaces in particular (as discussed in chapter 2). Secondly, the academic literature provided an initial understanding of data sovereignty and its antecedents. The output of this literature review was used to answer the first sub-question. Next, an analysis of the current state of data sovereignty in the context of data marketplaces was performed. Industry documents from data spaces and data marketplaces were combined with insights from expert interviews. This research step aimed mainly to better understand how the high-level data sovereignty antecedents translated in practice. This understanding was used as a baseline before researching the context of DMMPs and to help contrasting DMMPs to data marketplaces and data spaces.

Additionally, the expert interviews were used to better understand the DMMP-context vis-a-vis the data marketplace-context, both regarding emerging challenges and possible solutions in relation to data sovereignty for data providers.

3-3-1 Literature review

The goal of the review of academic literature was three-fold. First of all, academic literature was used to conceptualise data marketplaces as an instance of digital platforms. This meant for example to identify the different actors related to a data marketplace such as platform operator, data provider and data consumer. Secondly, by aiming specifically at digital platform governance literature, a first understanding was obtained of how governance of digital platforms differs from traditional organisations and how research has studied governance mechanisms in a digital platform environment and for which purpose those mechanisms are used. Thirdly, the academic literature provided insights regarding the antecedents of data sovereignty.

The Scopus-database was used as main literature database as this database contains mainly high-quality journals and conference papers, often peer-reviewed. Several queries were used to find relevant material, each with a different scope. In general, queries were performed by combining two strings of related concepts. Table 3-1 summarises the different queries.

Scope	String 1	String 2
Organisational governance & control	Organizational, theory	control, governance
Digital platform governance	Digital platform*, digital ecosystem	governance, control, polic*
Data marketplace	Data	marketplace*, space, ecosystem, sharing, exchang*
Meta-platforms	Meta, platform-of	platform*,
Data sovereignty	Data, digital	sovereignty, control, governance

Table 3-1: Literature search queries

The return of the queries were sorted by number of references first, to find influential and highly-cited work. Secondly, the results were sorted by age, where newest publications were ranked highest. This second sorting of the results had the goal to find the newest publications, with potentially the newest academic insights. The results were scanned by title, and in most cases abstract too. Based on title, abstract and author keywords, relevant work was saved in a reference manager. These articles were reviewed in full, also to find frequently cited work that was still missing from the literature set. This backward snowballing-approach led to around 40 additional articles. The use of this snowballing approach was also used to mitigate the risk of missing articles not present in the Scopus database. The full-text of these articles found secondary were saved from either Google Scholar or the TU Delft Library portal.

3-3-2 Analysis of current state

After the initial understanding of data sovereignty in a data marketplace context was obtained from the literature, several data marketplaces from practice were studied. However, several industry initiatives identify themselves as data spaces, which differs from a strict data marketplace-definition in the sense that data spaces focus more on inter-operability and soft agreements instead of data trading. The goal of this step was to determine the current state of data sovereignty in the context of data spaces and data marketplaces. Both websites and industry documents (e.g. white papers) were reviewed to understand how these initiatives and platforms enhance control over data for data providers. The paper by Spiekermann, 2019 provided as a starting point for this analysis, as this work provides a relatively recent overview of data marketplaces.

The current state analysis was supplemented by insights from the expert interviews, as many experts are personally involved in data sharing initiatives and were able to share personal experiences regarding data sovereignty. The process of the expert interviews is discussed next in section 3-3-3.

Participant	Meeting duration (minutes)
T1	60
T2	60
T3	30
T1 (follow-up)	30

Table 3-2: Overview of interview protocol pre-testing meetings

3-3-3 Expert interviews

The expert interviews were conducted in a semi-structured manner to include certain topics, but to leave open the possibility to elaborate on interesting or according to the interviewee essential elements. A total of eleven interviews were conducted with ten interviews using a virtual video-call connection and one interview in a physical setting. As both setups included face-to-face connection, the researcher was able to visually observe respondents in addition to the audio. All interviews were recorded for later analysis, after explicit permission by the interviewee. All interviewees were asked to read and sign an informed consent form in advance of the interview, with the possibility to ask questions before the interview. This form is included in appendix C and included the goals, risks and risk-mitigating steps of the research project. Additionally, it informed participants about their rights of withdrawal and the data storage and processing measures. There were no remarks or questions by participants regarding the informed consent form.

An interview protocol was used a basis for the interviews. This protocol was developed based on the findings from the literature-based research steps. However, before conducting the actual expert interviews, the protocol was developed and refined iteratively. This approach was adopted to minimise the risk of ambiguity in the wording or visuals of the protocol. More specifically, the protocol development included a total of 4 interview protocol pre-testing meetings that were organised to obtain feedback and to grasp initial responses by staff from the internship company. These pre-testing participants were no experts regarding the research context, but all had a background in IT and/or cybersecurity. An overview of these pre-testing meetings is included in table 3-2.

Pre-testing participants were asked to read to protocol and to comment on all things that were unclear, interesting, or could be improved in their opinion. In addition to the protocol text, they were asked to comment on the slides that displayed the visuals of the data marketplace, data marketplace and data sovereignty antecedents. Their feedback was used to improve the protocol, for example to improve the bridging segments between the different topics of the interview. All in all, this resulted in several improvements to the slides with the visuals, especially for the DMMP-slide. A follow-up meeting with one of the pre-testing participants was held to validate the improvements to the protocol and slides. During these pre-testing meetings, both the participants and the researcher used a paper copy of the protocol and visuals to make instant notes of the feedback.

The questions of the protocol were deliberately kept open, to offer expert interviewees the possibility to express their opinions and personal experiences. Furthermore, as it was important to make sure that interviewee and researcher had a common understanding of data marketplaces and DMMPs, the visuals on the slides presented a relatively specific overview of these concepts, to improve mutual conceptual understanding. This comes with a risk as well, as it can steer the interviewee. To mitigate this risk, the interview protocol was balanced (e.g. asking about advantages, but disadvantages as well). Furthermore, questions were included were interviewees were asked explicitly if they missed items on the slides. The visual and questions related to the data sovereignty antecedents were very open, partly because this approach allowed for rich and deep understanding of the experts' viewpoints. The final interview protocol is included in appendix D and the final version of the slides with the visuals are included in appendix E. A total of eight iterations was performed after no further comments were given by the pre-testing participants.

Sampling method

Researching data sovereignty in the context of DMMPs requires relatively specific knowledge from interviewees. This knowledge is not widely available and had to be acquired from experts with specific experience. As a result, purposive sampling was used which is used to obtain information from specific target groups (Sekaran & Bougie, 2016, p. 248). Based on the literature review, a profile was sketched to identify expertise-areas for potential interviewees. As the group of experts with experience regarding data marketplaces is relatively little, the scope was broadened to the field of business-to-business data sharing, digital platforms, data sharing barriers and data sovereignty. These fields of expertise were also included in the thesis project one-pager, which will be discussed later in this section. Using these selection criteria to find people within a limited category of appropriate experts is a form of judgement sampling (Sekaran & Bougie, 2016, p. 248).

Experts were approached using four channels:

- Direct e-mail invite
- The network of one of the university supervisors
- The internship company
- The 2022 Hannover Messe fair

By researching current data sharing and data marketplace initiatives, several potential interviewees were identified and approached by emailing the interview invitation in appendix A. The thesis project one-pager in appendix B was attached to provide additional background of the research project. The professional network of one of the thesis project supervisors also resulted in potential leads to interviewees. Thirdly, people within the internship company were contacted to get in contact with potential interviewees. Lastly, the 2022 Hannover Messe fair in Germany provided an interesting opportunity to get in direct contact with experts,

as several European data sharing initiatives were present with a stand. These people were approached during informal talks with follow-ups via email including the one-pager.

The interviewee selection process resulted in a total of eleven interviews as presented in table 3-3.

Expert identifier	Professional background/ experience	Interview duration (minutes, excl opening)	Location	Dutch/English	Virtual/ physical
Expert 1 (E1)	Data sharing and digital identity consultant	77	Netherlands	Dutch	Physical
Expert 2 (E2)	Date e-commerce project manager, experienced professional in the telecommunications and financial industry	70	North-America	English	Virtual
Expert 3 (E3)	IT Architect/software developer, data sharing expert	51	Netherlands	Dutch	Virtual
Expert 4 (E4)	Experienced IT and project professional	62	Europe	English	Virtual
Expert 5 (E5)	Experienced professional in financial services and management consulting, currently leading the data practice of a North-American data e-commerce company	57	North-America	English	Virtual
Expert 6 (E6)	Senior research specialised in trusted data sharing and business ecosystem architecture at a Dutch research institution	56	Netherlands	Dutch	Virtual
Expert 7 (E7)	Director of pan-European trust and data sovereignty framework, combined with broader experience leading IT- and technology-driven companies.	53	Netherlands	Dutch	Virtual
Expert 8 (E8)	Board member of regional collaborative organisation, specialised in future affairs including digital and data-related topics	47	Netherlands	Dutch	Virtual
Expert 9 (E9)	Data management expert at a global professional services firm	52	Europe	English	Virtual
Expert 10 (E10)	Data expert and research engineer at a German research institution	37	Europe	English	Virtual
Expert 11 (E11)	Developer and semantic web expert, data sharing initiatives expert, data engineer	47	Europe	English	Virtual

Table 3-3: Overview of expert interview participants

3-4 Data analysis

The analysis of the individual interviews was started as soon as possible after finishing each interview. This means that transcription was mostly started directly after the interview, to make make sure that the memory of the conversation was still vivid. During the interview, a paper copy of the interview protocol was used to make instant notes during the conversation as well. Each interview was transcribed using the interview recording. Interviews conducted in English were transcribed automatically using transcription software. However, the automatic transcripts were manually compared to the audio from the video recording and corrected where necessary. The Dutch interviews were transcribed directly by re-playing the recording. Subsequently, the transcripts originating from the recordings in Dutch were translated to English using translation software with a manual review of the translation afterwards.

All interview transcripts were coded using Atlas.ti software. In the first step of the coding process, open-coding was applied to find the concepts and challenges that were mentioned by the interviewees. In the second coding-step, codes were grouped and re-grouped, and codes for certain quotations were changed to fit in the new code structure. During this

process, earlier transcripts were re-reviewed multiple times to apply changes in the coding structure. Recurring concepts and viewpoints were linked and conflicting viewpoints between participants were noted. Although the open coding-phase resulted in several new codes, the findings from the literature provided the first starting structure for the codes. Codes were for example linked to data marketplaces, DMMPs, specific data sovereignty antecedents or elements of digital platform governance.

Lastly, to protect the identity of all participants and to adhere to TU Delft standards regarding research involving humans by the Human Research Ethics Committee, all transcripts were anonimised. Names of respondents were replaced with pseudonyms (EX for Expert X) and specific job descriptions were generalised. Furthermore, as some interviewees mentioned very specific personal experience with particular projects, these quotes were either generalised or removed entirely. Given the network of experts in this field, and the mutual connections that these individuals have, this was considered necessary to avoid identity recovery by triangulation. Additionally, all anonimised transcripts were summarised in a few pages to be made available to the public, as described in appendix F. Both the anonimised full transcript and anonimised summary were send back to the participants. Participants were asked to reply in case they feel that there is deviation between their memory of the conversation and the written transcripts and summaries. This step was included to make sure that interpretation of the interview during transcription would not have led to a different overall conclusion of the conversation with the expert. No deviations or complaints were returned by the experts from the interviews.

4

Data sovereignty

This chapter aims to answer the first sub-question: *"What are the antecedents of data sovereignty in the context of a data marketplace?"* Data sovereignty is about staying in control as data provider, but to achieve this, data providers need control over several antecedents. This section identifies data sovereignty antecedents in the context of data marketplaces by analysing existing literature. Additionally, potential challenges for a data marketplace context are provided.

Firstly, each antecedent will briefly be introduced by a description. Subsequently, potential challenges that might arise for these topics in the complex multi-actor environment of data marketplaces are discussed. To guide this section on data sovereignty, the different facets of data sovereignty are discussed from the perspective of the original data source (i.e. data provider). As this research is on business-to-business data sharing, the original data source is the organisation that has gathered the raw data. An example could for example be the company that owns a factory where assembly line sensors measure failure rates of new products. Another example could be an insurance company that gathers data on the financial impact of natural hazards. Based on the literature, four antecedents were identified: data ownership, data access, data usage (& processing) and data storage. These antecedents will be discussed accordingly.

4-1 Data ownership

The term "data ownership" could appear to be simple, just as it is for example when defining "car ownership". However, data ownership is not just about property, but is a concept that entails responsibilities and expectations of data owners as well (Hummel et al., 2021). Being the owner in this case means which party has generated the data, has control over the data,

can grant and withdraw rights to others over the data (Asswad & Gómez, 2021) and is able to decide on the storage of the data (Peterson et al., 2011). Before several challenges for data ownership in relation to data sovereignty can be discussed, it is important to deal with the unique characteristics of data first. In contrast to "normal" goods, data are intangible, non-exclusive (data can be duplicated almost endlessly), do not deteriorate with use but can become less relevant as they age (Hart, 2002, p. 25-26). Due to the intangibility of data, a legal basis for data ownership can also be problematic (Lauf et al., 2022). Furthermore, data can be both the final product, an input for a product or service or an intermediary (Aaronson, 2021, p. 7). These characteristics of data lead to challenges when it is traded using a digital marketplace. To address these challenges, governance mechanisms have to deal with the issue of data ownership.

4-1-1 Data ownership challenges for data marketplaces

One potential issue is the way how ownership is treated after the sale of data. For 'normal' goods, a sale means that ownership is transferred from seller to buyer. Due to the unique characteristics of data, this might not be that straight-forward. The intangibility can lead to difficulties for data consumers to assess the quality of the data. The non-exclusivity of data can lead to numerous copies of a dataset with unclear ownership of these copies. Lastly, as the value and relevance of the data depends heavily on age, ownership might become less important as the dataset ages.

Additionally, data ownership solely might not be able to solve the challenges of the data economy (Koutroumpis et al., 2020). For example, what if an organisation decides to sell parts of their data, are they still the owners? Furthermore, what if the organisation owning the data no longer wants to take care of it. Who is responsible in that case? Lastly, what if the party that has sold the data wants to revert the sale in the future? This might sounds hypothetical, but an issue with sharing data is that an organisation can not know what becomes possible in the future.

4-1-2 Data ownership implications for data marketplaces

From a data marketplace-perspective, these challenges offer directions for potential solutions. Firstly, for data ownership, it has to be clear who is the current owner of the data. This is important to determine which stakeholder is currently responsible for curating the data. Secondly, for a particular dataset, the original owner has to known. This might be important if the platform wants to offer the original data source the option to withdraw data from the market. This could for example be done by making a distinction between the original data owner (i.e. data creator) and current data owner. Thirdly, a data marketplace might want to distinguish between who owns the data and who can access the data. The data owner could be granting these access rights to other parties (Asswad & Gómez, 2021). Data access as a

factor for data sovereignty is discussed in section 4-2.

4-2 Data access

The second antecedent of data sovereignty is data access. This includes how and when other parties can access the data of the data source. It is also about how access is arranged after a data transaction has taken place via the data marketplace. This granting of data access rights to other stakeholders could be in control of the data owner (Asswad & Gómez, 2021). Secondly, access to data can be important for data trading, as data consumers need to review the data for quality assessment. Furthermore, scholars separate access to data and metadata in the context of data trading (e.g., Koutroumpis et al., 2017; Lawrenz et al., 2019). Whereas the data are the core product, and form the basis of additional data-based services as well, metadata are data about the data. These metadata contribute to to the provenance of the data as it provides information about the origin and quality of the data (Koutroumpis et al., 2017, p. 12).

4-2-1 Data access challenges for data marketplaces

For data marketplaces, dealing with access to data can lead to several challenges. Firstly, data consumers could want to access the raw data for quality assessment. However, full disclosure of the data for quality assessment makes the data provider give away the value of the data. Furthermore, Spiekermann (2019, p. 209) argues that this inability of data consumers to assess the full data set can lead to data consumers that are not able to recognise the full potential value of the data. Secondly, there can be challenges from the perspective of the original data source. Is the original data source still able to access the data after the transaction? Or should part of the transaction address the access rights of the original data source? This could mean that if both parties prefer to do so, it could be possible for data providers to sell not only the access rights, but also the ability to grant access rights. Furthermore, especially for highly-sensitive data (e.g. health data), care has to be given to who is able to consume and access these data-sets. How can unwanted access to data be prevented? How should access to personal data (i.e. data that can be traced back to human beings) be handled differently than less sensitive data? Lauf et al. (2022), for example, propose data access with a time constraint. This can mitigate the risk of a data provider giving access now, without being able to oversee what future technologies might be able to do with it (Lauf et al., 2022, p. 9).

4-2-2 Data access implications for data marketplaces

The data access antecedent has several implications for data marketplaces. Firstly, data access has to be governed both before and after a data transaction that happens on a data

marketplace. In the former scenario, data access can be necessary for data consumers to become informed about the potential data goods to be acquired. However, in the latter scenario data access has to be governed in a way that ensures the interests of the original data provider. Metadata is an important complement to the raw data. It can help to improve data provenance and the governance of the data marketplace needs to address metadata explicitly. It could for example be useful to separate the data marketplace in a platform for metadata and the data itself. Koutroumpis et al. (2017, p. 28) propose a decentralised marketplace design where a separate layer is used for data tagging by the data providers and the validation and lineage of the data which is visualised in Figure 4-1.

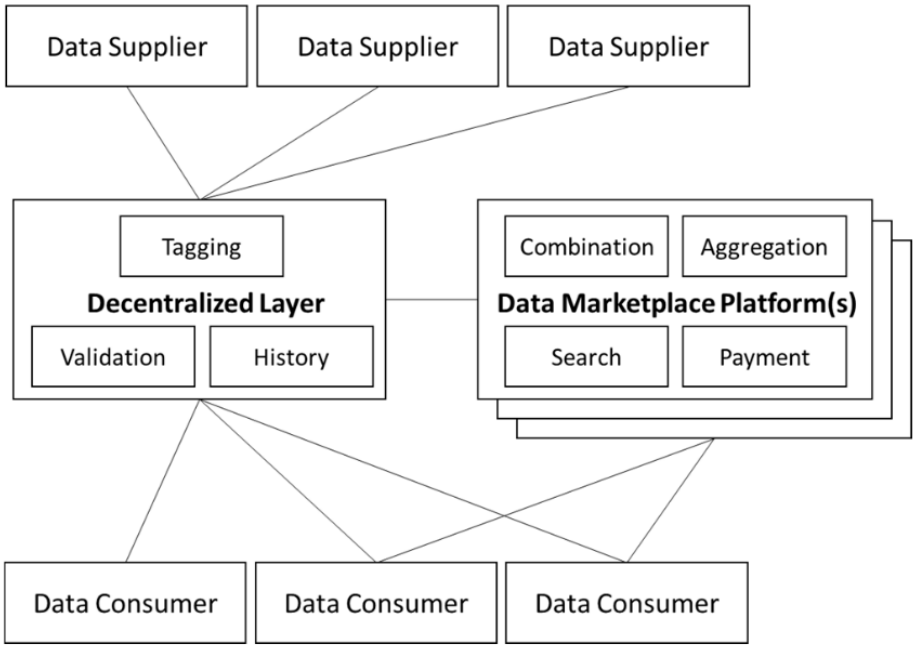


Figure 4-1: Decentralized data marketplaces, adopted from Koutroumpis et al. (2017, p. 28)

4-3 Data usage and processing

Thirdly, data usage and processing determines if data sovereignty is achieved or not as well. One of the goals of data marketplaces is the transfer of business data in order to create value for the different actors in the network. This can be achieved by trading raw data, or by trading processed data using "data processing and analytics tools" as well (van de Ven et al., 2021, p. 321). After a processing step, raw data can be turned into something useful, e.g. a business intelligence report, management information dashboards and the like. In summary, data marketplaces can add value by aggregating different data sources to provide novel (business) insights (Koutroumpis et al., 2017, p. 19). An example could be that car insurance data from France and Germany are combined in a report that provides insights on

the driving behaviour of the owners of particular car brands.

4-3-1 Data usage challenges for data marketplaces

This combining and aggregating of raw data in processed data products creates value, but comes with difficulties from a data sovereignty-perspective. Firstly, the way how data is processed determines partially to what extend the original data source can be traced in the final product or service (Moreau et al., 2008). Secondly, combining different data sources in finalised products could lead to contracting issues because each data provider offers data using different terms (Koutroumpis et al., 2017).

For example, if a data marketplace wants to combines data-sets from Dutch and Spanish energy companies to offer business intelligence products for a British solar energy company, both of these data providers have to allow their data to be used by a data consumer outside the EU. This example highlights the challenges that can arise when combining separate data sources. Data sovereignty of data providers is only ensured when they are able to share data by their own preferences, or have at least a high level of control over it. These conditions should not only be maintained while sharing raw data, but also when this data is processed in data products.

The processing of data can also be problematic when data has to be withdrawn after transactions have already taken place. An example could be that a set of raw data turns out to be erroneous or that a data providers decides to withdraw data for business strategy reasons. Withdrawing data ex-post is already challenging, but can be even more difficult if data has find their way in processed data products.

4-3-2 Data usage implications for data marketplaces

These challenges highlight that it can be difficult to maintain or at least enhance data sovereignty and at the same time use the full potential of data marketplaces by combining and processing different sources of data. To start with, a data marketplace needs to install governance mechanisms that enable the use of diverse data sources for aggregation in data products. These measures are mainly installed before a transaction happens, to make sure that risk of future conflict between the data providers is minimised.

Secondly, there has to be a mechanism that ensures that the original data sources of finished data products can be traced back. This is mainly to make sure that data providers can withdraw data ex-post. This does not only serve the interest of data providers, but of data consumers as well. Although withdrawing data ex-post is not beneficial for data consumers necessarily, withdrawing data that turns out to be erroneous is of interest for data consumers as well. Someone can see parallels between this scenario and for example recalls by car manufacturers. Although it can be annoying for car owners to bring their car in for a service

due to a recall, it is in the end at their best interest to have a car that is fixed and safe.

4-4 Data storage

The way how data is stored has an impact on the level of data sovereignty as well. For example, de-centralised data storage that keeps data at the source might be favourable for the data providers, but might be problematic when large data volumes have to be traded over the data marketplace. Secondly, central data storage could enable quicker transactions. The intensive use of cloud services can make data localisation challenging as well (Peterson et al., 2011). Thirdly, there could be a distinction between the storage of the raw data and the metadata. Meta-data could be used to provide data consumers with information in order to assist the data acquisition process in data markets (Lawrenz et al., 2019). The value of metadata for provenance of data goods was already discussed in section 4-2. The issue of data storage is not only about the localisation of the data, but also about responsibilities of protecting the data. In a centralised scenario for example, the cyber security responsibility could shift from data providers to the marketplace. Storage costs of large data volumes is a major factor that impacts big data governance decisions in a single organisation (Tallon, 2013). These storage costs apply in a data marketplace-context as well.

4-4-1 Data storage challenges for data marketplaces

For the stakeholders of a data marketplace an issue regarding data storage is that unwanted copies of the data can exist. This can be both harmful for data providers: the data consumer could have created copies for purposes that were originally not agreed on. Conversely, data consumers might pay data providers for unique access, whereas in reality other parties use illegal copies of the data. Although data sovereignty can not guarantee that there are no additional copies of the data (Peterson et al., 2011, p. 2), a data marketplace could use a system of certification for data suppliers and consumers to enhance trust. Additionally, a data provider might prefer to keep track of the location of the data after it is sold on the data marketplace. However, a data consumer might prefer to not disclose the location of the data that was acquired.

4-4-2 Data storage implications for data marketplaces

It has become clear that the choice for the storage of data brings both opportunities and responsibilities. The storage configuration can also impact the willingness for data providers and consumers to participate in the data trading. For the data marketplace, it has to be clear which party is responsible for the storage of data, both before and after a data transaction is made. The responsibilities of the platform have to be clear as well. These responsibilities include the protection of data, the retrieval of data and the disposal of data. Additionally,

governance is necessary to enhance trust between data providers and consumers. For example, data consumers that are informed before the transaction about how they have to handle data that they are about to acquire, could be more willing to adhere to these directions. At the same time, it can help data providers to supply more and better data because they know that they can set data storage preferences before a transaction takes place.

4-5 Chapter conclusion

As discussed in the introduction of this chapter, the goal of this chapter was to answer sub-question 1: *"What are the antecedents of data sovereignty in the context of a data marketplace?"* Based on the literature, **data ownership**, **data access**, **data usage** and **data storage** were identified as antecedents of data sovereignty. As a result, these four antecedents will also form the basis for that particular section of the expert interview, using figure 4-2

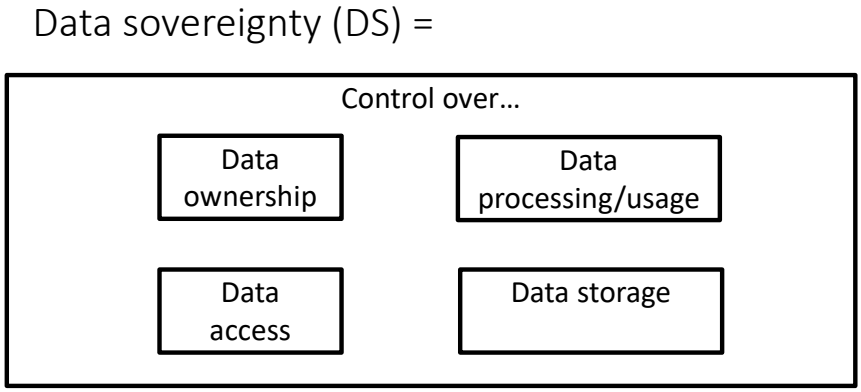


Figure 4-2: Data sovereignty antecedents in a DM-context

5

Data sovereignty and sharing: current state and expert perspectives

This chapter addresses the second sub-question: *"What are the current strategies to enhance data sovereignty in data marketplaces and data sharing initiatives?"* Firstly, an overview will be given that describes current data sharing initiatives and the data sovereignty related governance mechanisms that these initiatives have adopted or are currently developing. These insights are partly based on sources such as white papers and reports. Additionally, experts were also asked about their experience with these data sharing initiatives and their view on data sovereignty in that context. By sketching the current situation, a baseline is developed to contrast DMMPs to in chapter 6. Section 5-1 provides an overview of several data sharing-related initiatives. Secondly, section 5-2 discusses the data sovereignty-related strategies that were identified in these initiatives.

5-1 Overview of current data sharing landscape

Currently, the data sharing landscape is still fragmented with several initiatives in different stages of development. Some initiatives have started recently, whereas others are currently being adopted by the first users. As a result, several different terms are used for these initiatives. Whereas several projects are developing a data space, others are focusing on developing data marketplaces. Additionally, the term data ecosystem is also used frequently. Although these initiatives differ in scope and use different terminology, they are all focused on improving (cross-organisational) data sharing. Some data sharing initiatives started around a particular application, for example the Dutch AI Coalition (NL AI Coalitie), focused to enable data sharing as this is an important boundary condition for training Artificial Intelligence (AI) algorithms (Nederlandse AI Coalitie, 2020, p. 7). Other initiatives focus on solving a specific

problem in a specific sector, for example Smart Connected Supplier Network which works towards inter-operability of Enterprise Resource Planning (ERP) systems to achieve more efficient data sharing between primarily Dutch manufacturing companies by developing shared data standards[1].

5-1-1 International Data Spaces

Several initiatives are operating at a European level. One of these is the International Data Spaces (IDS)-project operated by the International Data Spaces Association (IDSA), an association consisting of 20 companies and research institutions. IDS works towards secure and sovereign data sharing between participants by creating data spaces. Data spaces are sets of 'soft' agreements to realise data sovereignty for data providers and secure data exchange for all participants[2]. One of the key components of IDS is certification of participants and their components that are used for data exchange (Steinbuss et al., 2019). According to the IDSA, certifying a participant "demonstrates a level of security regarding availability, confidentiality and integrity to all other participants and stakeholders" (Steinbuss et al., 2019, p. 4). Certification is developed with several assurance levels depending on the profile of the participant. Furthermore, the lower levels of the certification ensure that a lower barrier of entry still enables Small and Medium-sized Enterprises (SMEs) to participate (Steinbuss et al., 2019, p. 5).

5-1-2 GAIA-X

Another large European initiative is focused at developing the infrastructure for sovereign data sharing ecosystems: GAIA-X (Asswad & Gómez, 2021). GAIA-X is developing a federated data ecosystem, which means that there will not be one large cloud, but that existing clouds will become inter-operable. The goal of a federated ecosystem is also reflected in the organisational structure of GAIA-S: a central GAIA-X association, with national GAIA-X hub and lastly, the GAIA-X Community[3]. Similar to IDS, data sovereignty is one of the primary goals. According to Lass and Bender (2021, p. 338), this translates into "complete control over stored and processed data and also the independent decision on who is permitted to have access to it." Eventually, the IDS reference architecture, including the "blueprint" of the actual data exchange, will become part of GAIA-X[4].

[1] https://smart-connected.nl/en/about-scsn/whic-problem-do-we-solve
[2] https://internationaldataspaces.org/why/data-spaces/
[3] https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html
[4] https://internationaldataspaces.org/we/gaia-x/

5-1-3 Data Sharing Coalition

Furthermore, another initiative is currently developing specific use cases to make the value of data sharing visible and to help organisations translating this to their specific needs. The Dutch, although operating internationally, Data Sharing Coalition (DSC) is currently using this use case-driven approach with a large number of participating organisations. One of the goals of this initiative is to help interested organisations to develop data sharing use cases. One of the interesting aspects of the DSC is that it is focusing on cross-sector data sharing[5]. One of the solutions that the DSC is exploring is the *proxy model*: "systems which are to be used by every Domain with the function of translating between Domain specific specifications and common, Harmonised inter-Domain specifications" (Data Sharing Coalition, 2021, p. 36).

5-1-4 iSHARE Trust Framework

The iSHARE Trust Framework is developed by the iSHARE Foundation in close collaboration with users, originally in the logistics sector[6]. This trust framework enhances inter-operability among data spaces in different industry sectors by offering several building blocks, both technical and legal. The iSHARE framework includes an authorisation registry as well, where participants are registered with their authorisation policies and where they can receive data access requests[7]. Additionally, the organisational structure of iSHARE reflects its intended federated use: members of the individual data spaces that are connected to iSAHRE get a seat in the iSHARE Foundation to co-create new releases and contribute to the decision-making.

5-1-5 Advaneo

Advaneo is an organisation that offers a solution to help parties setting up their own data space or even data marketplace, using four building blocks to exchange with external parties: the Data Catalog (to index data assets), the Data Marketplace (to broker data using metadata from the Data Catalog), the Trusted Data Hub (to perform the actual exchange in a secure and collaborative manner) and the Connector to connect the three other blocks and to connect to the external parties [8]. Furthermore, Advaneo has developed its solution using the existing reference architecture of GAIA-X and the standards developed by IDS.

[5]https://datasharingcoalition.eu/about-the-data-sharing-coalition/
[6]https://ishare.eu/ishare/the-foundation/overview/
[7]https://ishare.eu/ishare/authorization-registry/
[8]https://www.advaneo.de/en/data-spaces-solution-en/

5-1-6 TRUSTS

TRUSTS is a pan-European project that is ran by researchers from academia in collaboration with industry players. One of the main objectives of this project is to develop a platform that can both be used independently, but more importantly, also be used to integrate separate data platforms while maintaining security, trustworthiness and General Data Protection Regulation (GDPR)-compliance[9]. This project will also use existing standards and data sharing initiatives, such as IDS and GAIA-X (Simon & Natalia, 2021, p. 13).

5-1-7 Expert views on the data sharing landscape

The experts, having both direct or indirect experience with data sharing initiatives, were also asked about their view on data marketplaces. They expressed several advantages ranging from improved data discovery, harmonised supply of data, cost reduction, enabling new business models and as useful part of a larger data space. However, several experts also expressed their scepticism regarding data marketplaces. More specifically, several experts feared that a data marketplace could lead to vendor lock-in, would be difficult due to privacy concerns and would be difficult to implement due to technical complexity. Additionally, several experts also sketched the current perspective of industry players regarding sharing of data that can be interpreted as risk-averse and that there is currently not a clear business case. According to some experts, the lack of a business case is the reason why there is a mismatch between market sides (data providers versus data consumers) which makes it difficult for data marketplaces to reach sufficient scale. For the remainder of this section, please consider the discussed data marketplace-configuration of figure 5-1.

5-2 Strategies to enhance data sovereignty

Experts also mentioned the data sovereignty concerns they had. Furthermore, these experts provided insights in the current practice of data sharing and which strategies are being adopted to enhance data sovereignty. Firstly, the expert insights regarding data access and data usage will be discussed in section 5-2-2. Secondly, section 5-2-1 will further complement section 5-2-2 by elaborating on how agreements between data providers and data consumers are currently being formalised. Thirdly, the experts emphasised on the role of trust in the current state of data sharing which will be discussed in section 5-2-3. Lastly, during the discussions about data sovereignty in the current data sharing landscape, experts often mentioned that currently several data sharing initiatives are exploring the balancing between specific and generic agreements. Considering the DMMP-context that will be discussed in chapter 6 which could potentially further drive cross-sector data sharing, the current perspectives on this topic are included as well.

[9] https://www.trusts-data.eu/

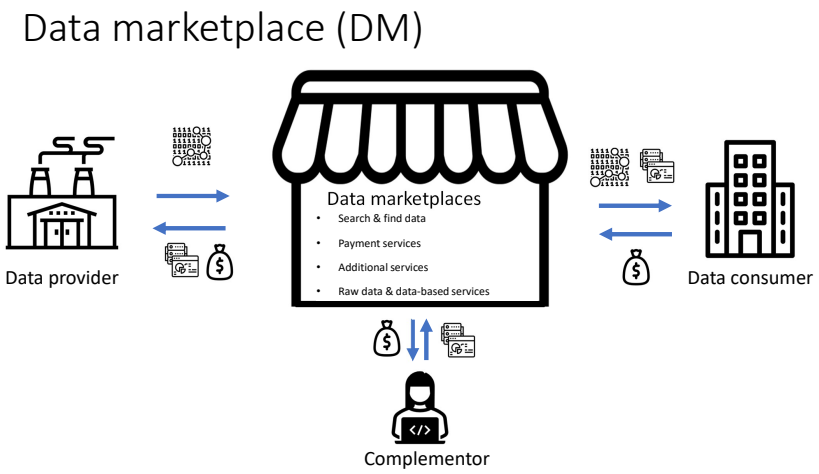


Figure 5-1: B2B data marketplace

5-2-1 Operating on a common legal basis

The experts that described the data sharing initiatives which they were involved in often mentioned that the first step was to create a common legal basis on which all parties can operate. This means that all parties involved have committed themselves to a shared set of rules and agreements. This legal context provides the first layer of protection for data providers (but for data consumers and complementors as well). Due to the different types of data that are exchanged in these initiatives and the different types of use cases, there is a set of shared agreements across all parties, and specific agreements for each particular exchange as explained by E7:

> *"The basis is that all parties operate on the same legal basis, and on top of that you have a specific agreement for each data exchange. And that in connection with the specific purpose, I request access, and for this purpose you then get access with this license. With this you connect purpose and license and then you have arranged that transaction."*
>
> Expert 7

Please note that E7 is also mentioning the 'license'. This partly addresses the data ownership

challenge of data sharing. By agreeing on a license, the owner of the data that provides it to the consumer does not give up ownership, but is only granting access under certain conditions which are registered in the license. During this conversation with E7, this expert also noted that the term 'data provider' is not always the party that can be considered the owner. According to E7, due to the legal context, data owners are described as 'entitled parties', which could be translated into rights holder to the data. The party that in the end actually delivers the data to the data sharing platform could be an external cloud provider as well:

> *"If you look at where the data is and what the primary trigger is to carry this further. At the moment, we have mountains of data in the cloud, the cloud suppliers have the data. But the data belongs to their customers, or to the data entitled party. And it is mainly about the data sovereignty of the entitled parties, whereby they control which [of their] data is on the platform."*
>
> Expert 7

However, several experts also indicated that as the scale of data sharing platforms increases, a purely legal approach will become more difficult and that in the end it could lead to risk versus reward considerations for data providers when data consumers are in breach and have to determine whether to take legal action. E5 described this trade-off:

> *"And again, it's all scale too. I mean, if it's something minor, you may just say: it's not worth our trouble to chase, you know? But if it's something big and it really is undermining your business, then you should have the ability and the right to go after it fairly aggressively."*
>
> Expert 5

However, this could mean that smaller data providers with limited resources regarding time, money and legal assistance could be vulnerable for misbehaviour on the platform. There could also be a role for the data marketplace-operator (or more general: the organisation behind a data sharing initiative) to help in case of disputes, specifically to assist data providers to stay in control. This was illustrated by E8:

"But to be able to record that properly, to be able to have a notary role, infrastructure is missing there, there is actually missing an organisation that can insure that, you could call it a kind of trust provider. Which ensures that agreements between parties in a consortium are also complied with and can be enforced."

Expert 8

This section discussed the legal basis that is currently one of the important components to make sure that agreements between parties can be made, and can legally be enforced. However, experts also provided insights about the process of usage and access control of data, which is discussed next.

5-2-2 Granting access and controlling usage

As discussed in section 4, data access and data usage are antecedents of data sovereignty. The interviews with experts confirmed that these two antecedents are highly related in current practice. More specifically, data access is often granted for a particular purpose to a specific consumer and the process in which data providers grant access often includes the approval for data usage as well. Furthermore, some experts also explained how this process is currently designed in specific data sharing initiatives. Although there are many configurations to arrange the granting of data access and data usage, several initiatives have developed a methodology where the data consumer makes the initial request to the data provider. For example:

"It often starts with a question. [...] Can I have this data? Then the specific notification to the organisation in question and you get a request for a specific purpose and dataset. For example, if you want to give access to these three specific fields, you want to give them access for the next two years."

Expert 7

Please not that this expert also illustrates that several data access related parameters are proposed by the data consumers, in this example access to specific fields in the dataset and for a period of two years. Although specifying access to the actual parts of the dataset and the time duration are from a high-level already related to data sovereignty of the data provider, there could be much more parameters to include in this request, for example agreements about the persons of the consumer's organisation that get access specifically or for example security-related measures that the data consumer will implement to protect the data that is provided. Furthermore, one finding from practice related to data access is that data consumers often need access to the data not only for their own organisation, but also for third parties they are collaboration with. E7 described this situation, which he explained are delegation levels:

"And one of those elements could be, naturally, and think for example logistics, often the most complicated use case, I have a transport company and I get data access from you. I'm hiring a third party in my turn because I am out of trucks. Then I have to be able to delegate it for a while."

Expert 7

This element further complicates the relation between data provider and data consumer, as the data consumer is not a single entity in practice. Whereas the data provider can already raise questions regarding whether the data consumer is trustworthy, the inclusion of third parties on the consumer's side can further amplify these.

5-2-3 Shift to trust-based data sharing

Although a common legal basis and an effective process to grant access to data for a specific usage purpose can ensure a first level of data sovereignty for data providers, the experts also emphasised on the increasing role of trust. Of course, a network of parties that have all committed to the shared legal agreements and use the installed process to request access for data usage can enhance the level of trust already. However, some experts explained that currently, several data sharing initiatives are build around a shared problem or ambition in a particular sector. This means that several organisations have an interest to share data, namely to help solve a problem or to work towards a future ambition. When asked about his experience with current data sharing initiatives, E1 illustrated this with an initiative by a consortium of organisations in a specific sector that was already working together before the idea for a data sharing imitative emerged:

"Yes, then you can think of the construction sector, for example. That is a clear example of a shared ambition within the sector. An example of something we often run into is that parties would have preferred to share more data in hindsight."

Expert 1

Illustrated by this example is that the current situation where no or less data is shared has shown the parties involved that data sharing can improve business operations in the sector. As a result, the mutual ambition to share data in the consortium of existing partners arose. Similarly, E1 also gave an example that an ambition to solve a particular problem can arise in a single organisation, which will then collaborate with existing partners to improve data sharing. In this situation, the focal organisation can be considered as data marketplace-operator and motivates existing partners that can provide data. The focal organisation then redistributes the data among other partners to actually use it, as also explained by E1:

"Yes, or you could say that the data consumer is setting up a data marketplace for his own business. [...] I think in this case they have data providers that collect data about the building. But the data consumers are other companies that also need data. The data provider can, for example, be the lamp supplier, who supplies data about the use of the space. And the cleaning company, the data consumer who actually uses the data, then knows whether the space needs to be cleaned."

Expert 1

This situation where individual companies are setting up data marketplaces or at least data sharing platforms, was also described by E2. What this expert further added was that highly tailored data marketplaces enhance data sovereignty for data providers because they build up their own platform (with use of external sources, e.g. tools) according to their own terms, conditions and preferences. As a result, the control of the data provider is anchored in the platform from the beginning. For example:

"Essentially, the data provider, if he wants to control and he is, as said under his terms, well, then define the terms of the platform. And that's what the company I work for offers. We offer the tools, and the provider defines the terms of the usage. So, we can manufacture the data as you want. You can package it as you want. Price it as you want. Sell it under the contracts, the type of contracts that he defined, billed for it that he wants, distributed as he wants. So, it's data from him to anywhere under his conditions."

Expert 2

5-2-4 Generalise if possible, specify when necessary

During the expert interviews, several experts described the current process in practice where several initiatives try to balance between generic and specific agreements and standards. For example, standardised identity is currently being developed to make sure that data providers and consumers can trust that they are interacting with a valid party. A similar process is happening regarding payments, although several data sharing initiatives are currently not including an actual financial exchange for data, but are trying to solve business problems, as discussed in section 5-2-3. This means, data providers are not rewarded directly by a financial fee by data providers, but can for example benefit from improved operations, more efficient energy usage or other rewards that could indirectly lead to cost reductions or achievement of social goals such as employee satisfaction. Several experts also noted that the process of making generic agreements is currently mainly dealt with at the level of an industry sector, for example E1:

> *"It is really at the highest level of abstraction in very abstract agreements. The more you go to specific things, the more specific you can organise it. At sector level, you can agree very generic things that we have not yet solved, and for very specific use cases you can make very detailed agreements. For one sub-sector, the agreements of the other sub-sector are not relevant."*
>
> Expert 1

Furthermore, experts also warned that over-standardisation will both be difficult and that it is doubtful whether this will actually improve the data sovereignty of data providers. However, they argued that generic agreements at a high-level for trust-related topics is what practitioners are currently focusing on. E7 illustrated this for example:

> *"It is good to realise that [our initiative] is not going to make all kinds of agreements. Data spaces decide for themselves: we will exchange data around this standard. And we're going to use [our initiative] as a baseline discovery."*
>
> Expert 7

This also relates back to the same legal basis that is important for parties to operate on within one data sharing network. Currently, fragmentation between different data sharing initiatives and participants is reduced by balancing between these generic and specific agreements.

5-3 Chapter conclusion

The aim of this chapter was to answer sub-question 2: *"What are the current strategies to enhance data sovereignty in data marketplaces and data sharing initiatives?"* After the brief analysis of current data sharing initiatives and how they are related combined with the expert perspectives on the current state, several strategies to enhance data sovereignty can be derived. Firstly, most data sharing initiatives have created a set of agreements that all parties have to agree and commit to, before entering the network. This creates not only a first layer of security and potentially trust, it also makes sure that there is a legal basis which can provide a fallback scenario in case of disputes. Although this legal basis could affect all four dimensions of data sovereignty, based on the expert interviews, it appears that this basis which is generic, primarily addresses the data ownership antecedent. It provides a preliminary understanding of how the relationship between data provider, platform and data consumer works regarding responsibilities and rights. This basis is supplemented by a specific agreement between only data provider and data consumer for each transaction.

Secondly, certifications are used in practice to enhance trust among participants. These

certifications are not only a first barrier of entry for potential participants of lower quality, but can also help to communicate trustworthiness of certain participants across the data sharing network (both for data space or data marketplaces). However, the level of assurance depends on the certifications process, as was illustrated by the section on 5-1-1. When the certification is designed in a way that new entrants are actually audited at entry with periodic audits afterwards, it can be considered as process control. Primarily, because the auditee has to repeatedly show and prove that the process complies with the agreed standards. However, a certification could focus solely on the moment of entry to the platform as well. This means that at the moment of entry new participants are checked and verified, but that there are no further checks afterwards. This second scenario is mainly an example of gate-keeping and can be seen as input control.

Thirdly, experts mentioned the early phase of data sharing in practice. To fuel data sharing and to develop methodologies that enhance data sovereignty, several data sharing initiatives are currently working with a non-profit model or are funded by governments. This could explain why several initiatives are currently organised in the form of an association or foundation which means that the data marketplace-operator (or organisation behind a data space) is not primarily using mechanisms to achieve financial goals, which is frequently the focus of digital platform governance literature, as discussed in section 2-2-2.

6

DMMPs: data sovereignty challenges and potential solutions

This chapter introduces the findings from the expert interviews regarding data sovereignty challenges and potential solutions to address these challenges. Whereas 5 discussed contemporary data sharing initiatives and data marketplaces, this chapter focuses primarily on DMMPs. This chapter starts with a brief note to inform the reader in section 6-1 about the DMMP-configuration that was displayed during the interview discussions. Secondly, section 6-2 briefly addresses the anticipated advantages by the experts of a DMMP for data providers in general. Next, the chapter will further focus on data sovereignty-related challenges and potential solutions to address these.

Next, this chapter will discuss the findings to answer sub-question 3: *"How do the challenges of SQ2 translate into a DMMP-context and what additional data sovereignty related challenges can arise?"* Firstly, section 6-3 discusses the challenges that data providers face when considering the whole ecosystem of the DMMP. Secondly, section 6-3-2 elaborates on the challenges that data providers face internally considering data sovereignty in a DMMP-context. Thirdly, experts mentioned challenges that they foresee at the interplay between data providers and data consumers as well. This group of challenges was linked to the level of the data-transaction and will be discussed in section 6-3-3.

Furthermore, the potential solutions that were identified will be discussed to answer sub-question 4: *"Which potential solutions can address the data sovereignty related challenges identified for the DMMP-context?"* As these solutions could potentially address more than one challenge, these solutions were grouped around themes. Section 6-5-1 discusses the role of the governance structure and associated ownership of DMMPs. This is the group of architectural solutions. Secondly, experts related trust very frequently when asked about their view on improving data sovereignty in a DMMP-context, which will be discussed in section 6-5-2.

However, trust will often follow from other aspects, one of which is transparency and visibility. Additionally, being able to track what is happening appeared to be important in general, as a complement to trust as well, which will be discussed in section 6-5-3. Although this research project is mainly centred at governance-related challenges and solutions, the use of technology will still play an important role to address data sovereignty challenges as discussed in section 6-5-4. Lastly, section 6-5-5 discusses the role of use cases and the development of them to address data sovereignty challenges. The chapter is concluded with a brief overview of the challenges and solutions that were identified.

6-1 Note to reader: DMMP archetypes

The remainder of this chapter discusses the expert interviews and their view on the data sovereignty related challenges and potential solutions in a DMMP-context. However, several configurations of DMMPs are theoretically possible. Firstly, a DMMP could mainly serve data consumers by offering a data discovery entrance point to search existing data marketplaces in their quest for data. Furthermore, a hybrid setup is possible were theDMMP combines the previous configuration with servicing data providers by redistributing their data to data marketplaces. However, considering that this research has used a data provider's perspective primarily in relation to their data sovereignty, the interview protocol included a visual which placed the DMMP between data providers and existing data marketplaces. The data consumers where still linked to existing data marketplaces, as shown in figure 6-1. The visual is included with the other visuals used in the interviews in appendix E as well. The reader is advised to consider this DMMP-configuration as the baseline where experts have based their responses on.

6-2 DMMPs and potential advantages for data providers: expert perspectives

Before experts were asked more in-depth about data sovereignty-related challenges for DMMPs, they were also asked to provide their first perceived advantages and/or disadvantages for data providers. Some experts also shared their view on DMMPs in relation to data consumers, although the value of DMMPs for data consumers is highly dependent on the configuration as discussed in previous section 6-1. The perceived disadvantages for data providers are integrated in the remaining sections of this chapter. This section will briefly discuss the perceived advantages for data providers. The advantages mentioned by the experts can be linked to three broader topics: 1) Increasing reach for data providers, 2) DMMPs as data sharing outsourcing partner, 3) Improving data sharing operations. These three topics will be briefly discussed in similar order in sections

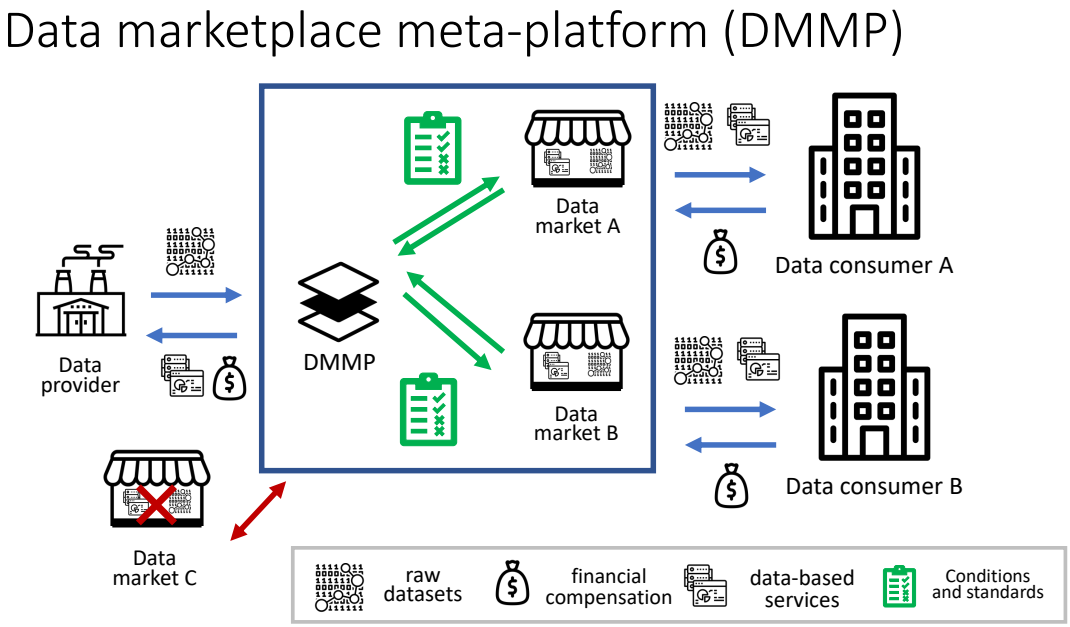


Figure 6-1: DMMP overview including participants

6-2-1 DMMPs to increase reach to data consumers

The experts that shared their perceived advantages of DMMPs frequently expressed that they could imagine that a future rise of DMMPs could enable data providers to reach more potential data consumers, because the DMMP offers them better access to data marketplaces. Among this group, some mentioned the potential increase in financial revenues as result as well. The quotes below illustrate these views, although this overview is not exhaustive:

"I want as many outlets as I can. Right? I mean, I want to make sure that my data is out there to be sold and it's available to whomever I want. You know, it's why do you advertise? You know, why do you hire salespeople? You know, it's the same concept you want as many channels as possible."

Expert 5

"I could spread my reach, reach a wider audience, either geographically or topically."

Expert 9

"Maybe in economical terms, I could make more money of it because I have more access to data consumers."

Expert 10

Although most experts mainly expressed this using a quantitative perspective (i.e. more potential consumers, more data marketplaces, more financial revenues), some experts also mentioned that DMMPs could also lead to a better fit between the type of data that a data provider has and the data consumers that the data provider is trying to reach. Some experts argued that one interesting service of a DMMP could be that it helps data providers to make better decisions regarding where to market data offerings, as illustrated by E8:

"And why you do include something under marketplace A and B, but not under C? Because I don't think the choice not to include it under C is not a conscious choice by the data provider, but something that is determined based on which input the provider delivers to the platform. The meta-platform will have to make that choice, otherwise it still won't work. Although, it can still work, but if there are thousands of data marketplaces, it can no longer be determined manually.[...] That if you offer this, then I can offer it on these data marketplaces, but not on this one. And that is a very interesting functionality, if you get that done."

Expert 8

6-2-2 DMMPs to outsource data sharing operations

Similar to the quote by E4 in the previous section, some other experts mentioned an interesting proposition that a DMMP could offer as well, as outsourcing partner. What this means is

that a DMMP gives advise and operates on behalf of organisations that want to share data with data marketplaces, but lack the internal capabilities to start sharing on their own. In this scenario, activities related to data sharing, e.g. selecting data marketplaces, determining the necessary level of control for a particular data asset, verifying credentials of potential data consumers, are taken over by the DMMP. Please consider these two illustrative quotes:

"I think there will be parties who say, I want to have that data with the meta-platform and then it is arranged."

Expert 1

"I could think of added value, added services. So basically, stating; I give you this kind of data and you cater for the rest, including, for example, legal aspects, contractual aspects, quality aspects or storage."

Expert 9

6-2-3 DMMPs to improve data sharing operations and efficiency

In addition to the two advantage-related topics discussed in sections 6-2-1 and 6-2-2, a third perceived advantage of DMMPs for data providers was primarily related to savings of resources such as time and costs. These savings would primarily be achieved by economies of scale, for example regarding legal fees:

"I also think that if you arrange it properly, the advantage could also be cost savings, especially regarding legal. That is in any case, if you can digitise agreements with each other and have them automatically set up, then there is less and less need for a lawyer and notary."

Expert 8

Please note that the expert mentions use of digital technology as a boundary condition to achieve the legal cost savings. Furthermore, although the central DMMP-approach was received with scepticism by many experts due to for example fears for lock-in, as will become clear in later sections, E9 described how this centralised configuration could actually mean an advantage for data providers as well:

> *"At first glance, it would seem that I don't need to cater for a lot of input formats that I need to supply for. [...] I would have a single point of contact, which makes customer relationships easier."*
>
> Expert 9

Lastly, the standardisation that a DMMP could offer can lead to time savings for data providers, according to E10. Please note how he relates the concept of *'enter once, use many'* to providing descriptive data (i.e. metadata):

> *"...I can imagine that it would be much easier to distribute my data to different marketplaces, would be like less workload than access every data marketplace and provide my data and type in the information every time, would be easier to find actual data marketplaces*
>
> Expert 10

Although this section provides a couple of quotes that can be considered reflective for the perceived advantages for data providers according to experts, several experts expressed critical views as well. More specifically, they shared challenges at three levels, which will be discussed next.

6-3 The data sovereignty challenges in a DMMP-context

6-3-1 Challenges at the DMMP-platform ecosystem level

Data sovereignty for data providers is impacted by multiple factors. One of them originates from decisions and configurations at the high-level of the DMMP-ecosystem. When data providers are faced with a DMMP, their decision to share data via such a platform is influenced by how the platform is designed and operates and what the impact is on their data sovereignty as a result. More specifically, the interviews indicated that a DMMP that is too dominant will diminish perceived and actual data sovereignty for data providers. This challenge is discussed in section 6-3-1.

Fears for DMMP-dominance

The challenge of fear for DMMP dominance was one of the most frequently mentioned topics during the expert interviews. Firstly, interviewees feared that a DMMP that is managed very centrally could impact the level of autonomy for data providers. This could for example mean that data providers are being limited in setting their conditions regarding data access for

certain data consumers. Additionally, a dominant DMMP could mean that data providers are forced to adhere to platform conditions (e.g. regarding data storage) or could miss revenues from their data as the DMMP adopts an aggressive fee-structure once it has captured a sufficient share of the market. For example:

> *"I can well imagine that a platform will be created that organises it very well from a technological perspective, but then starts to exploit participants in the platform."*
>
> Expert 1

> *"Hmm what you see is that the winner takes all platform is going to win. [...] Due to the market forces that we have in the Netherlands, for example, dominant players could arise. But then the question is, do you want that?"*
>
> Expert 6

> *"Personally, I think it's all about, and this is a winner takes it all scenario. So that the biggest player wins and the player, either meta-platform or marketplace that offers the most convenient solution to get data in and out, has the most aggressive approach to the market and the most convincing business models, probably end up winning."*
>
> Expert 9

These fears for DMMP dominance were especially strong when the DMMP-operator would be a stock-listed or at least commercial party. This illustrates that fears for dominance by data providers do not only arise due to the future market share of a DMMP, but also because of the ownership-structure:

> *"If there is another listed party there, who says: we do have a marketplace meta-platform, then by definition they have a certain interest, and that is increasing shareholder value. And that does not mean providing the most reliable service possible to everyone who is connected to it, whereas that is where you will have to move toward."*
>
> Expert 8

What this quote by E8 illustrates is that the goals of the DMMP-operator and its platform impact data providers. More specifically, the DMMP-operator could design the platform to pursue financial goals instead of offering a platform where data providers can retain data

sovereignty. This aligns with the findings from section 5-1 where several data sharing initiatives are operated by an association of organisations and where a non-profit model is used.

Additionally, although the concept behind a DMMP is that it could help data providers to reach a larger and more diverse selection of data marketplaces to offer their data, E9 warned that a very centralised platform could integrate those data marketplaces and compared it with Amazon, what could be considered an envelopment-strategy (Eisenmann et al., 2006):

> *"You see it with Amazon being now the marketplace for all kinds of stuff. And even added services, they integrated back into the Amazon platform. But you just had like this complimentary services being at trust services and authentication services, whatever. If there's high enough value they can be reintegrated into such a platform. And a data market meta platform could also easily become a data market itself by scooping up all the players because, eventually, they would probably argue, why would I need a data market A, B and C if I could make better deals with the data providers myself?"*
>
> Expert 9

When such a scenario materialises, data providers become even more dependent on a DMMP and the power of the DMMP could further increase. This increases the risk of aggressive behaviour by the DMMP-operator and higher transaction fees for example. As a result, data providers have less control over their data which is a direct impact to their data sovereignty.

6-3-2 Challenges at the data-provider level

Whereas the initial expectation was that the data sovereignty challenges would mainly appear at the DMMP itself or at individual data transactions between data provider and data consumer, several challenges were identified at the organisation of data providers, or even at individuals within these organisations. Firstly, experts sketched the current state of the business world regarding data sharing and data trading. This gave insight in what this could mean for a future scenario where business data is shared via DMMPs. Secondly, several discussions illustrated that business data is not a homogeneous class of data, and that both the potential value for the market and risks for the data provider gradually increases as data becomes increasingly sensitive.

Lack of data sharing capabilities

In general, experts emphasised that decisions regarding sharing of business data are made by individuals and that currently there is a gap to be even able to indicate as a data provider which data could be shared. Although the antecedents of data sovereignty discussed in section 4 are all possibly important for data sovereignty, a boundary condition is that organisations

that are data provider are able to determine how their data and business strategy relate to the DMMP and data consumers. This human element at the level of the data provider was mentioned by several experts:

> *"Because of course, there are companies who want to interact with companies. But what are companies? Companies are a bunch of people."*
>
> Expert 4

> *"Because it's in the end, every data provider, every entity, it's run by people and they have personal relationships with each other. So a company is not an abstract thing like Company A and Company B are doing business. But it's a person here and the person there."*
>
> Expert 9

Furthermore, the lack of capabilities also differs with the size of organisations. In general, experts mentioned that larger organisations might be better positioned to share business data via DMMPs. Some experts related this to data management practices that are more formal and mature in general in larger organisations. For smaller, or at least less mature, organisations it could mean that using DMMPs is more challenging. For example, SMEs could face difficulties when entering a DMMP to get their mix of conditions right for their type of data to retain data sovereignty. Beyond just difficulties, it could also lead to severe impacts to the overall business, especially with business-sensitive or personal data, which will be further discussed in 6-3-2.

> *"But I don't think SMEs are going to share raw data on such a marketplace."*
>
> Expert 6

> *"Knowing that, and then I look at the larger market of the SMEs, well, there is still a lot going on with smoke signals and faxes. All old junk."*
>
> Expert 7

However, other experts also explained that external parties could actually help organisations with lack of data sharing capabilities to improve their practices and mitigate risks for example. While discussing data storage, one expert compared this situation with how SMEs for example use external providers to store data:

"You see this a lot in larger organisations that also have those capabilities. The smaller parties become, the more that lies with external parties and IT parties."

Expert 1

Additionally, the challenge that data providers lack the capabilities to start sharing data and assess their data sovereignty needs is not only dependent on the size of organisations. Expert insights also illustrated that it very often depends on the type of industry sector. In this light, the conversation with E5 gave insights about businesses in the capital markets industry. This expert, with a long career in this industry, sketched how the practices in trading data in this sector have matured and evolved over the decades. As a result, the discussion made clear that this industry is highly mature regarding data practices, and the legal constraints:

"That said, you know, the capital markets as a data provider area is fairly mature [...] And I think is really interesting. And what is also interesting is because the financial and capital markets industry of course, is highly regulated one. So they are really mature in compliance practices and stuff like that."

Expert 5

This conversation with E5 also showed how Bloomberg operates as one of the largest re-distributors of capital markets data. This company does not only redistribute data from providers, but also builds additional services on top of them. E5 described how this platform is structured as well:

"Once you're a Bloomberg client, the friction goes down considerably because all it is, you know, you click requested entitlement, and then the next thing you know is you're getting the data and then it'll be reflected on your bill. And then, Bloomberg will go back to the provider and provide a royalty or whatever it may be."

Expert 5

Furthermore, experts also mentioned that it is difficult for data providers to estimate the value of their data. This links to data sovereignty, because even if a data provider is able to control their data, from a business-perspective using a DMMP makes no sense if they are not able to capture sufficient or appropriate value of their data. Estimating data value ex-ante was not only frequently mentioned during the sections of the interview specific for data sovereignty, but in the discussions about factors that are barriers to data sharing in general. The value issue is also highly linked to whether data providers have an understanding of potential data consumers or not:

"Also the question is, what data do I now have available to actually share that would benefit others?"

Expert 3

"They're trying to estimate, do I have any value in actually extracting this raw data and trying to package it for this type of use case or this type of systems? And that's what is the actual situation of the economy right now on the data economy, which is that we don't know."

Expert 2

As the quote of E2 illustrates, the challenge regarding value anticipation is also related to the usage and use cases of data, which will be discussed in section 6-5-5.

The challenge to estimate data value is related to the type of data as well. For example, a dataset that could help data consumers to double market share in a highly-profitable market segment is much more valuable than weather data that is generally available. Section 6-3-2 will further discuss the interview results regarding this challenge.

Sensitive business data

While discussing data sovereignty challenges and potential solutions for DMMPs, experts frequently mentioned that the necessary measures highly depend on the type of data that is traded. Type of data is a very broad definition and can be broken down in several dimensions, for example historic or live data, personal and non-personal data, industry-specific, region-specific and data versus metadata. However, during the conversations with the experts it was mentioned in advance that the scope of the research is mainly on non-personal or at least anonymised business data. One of the findings from the interviews was that there still is an substantial refinement that has to be made when sharing business data via DMMPs. The experts illustrated the spectrum of business data ranging from open data to highly-sensitive proprietary data:

"Simply put, you have open data, you have condition-based shareable data, and you have classified proprietary data."

Expert 7

The experts also emphasised the relation between these classes and value at the marketplace:

"But when it comes to closed data, data with consent, data that I don't just want to make available, but that I want to know that you are you. That is often data with the most value, then it is a lot more complicated."

Expert 1

This relates back to the complexity challenges that are amplified in a DMMP-context, especially compared to data marketplaces or even bilateral data sharing. Furthermore, sharing sensitive business data can pose risks to the business of the data provider:

"Obviously, the more damaging or sensitive data could be to you as an organisation, the more expensive it's getting."

Expert 9

"I could be scared of sharing my data because, first of all, I could give away information that is crucial to my company, or not crucial, but also secret, company secrets."

Expert 10

These quotes illustrate that the type of data that is shared via the DMMP partly determines how challenging the scenario can get. Furthermore, especially considering the lack of capabilities mentioned in 6-3-2, classifying data assets could be a challenge in itself for potential data providers. The expert-interviews also returned challenges that could arise for data sovereignty at the level of an actual data transaction between data provider and data consumer. These challenges are discussed in section 6-3-3.

6-3-3 Challenges at the data transaction-level

Whereas section 6-3 elaborated on challenges that arise at the larger DMMP platform ecosystem, and section 6-3-2 elaborated on challenges at the level of an individual data provider's organisation, this section reports on the findings from the expert interview related to the level of the data transaction. This means, exploring potential challenges that arise when data providers share data via the DMMP with data consumers. Firstly, data access and data usage in are highly inter-twined, with data storage as well. Section 6-3-3 will discuss the fears and challenges related to data access. Similarly, section 6-3-3 elaborates on the expert views on challenges that are linked to data usage and storage. Lastly, section 6-3-3 discusses an issue that could arise after transactions: disputes.

Fears to grant access

Many experts feared that using a DMMP could mean that unwanted parties could get access to data. This fear is dependent on the type of data as well: the more sensitive the data, the greater the consequences of unwanted access. When asked about their view on the relation between data access and data sovereignty, the majority of the experts mentioned that data access should always be granted by the data provider. They also related granting data access back to the structure of the DMMP, both from an architectural and governance perspective. In this similar direction, experts often argued that a more dominant DMMP would mean that it would become more difficult for data provider to keep data access control in their own hands. These two illustrate general comments regarding data access:

> *"In my opinion. Data access, well I assume with the data access, that the data ultimately remains with the data provider. And that the meta platform and also the marketplace itself will never have the data, so in that respect you have quite good control over access to your data."*
>
> Expert 8

> *"I mean, I would say me thinking about being a provider. I would need insurance that data access is in the way I need it to be."*
>
> Expert 9

During the conversations about data access, experts also emphasised that data access is a challenge because data access can be sub-divided in several finer categories. For example, it could means access to metadata or the data itself. Furthermore, experts also mentioned that data providers would need several options to grant access and set specific conditions. Firstly, this could relate to time- or count-conditions regarding data access:

> *"So for let's say the data access to like also you divide the access. Yeah. So maybe how many times you can access the data."*
>
> Expert 11

Whereas the previous three quotes are mainly addressing the challenge of 'how' and 'what' can be accessed, several experts also expressed their anticipated challenges regarding 'who' can access. This means that, for data provider's data sovereignty, an issue could be that it is unclear who is accessing their data when using a DMMP. During the data access section of the interview, this concern was mentioned frequently, as the quote below illustrates:

"For a provider, I would like to know who can access my data"

Expert 4

Section 6-3-2 described how organisations are in the end build up on people, the same applies to data access challenges as well. E1 mentioned that data providers would also prefer to grant access not only to specific data consumers, but even to individuals within the data consumer's organisation.

"For example, you may only view that data for a specific time, for a specific use case and only by a specific person of the purchasing party. You can make that very fine. You could think about that. For some data, this may also be much less relevant, such as geodata, geolocations."

Expert 1

Furthermore, this quote underlines the specific difficulties that each type of data brings for data providers to stay in control, as discussed in section 6-3-2. However, what the interviews made clear was that fears to grant access were also highly related to fears for not knowing who the data consumer is. Whereas in a data marketplace-context the relationship between data provider and data consumer can already become unclear, the interviews showed that these challenges would only become bigger for a DMMP-context. The quote below gives an example:

T: *"Then actually, if I understand correctly, does the meta platform increase the fear of losing control? Compared to a single data marketplace."*
E1: *"This is already the case with a single data marketplace."*
T: *"So it gets even worse?"*
E1: *"The further away it is, the more complicated it actually gets."*

Expert 1

Another expert described this as a proximity-issue:

"I notice now that I go through this, that proximity is a very important one. The further away something is from you, the less you trust it."

Expert 8

The importance of trust to enhance data sovereignty, and what it needs to establish will be discussed in section 6-5-2. Lastly, as data usage control was often mentioned together with data access control when discussing challenges for a DMMP-context, for example by E3:

"Well, these are terms that make sense. Specifically, for data sovereignty, keeping control over your own data, but I think it's about two things: access control, someone needs access, or usage control, because a party needs it for a certain thing."

Expert 3

As a result, challenges regarding data usage will be discussed next in section 6-3-3.

Staying in control over data usage and storage

Similar to data access, the majority of experts included data usage in their response when asked which antecedents of data sovereignty would be the most critical in a DMMP-context. Additionally, they mentioned there are multiple challenges. The quotes below illustrate some of these initial responses:

"I think the most important thing you need to arrange is the right side, so data processing/usage and data storage."

Expert 8

"It would be even more difficult to track down the use of my data. Yeah. Basically, the data sovereignty aspects. That there's another layer which has to track the whole provenance of the data usage."

Expert 10

As illustrated by the quote of E10, the DMMP creates another layer that increases the complexity for data providers to track data usage, both by data consumers and potentially

complementors as well. This links back to the increased complexity at the DMMP platform ecosystem level and the challenges it brings for data sovereignty. Furthermore, the comment by E10 regarding provenance indicates that data providers fear not only the actual usage, but also that they lose visibility over the usage. As a result, it became clear that the data usage antecedent of data sovereignty comes with information needs by data providers. As discussed in section 2-2-1, formal control mechanisms often have certain information requirements to compare actual behaviour with agreements, standards and procedures. Tracking of data usage does not only include the actual data, and refined outputs of it, but also metadata as was explained by E6:

> *"But you want to be in control of your own data, but also of your metadata. That does not mean that I always have the data with me, but that I can verify what happens to my data, that's just not the case now."*
>
> Expert 6

Several experts mentioned data storage in a close relation to data usage. As a result, data usage challenges are included in this section as well. Firstly, several experts argued that for the centralised DMMP that was displayed on the visual (please refer back to figure 6-1), storing data at the platform was too challenging for data providers regarding their data sovereignty, and that at least keeping the storage of actual data at the data provider would enhance control:

"You want access to that data at the right times. I think that is crucial in a data sovereignty context, you don't want some kind of central database"

Expert 1

"And in terms of data exchange. Well, I can imagine that the decentralised solution, where my data stays, where it is like with me, and it is only exchanged to the data consumers, that would be the better solution than uploading data to the cloud."

Expert 10

"But if you have the data on your premises, then you have more control."

Expert 11

The storage of data in a DMMP-context is not a single decision that has to be made. The expert interviews provided insights regarding which additional challenges can arise due to the additional services provided by the platform and the functional requirements such as convenience and speed:

"What you can of course imagine is that when it comes to such a marketplace, it is stored in multiple ways. We also want to be able to run analyses on this, we want to combine things. Then it is stored in several ways to make retrieving that data easier. [...] If you store it centrally, that central storage is an important part of the entire link. That makes it more difficult to have the confidence again to be able to assume that everything is going well."

Expert 3

Furthermore, Expert 1 noted that when designing for convenience and speed of services offered by a platform, the potential is affected by storage decisions:

> *"The difficult thing about that is often: how do you keep that data up to date if the data is stored somewhere else? Because often when you copy it to a database, the data is already outdated. So how do you ensure that it is up to date, and thus retains its value."*
>
> Expert 1

It turned out that there was no real consensus regarding data storage in relation to data sovereignty. In contrast to the experts that were very critical about centralised data storage. For example, E2 argued that there was not one specific challenge, but that at least data providers should be able to decide:

> *"You should have your data where you want, with whom you want to actually store it."*
>
> Expert 2

Another expert mentioned that the data storage challenge highly depends on how the platform operates and whether the platform is able to show that it operates in the interest of the data provider:

> *"If the meta-platform is able to control the data usage or data processing. And if there was a good solution for that, then I wouldn't be so scared of uploading my data because even if I stored my data on a meta-platform and the platform is able to control very nicely the data usage, then I'm not scared of giving it away."*
>
> Expert 10

Similarly, E4 weighted data storage against data usage, but also access, and also argued that data storage was a challenge to a lesser extent:

"I think access and usage would be the first to important thing I think. Storage is something were several opportunities are. I think storage is more like okay, it could come from here. It could be there. I can imagine it's okay to have this, to have this, to have this: it's not so hard to define like usage an access. I want to be really sure that not an opponent is accessing my data. I want to be really sure that this data is not used for, I don't know, whatever. Yeah. So I think access and usage feel like having more weight than the others because ownership to me it's clarified before sharing. Yeah. I think storage has more possibilities than access."

Expert 4

Please note that this comment by E4 also includes the fear for sharing business data with competitors ("opponent"), which was also discussed in section 6-3-2.

Solving disputes

The experts did not only mention challenges they anticipated in advance or during data transactions, but also potential challenges that could arise afterwards. More specifically, several experts mentioned the issue of disputes that data providers could face in their relation with data consumers. They provided several examples about topics that could lead to disputes, although these mainly focused at data access and usage. Due to the scale of a DMMP, they also feared that there could arise more disputes. Some experts feared that solving disputes could become very resource intensive. Considering data sovereignty, the challenge of disputes is an integral of the problem because even if data providers can set all the terms and conditions they prefer, they can still lose control when there is no fallback when mistakes or errors are made. These can happen both deliberately or by accident, as is illustrated by E3:

"It's partly intentional, but it could also just be that something goes wrong by accident. That a logging server was accidentally turned on that is being hacked. However, it is getting more and more difficult."

Expert 3

During the discussions about disputes in a DMMP-context, experts sometimes also mentioned that this comes with responsibilities from the side of the data provider as well. More specifically, disputes can not only arise because the data consumer is not following agreements, but also because data providers can cause errors or non-compliance with standards, as was illustrated by E1:

"Everyone should have faith in it, for example the cyber security and metadata agreements. That everyone submits in the right way. But also trust in the legal agreements, what if something goes wrong?"

Expert 1

As is illustrated by the quote of E1 is that trust is also a component of this challenge. Furthermore, E1 mentions the legal aspects of DMMPs, which will be discussed in section 6-5. Several other challenges were mentioned in the interviews as well in relation to disputes. The quote by E8 summarises the aspects that were often mentioned in relation to the challenge of disputes in a DMMP-context:

"What I mainly foresee is indeed transparency and accountability, which will become a real problem. And how do you deal with a dispute resolution at a given moment? How do you deal with it when something goes wrong? Where can you go? Is that the meta platform? Are you going to the data marketplace?"

Expert 8

What this quote further illustrates is that this expert links this challenge to the governance-structure at the DMMP platform ecosystem-level, which was discussed in section 6-3-1. More specifically, how accountability and responsibilities are distributed across the DMMP and its participants. Additionally, E8, among others, linked this challenge to transparency, which will be discussed in section 6-5-3.

6-4 Linking the challenges to the data sovereignty antecedents

Whereas the previous section of this chapter discussed the data sovereignty challenges derived from the expert interview, this section will link these challenges to the data sovereignty antecedents discussed in section 4: data ownership, data access, data usage and data storage. Firstly, in their comments regarding fears for DMMP-dominance, experts described several potential outcomes of that situation that could harm data sovereignty for data providers. This challenge can impact all four data sovereignty dimensions. For example, a highly centralised DMMP could force data providers to store their data centralised, whereas providers could prefer storage at their own source. Furthermore, the DMMP-operator could enforce data providers to grant access to data for purposes which data providers do not prefer. For example, mandatory access to a sample of the data to improve the service level for data consumers. A similar scenario could happen regarding data usage: the DMMP wants to attracts as many data marketplaces as possible and could force data providers to provide their data to data

marketplaces that are build around a usage scenario which the data provider is actually not interested in or has risk-related concerns about. Lastly, DMMPs could adopt a strategy where they provide as many analysis services as possible to increase market share, which could lead to a loss of ownership for data providers.

Furthermore, the lack of data sharing capabilities at the data provider level does not directly impact the data sovereignty antecedents directly. However, it can lead to struggles for data providers to estimate their required levels of control. This applies primarily to data access and data usage, but potentially to data storage as well. In short, overall capabilities are missing to develop an effective data sharing strategy that includes an internal analysis of data assets, a process to determine which assets are business critical and which ones could be shared. However, data providers are still missing the capabilities to determine their preferences regarding under which data access and data usage conditions to share the data assets of the latter category. As a result, this challenge is mainly related to data access and data usage.

Sharing of sensitive, proprietary, business data which could theoretically have the most value at the marketplace as well, is related to the data sovereignty antecedents as well. Based on the expert interviews, these more sensitive data assets require higher levels of control over data access. Furthermore, due to the sensitive character, the requirements related to storage are higher as well. This includes security requirements for example. Furthermore, data providers are afraid to share this type of data with unwanted parties, which also indicates the challenge of identity. All in all, the challenge of sharing sensitive business data is primarily related to data access and data storage.

The fear regarding granting access to unwanted parties appears at the transaction level between data providers and data consumers as well. Furthermore, the experts indicated that data access should be subdivided in very fine-grained access controls. Currently, these fine-grained controls are lacking which increases the fears for data providers to grant access.

As already mentioned in section 6-3-3, data usage and storage are highly related to each other. The comments by experts also linked the data usage and data storage antecedents of data sovereignty to the lack of visibility and transparency.

Lastly, the experts views on the disputes issue for DMMPs were mostly centred around the position of data providers when things go wrong. Based on the examples that the experts gave, for example illegal reselling or illegal copy-making of data, this is primarily an issue impacting data ownership for data providers.

The challenges and sub-challenges can be linked to the data sovereignty antecedents as follows:

Level **Control problem dimension** **"Illustrative data provider's question"**	**Challenge**	Sub-challenge	**DO**	**DA**	**DU**	**DS**
DMMP-ecosystem: **Lack of control expectation** **"I won't have control over my data"**	**Fears for DMMP-dominance***		x	x	x	x
		Fears to lose autonomy for condition-setting		x	x	
		Fears for vendor lock-in	x			x
Data provider: **Lack of control awareness** **"Which data can we share?** **How do we control it?"**	**Lack of data sharing capabilities** **Moderating factors: size of organisation, industry sector**		x	x	x	x
		Lack of data sharing condition-setting maturity		x	x	
		Lack of or difficulties with data value estimation	x			
	Sensitive business data **Open, condition-based, proprietary >** **increasing value, increasing consent requirements**			x	x	
Data transaction: **Lack of access control effectiveness** **"Where did we lose control?** **How do we keep control in the future?"**	**Fears to grant access**			x		
		Access control has to be very fine-grained		x		
		Access control must be defined by provider		x		
		People aspect		x	x	
		Proximity issue		x	x	
	Losing control over usage and storage				x	x
		Tracking of usage is difficult				x
		Tracking illegal copies of data				x
		Data usage control configuration impacts storage possibilities and vice versa			x	x
Post-transaction control: **Lack of control resiliency** **"How do I gain back control?"**	**Disputes**		x			
		Large-scale dispute resolution	x			
		Restoring control for provider after dispute resolution		x	x	x

Table 6-1: The data sovereignty-related challenges in a DMMP-context. DO: data ownership, DA: data access, DU: data usage, DS: data storage, *Indirect effect on data sovereignty antecedents

6-4-1 Intermediary conclusion

The first half of this chapter has identified challenges related to data sovereignty in the DMMP-context to find answers to sub-question 3: *"How do the challenges of SQ2 translate into a DMMP-context and what additional data sovereignty related challenges can arise?"*

Firstly, three different levels of challenges were identified. Starting at the DMMP-ecosystem level, the primary challenge is the fear for DMMP-dominance by data providers according to the experts: data providers are afraid to lose their autonomy to determine their conditions when sharing data via DMMPs and are afraid to get locked-in. Compared to non-DMMP data sharing platforms, these fears increase. Secondly, individual organisations that could participate in the DMMP as a data provider can lack the capabilities to effectively determine their data control needs, although this challenge appears to be moderated by firm size and industry sector. This lack affects their capability to estimate the value of their data assets as well. Furthermore, organisations possess different types of data, ranging from open data which they could share more liberally, to highly sensitive proprietary data. At the transaction level between data providers and data consumers, there are fears to grant access, which are amplified in a DMMP-context compared to a smaller consortia network with a shared ambition as discussed in chapter 5. Additionally, data providers need very fine-grained access controls, which can be difficult to realise in a DMMP-context due to the large variety of data (e.g. different industries, different formats). Furthermore, data providers can lose visibility of data usage, which makes it difficult to detect illegal copies or data consumer behaviour in general. Data usage control and data storage are also highly interdependent: the storage configuration determines usage possibilities and vice versa. Lastly, after transactions have occurred, disputes can arise, where data providers have to protect their interests (e.g. to make sure that they receive financial compensation, or can withdraw data from the consumer). However, due to the scale of a DMMP, disputes will need to be resolved with relatively little manual involvement.

6-5 Addressing the data sovereignty challenges

In response to the main challenges that were identified in the interviews, this section will elaborate on the solutions that were proposed that could address these challenges. This section will start with expert perspectives on the larger governance and ownership structure of DMMPs and proposed directions that could help to increase control over data for data providers. Secondly, trust will be discussed and its importance to overcome data sovereignty challenges. However, although trust could be a nice goal to achieve in a DMMP, trust in itself is not directly an actionable solution. As a result, section 6-5-3 will elaborate on the role of transparency, tracking and visibility for data providers and their data sovereignty. Next, technology-enabled solutions to help addressing several of the challenges of section 6-3 are discussed in section 6-5-4. Lastly, although use cases do not directly improve data sovereignty

for data providers, they could have an intermediary role to help with addressing the challenges. Furthermore, use cases were also mentioned very frequently during the interviews by experts.

6-5-1 Finding the right governance and ownership structure

Several experts warned for a centralised DMMP-structure and its effect on data sovereignty for providers. Although at the highest level all these solutions had in common that there should be decentralisation, they can be categorised in different categories: limit the centralised functionality of a DMMP, organise decentralised decision making and shared ownership of the platform.

Limit centralised functionality

All experts mentioned that a DMMP offering all services, including data brokering, distribution of raw data, additional services and data storage would have a severe impact on data sovereignty. As discussed in section 6-3-1 this is related to vendor lock-in effects:

> *"First of all, it's vendor lock-in. Having a single, powerful platform makes for me as a data provider to be dependent upon that."*
>
> Expert 9

First of all, when asking what a DMMP could to to enhance data sovereignty considering DMMP dominance fears, most experts first stated that the actual data itself should not be stored at the platform and that the platform should not be able to access it. At least not without consent from data providers. For data providers, it is crucial to self-determine where to store data, and not to be forced to store it at the DMMP, as illustrated by these quotes:

"I think that is crucial in a data sovereignty context, you don't want some kind of central database."

Expert 1

"Everybody is doing their own marketplace of data. So, if there is not, and that comes in with your study, of do we need a uniform platform that gathers all the automotive data to actually get to work? That's an approach, but I don't think that's the right approach. It's not centralising the actual data in one place from one platform that will actually make it happen."

Expert 2

"If you store it centrally, that central storage is an important part of the entire link. That makes it more difficult to have the confidence again to be able to assume that everything is going well."

Expert 3

"By the way, I never say on a data marketplace, but through a data marketplace. In my opinion, such a data marketplace should in any case not become another data monopoly with everything on it."

Expert 8

"And in terms of data exchange. Well, I can imagine that the decentralised solution, where my data stays, where it is, like with me, and it is only exchanged to the data consumers that would be the better solution than uploading data to the cloud."

Expert 10

Overall, considering data sovereignty for data providers, experts were arguing that a DMMP should be a platform to improve the connection between data providers on one side and data marketplaces and data consumers on the other side. This limits the role of a DMMP from a full-service provider for data providers into a platform that mainly improves the level of connection between market sides and could help to improve interoperability between data marketplaces. This role as a data broker was referred to by E6 as the yellow pages of data marketplaces, data providers and data consumers:

> *"I would not at all share data via a data marketplace. I just started my answers earlier with it, but if you were going to start working with it, I'd like to stay in control. So that you have your data sovereignty in control, in other words that such a yellow pages can show that it will not touch your data and will not copy or store it."*
>
> Expert 6

Please note that this quote also includes that the DMMP should not be able to access data provider's data and mentions the characteristic of data that it can be easily copied. The latter will be discussed in section 6-5-3.

The quote by E10 also illustrates that the possibilities to decide to develop a centralised or decentralised platform are also related to how the data marketplaces are organised that the DMMP wants to connect to:

> *"Well, on data storage, it really depends on the meta-platform. And first of all, how the data marketplaces themselves work. If they are central, they have to have central storage or decentralised, then it really depends what a meta-platform should do here, as it also should store data or only the metadata to share and the actual raw data."*
>
> Expert 10

Overall, this quote illustrates the additional complexity that comes with a DMMP, as there is an interdependency with the structure of the data marketplaces that the DMMP wants to connect to. Furthermore, a decentralised DMMP could also be limited regarding for example additional services that can offered, as the actual data of providers is less accessible. However, experts did not mention this potential consequence of a more decentralised approach, with could mean that for data providers the benefits outweigh potential negatives.

Shared decision-making

Many experts also argued that data providers should be able to have a say in how the DMMP should function and for example which standards to adopt. One centralised entity that has the central power to make decisions was at least something very harmful for data sovereignty according to the experts. It could also mean that data providers would not even consider a DMMP as illustrated by E2:

"You won't get any provider interested. They want to know in advance what they're getting. So, in our view, building the platform without the provider's interests at first hand, it doesn't make sense"

Expert 2

It is important to emphasise that shared decision-making also relates to the decision making process itself and the distribution of power across the DMMP, as illustrated by E6:

"This in turn has to do with the power of different parties. That is more about the governance model. Here's I think a 15th god mode doesn't work but making agreements does."

Expert 6

One expert also mentioned the decision structure were existing data spaces or marketplaces are involved in decision-making:

"That you do it more in a federated way, so without 1 party standing above it and dictating who can and cannot join, but that it is the collective group of marketplaces that say, okay you meet a common set of requirements."

Expert 3

Furthermore, this quote also gives an example of a decision-area that could also be impacting data providers: requirements to enter the platform. In addition to shared decision making and distribution of decision rights, E9 also mentioned that the whole decision-making process should be transparent:

"You can make your governance processes open. See, we have a meeting every month. This is the minutes. This is our shareholders, stakeholders, participants. This is all open. You built your trust with your name"

Expert 9

What this quote further underlines is that transparency of governance is also an enabler to create trust, which will be further discussed in section 6-5-2. Furthermore, considering the previous section 6-5-1 about a full-service DMMP versus a limited DMMP as data broker and the findings discussed in this section, the findings align with Tiwana (2014, p. 131): "All the dimensions of platform governance must be aligned with its architecture."

More specifically, the experts indicated that the centralised functionality (e.g. data storage) of a DMMP should be limited and they preferred shared decision making. Whereas the former relates to the platform architecture, the latter relates to the governance structure.

DMMP ownership structure

Next to views regarding which functionality and governance structure could enhance data sovereignty for data providers, some experts also linked the fears for dominance as discussed in section 6-3-1 to the ownership structure of a DMMP. As this topic also has implication for the structure regarding governance discussed in section 6-5-1 it is discussed next. The experts that discussed this issue and their views on a potential solution were all against one centralised commercial organisation behind a DMMP. Mainly because a commercial DMMP-operator could exploit data providers or at least limit for example the decisions that providers can take autonomous or democratically. For example:

> *"But what makes me doubtful is that such a marketplace will, all the time and every time, be coupled to commercial aspects and capitalistic systems which are inherently non-democratic."*
>
> Expert 9

When exploring which solution could be more preferable compared to a commercial DMMP-operator, two potential options appeared. The first one is shared ownership by all DMMP participants. This means that data providers are partial owners as well, which often automatically gives a minimal level of decision rights as well. This can be compared to shareholders participating in the annual meeting of a stock-listed company. Experts described this solution as ownership by a cooperation, association or foundation, at least a solution were data providers are substantively more involved than when a DMMP is owned by a single commercial party. For example:

> *"Because then you become dependent on a commercial party in the middle again, that doesn't make any headway, so it has to take a cooperative form."*
>
> Expert 8

During the discussions, some experts also mentioned the scenario in which a governmental organisation own and/or operates the DMMP. However, when asked whether governments would be better for the ownership-model than commercial parties, experts indicated that governments should own or operate a DMMP neither. For example:

T: *"Do you think the government is really a major driver to get this off the ground?"*
E7: *"Definitely, at least, a major user of it. It should not be the owner of the system, so to speak, because then you will again have insufficient trust."*

Expert 7

Considering, the quote by E7 and the comments by experts regarding commercial parties, the majority of experts indicated that a DMMP should be organised and governed by a federated or at least decentrally-owned organisation. What the quote by E7 further illustrates is that trust is an important topic, not only regarding the DMMP-owner, but in general for the data sovereignty challenges. Trust is discussed next in section 6-5-2.

6-5-2 The importance of trust

The vast majority of experts emphasised on the importance of trust for data providers and their data sovereignty. More specifically, data providers must have sufficient confidence that a DMMP will operate in a way that addresses their needs. In addition to confidence, based on the comments by experts, trust is also related to the level of autonomy that data providers have. This autonomy could for example apply to the conditions that data providers can attach to the data they offer on the platform. The conversations underlined that a very centralised DMMP with a commercial DMMP-operator would eradicate trust beforehand as these quotes illustrate:

"That's right, if it is an unreliable party that manages this meta-platform, then it becomes a completely different discussion."

Expert 1

"So as long as all the conditions are defined by me and no one other. Also not the platform of course."

Expert 4

Furthermore, when experts were asked to compare the centralised DMMP with all functionalities to a configuration that would lead to more trust by data providers, they often referred back to a more limited DMMP, with a more decentralised structure as illustrated by E3:

"But I think the idea of a meta platform is actually a very logical one. The only question is, is that meta-platform a single entity floating above something, or is the meta-platform the combination of all the marketplaces."

Expert 3

Furthermore, experts argued that data providers should have sufficient trust in data consumers as well:

E1: *"At the highest level, for example, you can look again at how you organise identity and metadata."*
T: *"Especially because identity is important for trust?"*
E1: *"Yes, yes, identity is crucial and not really solved yet."*

Expert 1

This quote that links trust to identity illustrates that there is also a component of verification of trust, in this example regarding identity of data consumers. This will be further discussed in section 6-5-3. During the discussions about trust and its relation to be able to verify as a data provider, many experts linked this directly to transparency of what happens with their data. For example:

"But I have trust because I actually see what is actually being done with my information."

Expert 2

"So if I can provide any information I want to provide, because if I can control what's happening there, then I'm fine. I need, of course, to be able to check who is accessing there. So there we have the trust and verify the credentials..."

Expert 4

During the conversation, E4 also argued that verified identity could also enhance trust among data providers in addition to trust between data providers versus the DMMP and data consumers:

"If you define some rules like only people with a trusted identity from the following providers can interact. Of course, this improves a lot of things for providers and consumers."

Expert 4

Some experts emphasised on the role of data providers themselves even more, and argued that trust in the DMMP-ecosystem also comes with responsibilities for data providers as illustrated by E5:

"And, you know, conversely, it would be unreasonable to expect that the platform provider or even the consumers become liable for that because they're buying data from a trusted source. And a trusted source has to mean that as well, that, you know, that they are compliant in their data gathering practices."

Expert 5

Please note that compliance regarding data gathering practices implies that data providers are responsible for the process before providing the data. In the scenario that personal data is traded via a DMMP, this could for example mean that data providers ensure that they have received explicit consent from the subjects in the dataset in line with the GDPR.

In relation to trust in the DMMP and data consumers, experts were also asked to describe in more detail how this could be practically organised. Several experts mentioned the role of certification. Certification means that DMMP entrants have to complete a review process before being allowed to the platform. This review does not only include a review of the organisation, but also the technical components that are used for data sharing via a DMMP:

"As a rule, you have two things that you would like to certify if you want to have it completely, or as far as possible, conclusive is the piece of software. Not every party is going to write its own software to participate in such a marketplace. But they just use a product that is available, and they install it. That is the most important. If the software is properly configured in that sense, you already have a lot of that confidence. Then comes the question, can you trust that organisation yourself? The organisation where it is deployed, or a service provider that is linked to that consumer or that marketplace that provides that service for you to set up that connection, that it is certified. That is the next step."

Expert 3

When comparing bilateral data sharing agreements with the use of DMMPs, this expert also explained why the decreased proximity in a DMMP-ecosystem between platform participants

complicates trust and how certifications, among other control mechanisms and governance decisions, can address this issue:

> *"Exactly, then the contact is also more personal, so that trust may come sooner. And with a marketplace, it really has to be built up through provisions. Also thanks to those certifications."*
>
> Expert 3

Other experts linked trust in a DMMP to reputation. They argued that trust is in the end build up by showing and proving that the practices are in line with agreed standards. One expert for example made an analogy with Bloomberg, which is one of the largest distributors of capital markets data:

> *"I think you develop a reputation of fair dealing and reputation of, you know, making sure that the proper controls are in place, that you're not seeing any leakage in the market. Yeah, absolutely. I don't know how you get to that. I don't know when you get to that. But sure, I don't know too many firms who wouldn't want to distribute their data through Bloomberg, for example."*
>
> Expert 5

In this quote, the expert links reputation with "fair dealing", which relates to how for example revenues are distributed and whether conditions are in line with data providers' preferences. Secondly, the "proper controls" align with earlier expert statements that trust in a DMMP ecosystem is much more complicated and increases the need for control mechanisms. One expert related the reputation of a DMMP with the topic that individual data providers are in the end collections of individual people:

> *"You built your trust with your name. Because it's in the end, every data provider, every entity, it's run by people and they have personal relationships with each other."*
>
> Expert 9

This section has provided several connections between trust and control mechanisms, reputation and transparency. Furthermore, whereas control mechanisms come with information needs to compare actual behaviour with agreements or standards, transparency implies that behaviour should be visible. Whereas the former can improve data sovereignty because it can make controls enforceable, the latter is a basic condition to get the information required to verify compliance. As result, next section 6-5-3 will further discuss the solutions to address the information needs related to trust.

6-5-3 Trust, but verify

As mentioned in the previous section 6-5-2, experts argued frequently that for data providers and their sovereignty, trust is very important, if not a boundary condition. Furthermore, experts linked various types of information (e.g. data consumer behaviour) that would contribute to that trust. What the expert interviews made clear was that trust for data providers is not only related to the type of DMMP-operator and related structure, but also the information they can acquire to make sure that they can not only set their preferred controls, but can also verify that there is actual compliance to them. As a result, trust is build on evidence based on actual information.

These information needs can be separated in two categories. Firstly, data providers need to be sure that they exchange with trustworthy parties. This does not only relate to data consumers, but also to the DMMP-operator and potentially complementors as well. This relates to the certification of both parties and under-laying infrastructure components. Secondly, data providers that are entering an actual data exchange need information that can provide evidence that the data consumer can be trusted. Furthermore, during and after the transactions, data providers need information to monitor the data usage by the consumer. This also relates to monitoring data storage. In this case, it does not only include the data that was provided initially, but also where processed and analysed outputs are stored. According to the expert interviews, this entire chain of information is necessary build up the data provenance.

While discussing the problem of creating new data with raw data, which could be one of the goals of a DMMP, E5 described these data outputs as derived content:

> *"Where it gets tricky, is when people use your content to create derived content. And where do you draw the line to say, okay, the consumer who created this derived content, at what point is that their content versus you have a piece of it because your data was used in the creation of it."*
>
> Expert 5

What this quote further illustrates that a potential solution could be is to agree between data providers and data consumers beforehand not only data usage possibilities, but to also make agreements regarding these derived content. However, in order to make these agreements enforceable, there is still information needed for the data provider to know when a data consumer is in breach. Unfortunately, the interviews did not bring a clear solution forward how this could work out in the relation between a DMMP and the data marketplace which are both in the entire chain.

An example where monitoring can help data providers to stay in control, is to track the creation of illegal copies of data to overcome this issue. For example:

"So, you need to be a agreeable on both fronts as a seller and as a buyer that what you're getting is what you expect and what you're selling is what you can actually expect as the usage also. Yeah. So again, and that's a big complexity with data because data is actually really easy to replicate"

Expert 2

"Data in its nature is digital. Digital means like there can be a thousand copies without any overhead, without any cost."

Expert 9

Considering that data is inherently vulnerable for duplicates, overcoming this issue by sufficient monitoring can enhance data sovereignty for data providers. To improve the monitoring, many experts mentioned the labelling of data. This could enhance data sovereignty for data providers, as the label can make the tracking more reliable or transparent. Experts argued that labelling of data can also help to address the issue that data providers of raw data lose ownership when data is further processed along the chain.

"But there is a lot of discussion around tagging contents so, you know, who the ultimate owner is. And I think that's the answer. So that at the end of the day, if you were to review where the content comes from, you can say, ah, this came from data provider A and you owe data provider A a licensing fee because you use their content without their permission or now you have to have permission. To get a license to use it."

Expert 5

This quote does not only illustrate the role of tagging and labelling of data, but also explains which role this could play for data providers to protect their financial interests and to capture value from derived insights based on their data. More specifically, the legal aspects of data and opportunities for data licensing are discussed later in this section.

Regarding data labelling, E4 saw opportunities for the DMMP to orchestrate this process and to provide the necessary guidelines to get to a standardised and trusted labelling practice. Please mention that this expert did not only refer to labelling to address the data ownership issue, but also to make data marketplaces better searchable, which could be a thing were a DMMP can create value and improve the data discovery functionality:

"What do you just do with the meta platform? You make them searchable. And I mean having really good searches on this, because if you maybe find a solution for they have, you can access how they label their stuff. There are some labelling guidelines out there. If we if we just reached a point where we standardise labelling and meta platform to search, all this will be so helpful because then you don't have to register on every marketplace you're just seeing up in this marketplace to find the data I need."

Expert 4

While discussing tagging and labelling opportunities, experts often mentioned the legal enforceability of agreements and contracts as well. Currently, most data exchange is based on the use of a legal contract between parties. In general, experts mentioned that for data providers these legal aspects will remain important as well for a DMMP-context. More specifically, a legal basis provides a fallback scenario in case of disputes. This also relates back to the tagging and labelling of data. Namely, tagging and labelling can make the gathering of information to verify compliance with contracts more effective and transparent.

Some experts also emphasised that a data provider should not really sell their data. Instead, they argued, they should provide a license to data consumers to grant access under conditions. This license contractually restricts ownership of data to the original data provider, whereas the data consumer only acquires access and usage rights. This was underlined by E5 who also shared personal experiences how licensing of data is used for data in the capital markets industry:

"As an advisor to a data provider, I would never say to sell your data. I would always say you license your data. And then license means that's controlled by you. You own it, you're the owner. You never give up ownership." ... "And that licensing can certainly be handled by a marketplace."

Expert 5

Mention the potential role of a DMMP that this expert mentions related to this topic. However, this expert also elaborated on the boundary conditions that are necessary to make data licensing workable, such as regulation and audits. These conditions were met in the capital markets industry which this expert has most experience in. However, E5 also expressed that this is not the case for other industries. This relates back to the complexity of one of the applications were DMMPs are supposed to bring value: cross-sector data sharing. Please refer to the quote by E5 as well:

"Getting in control is very specific to the type of data, to the providers. In the financial market the contract-based approach is pretty much respected because there are a lot of regulations and audits that are done on the data controls. So, you can actually have validation and show people that are using the data that they should because it's all contract based and it's very regulatory. And other markets are different because you don't have those regulatory elements in place..."

Expert 5

Several other experts also expressed their anticipated difficulties with a legal approach to enforce control for data providers. Those concerns related besides cost-concerns mainly to whether it would even be feasible in a complex system that a DMMP could be. Please note that this is opposed to what other experts mentioned when discussing advantages of a DMMP which was discussed in section 6-2-3. In general, experts made a distinction between legal and technical approaches, for example:

"But you can do access control and usage control via legal enforce-ability via contracts, or via technical enforce-ability, which you can use to technically enforce that someone cannot do something."

Expert 6

"They absolutely go hand in hand, so both legal and technical assurance. That has to go hand in hand."

Expert 7

Furthermore, although these experts explained that the legal-approach is currently still dominant, they preferred a shift towards technical enforce-ability of contracts and agreements, for example:

"And then in the coming period the question is, how can you actually make that sliding scale from technical and legal as far as possible towards technically enforceable."

Expert 3

This means that compliance is monitored in a much more automatic and technical manner. For example, if a data provider has agreed with the consumer that its data is only used to test one algorithm in a predetermined shared data safe on the platform, the access to the

data consumer should automatically be blocked when an employee is even trying to copy data outside that safe. Additionally, this could also mean that the data consumer is automatically reported and can be reviewed for removal from the platform.

However, some experts were sceptical regarding a shift towards automating by use of technology, because they argued that this could have a negative impact on trust. For example E9 mentioned the importance of the human element for trust:

> *"So even as attractive as it might sound and seem in the future, everything is being done automatically by smart contracts and the likes. There's the human element that is missing, and that is crucial in forming trust relationships."*
>
> Expert 9

This section provided expert perspectives on the solutions that can address the information needs to establish trust in DMMPs, partially enabled by tagging and labelling. Furthermore, this tagging and labelling is part of a broader approach where technical solutions provide the foundation for a shift from legal to technical enforcement of data contracts and agreements. The next section will elaborate on other technological methodologies that could provide solutions to address data providers' needs regarding data sovereignty.

6-5-4 MPC, anonymisation-techniques and technology-solutions

In the conversations with experts regarding data sovereignty challenges and potential solutions for a DMMP-context, several technology-based solutions were mentioned. These included MPC, a technique that is used to make analyses on several data-inputs, without the need for the separate providers to access each other's data. Furthermore, it can be used to control data usage. Secondly, several experts indicated that anonymisation-techniques could be useful for data providers to enhance their data sovereignty. Whereas MPC can be used to control data usage, these anonymisation-techniques could mainly lower the impact in case of data leakage. Furthermore, anonymisation-techniques can protect the actual dataset, whereas they can still offer data consumers the possibility to conduct sensible analyses. Lastly, the third category of solutions is not primarily technology-based, but has at least architectural consequences or makes intensive use of technology. The solutions will be discussed in the next sections.

Multi-party computation

The use of MPC and its opportunities has been researched before in a data marketplace-context (e.g. Agahari et al., 2021). Several experts mentioned this solution as well during the interviews when discussing how data sovereignty challenges could be addressed. For example, when asked how technical enforcement of data contract could be executed, E7 mentioned that

MPC could be useful for usage control, but also emphasised that MPC still needs additional access control:

> *"Yes, of course you have multi-party computing methods that are coming. That is specific to the usage control side, the technical enforce-ability of usage control. But you still need access control for that."*
>
> Expert 7

Data anonymisation

The use of data anonymisation techniques was frequently mentioned by experts as well. These experts mainly linked these methodologies to address the challenge of sensitive business data, especially personal data. The goal of these technologies is modifying the data in a manner that the actual raw data that is provided is at least modified to an extend that the data consumer can still sensibly use it, but that the potential risks for data providers are lowered. For example, consider the scenario that a data leakage arises. This could have very severe impacts for the data provider, but the business critical information (at the data attribute) level is at least partially secured. Regarding privacy-sensitive data, anonimisation could be a boundary condition to even be able to share data, for example due to privacy-regulation conditions.

> *"Plus also like if the data provider also will have also the control on all of their data. So they will see like for example, I can share some of that. I will not share the raw data itself. I would share some kind of anonimised data."*
>
> E11

Analysis at the source

Several experts argued that another solution could be to not share the actual data itself, but to only return the actual insight that the data consumer is trying to get. For example:

"For now, I would imagine a technical solution, maybe, where the consumers don't get the actual data, but they and the algorithms go where data is. The algorithms go where the data is."

E10

"The application could also maybe be installed on the consumer side, but it will not lead to leaking of the data, but it will give you only the required output. So this could be also done."

Expert 11

However, one expert warned for the risk that outputs can sometimes be traced back to the actual data:

"So, for example, if we take machine learning and artificial intelligence were used where research institutions thought they protected their source data reasonably well by just giving access to the end result to the model, but by a really smart way the attackers were able to refer back to the original data by just the answers. By querying answers smartly."

Expert 9

What the quote by E9 illustrates is that when adopting certain technological solutions to address data sovereignty challenges, there still has to be a mix of other controls. The strength regarding data sovereignty of the complete DMMP-ecosystem is dependent on the weakest link.

6-5-5 The need for use cases

Before discussing the actual questions related to data sovereignty in a DMMP-context. The experts very frequently sketched a picture of the current state of data sharing, which is far from a functional DMMP, with even very few operational data marketplaces. During the conversations, use cases were mentioned when asked how this development of data sharing could progress. According to the experts, use cases have several advantages. Firstly, it can help to present the possibilities and potential that data sharing can have to solve specific problems. More specifically, as use cases are more specific regarding the problem they try to solve, the value that could be created gets more visible. This value could relate to cost savings or other financial goals, but also for example societal or environmental value.

Secondly, use cases help to provide more context that also helps to solve the technical unknowns, for example which data model. Another example could be to be able to determine

which data quality is necessary in the specific use case to fulfil the needs of participants to achieve the mutual goal:

> *"Data quality is always to me something use case related. So data quality is nothing that's absolutely and has to be. So it's very hard to rate quality because if you have some data and you want to ask: how many of my customers live within 50 kilometres? Then it's very important to have the address. If you ask: how many of my customers are over 30? It's completely irrelevant if their address is in there or not. So data quality is something use case specific and maybe you want to tackle this with the platform."*
>
> Expert 4

Please mention the role that the DMMP could have in solving this issue according to E4. This could for example mean that the DMMP acts as a body to develop standards that data providers and consumers can use for their data exchange. However, this immediately becomes difficult given the large variety of players, industries, regions and types of data a DMMP has to service. Somewhere there must be a middle ground that is still flexible, but specific enough to be workable. For example:

> *"Certain things can be done generically, others specifically. An example: in the logistics sector agreements had to be made about standards. Then it turned out that a certain sub-sector already had a set of specific standards. In this particular example, it makes sense that that sub-sector would continue to use those standards, but that's irrelevant for the entire logistics industry. Then it is important to see which part is relevant for all parties within that sector."*
>
> Expert 1

> *"For parties and their own data spaces, they can make their own additions independently, for example to make agreements in a certain domain about the quality of data."*
>
> Expert 7

6-6 Overview of solutions and linking to challenges

Section 6-5 has provided an overview of the expert views on potential solutions to address the data sovereignty-related challenges in a DMMP-context. Solutions are divided in four categories. The first category includes the architectural solutions, these solutions have an impact at a very high-level that can affect the broader operation of the platform. More

Architectural solutions	**Sub-solutions**
Limit centralised functionality	Only centralise data descriptions, only provide standardised contracts, only fulfil data brokering function, only as advisory body for data providers (e.g. DM selection guidance), only as independent trust provider (e.g. auditing consumers/providers), mixed
Shared decision-making	Foundation, association, federation, cooperation
Shared ownership	Shared stake & risk
Trust-building solutions	
Granting autonomy to data providers	Fine-grained data sharing condition setting by data provider. Trade-off with increased divergence
Identity management	Identity management, technical standards for identity
Certification	Of providers, of consumers, of complementors, of DMMP(-operator), certification before entry, periodic certification before participant license renewal.
Reputation-building	Third-party assurance (by external auditor), communicate, community-building, Track record of fair dealing.
Trust-sustaining solutions	
Monitoring	To verify usage control compliance, illegal copying tracking, supported by labelling, tagging of data assets
Licensing model	Adopting of licensing model
Technical enforcement	Automated notification of non-compliance, enabled by monitoring tools, digital contracts, smart contracts, data tagging/labelling.
Enabling technologies	
MPC	(usage control)
Anonymisation	Pseudonimisation, anonymisation, aggregation, synthetic/digital twin, differential proviacy
Analysis at the source	direct acces control as no direct access is granted, indirect usage control
Developing use cases	Cross-industry use cases, geographically-spread use cases, mixed

Table 6-2: Proposed solutions to address data sovereignty-related challenges.

specifically, architectural choices limit possibilities at lower levels of the DMMP. Secondly, a category of solutions that can help to build up trust on the DMMP is included. Thirdly, trust needs not only provisions to be build, it also needs repeated evidence among participants that the DMMP is indeed trustworthy. Solutions to sustain trust are included in this third category. Lastly, the fourth category includes all solutions that are of more technical nature, but can be supporting during the realisation of the solutions in the first three categories. For example, the use of data tagging could be helpful to build up a monitoring system to create data usage lineages. This overview is presented in table 6-2.

The remainder of this chapter will discuss data sovereignty related challenges and the options for addressing them with the set of potential solutions.

6-6-1 Addressing fears for DMMP-dominance

The fears for DMMP-dominance were frequently expressed by experts, and the proposed solutions were relatively diverse. However, most of the proposed solutions were either focused on reducing the power of the DMMP and its operator or on improving the autonomy for data providers. Firstly, an architectural solution will probably be needed first to ensure that data

providers can influence decision-making at the level of the DMMP-operator. Furthermore, this could be combined with limiting the centralised functionality of the DMMP, for example to offer data brokering services primarily at the DMMP, but keep other aspects such as the actual exchange of data decentralised. Furthermore, there has to be autonomy for data provider at lower levels, especially at the level of the relation between data provider and data consumer, as well. This means for example that data providers need to be able to be autonomous in setting their conditions at a fine-grained level. Adopting an architectural solution, combined with high levels of autonomy for data provider can in the longer term build up the reputation of fair dealing. However, the DMMP will also need to prove that its internal organisation and infrastructure is trustworthy. This could for example be achieved by audits by third parties.

6-6-2 Data providers: knowledge building and the value of communities

As indicated by the experts in the interviews, several organisation still have internal struggles regarding their data sharing capabilities, such as conducting an internal data inventory to assess potential data assets that could be shared or marketed. Although there are external factors that could drive the capability building of potential data providers (e.g. upcoming EC proposals to increase access to data and data sharing), building up communities of parties that are already sharing data or want to share data can help to improve the industry understanding of their needs regarding data sovereignty. Furthermore, specific use cases could help to make potential value of data sharing more clear. The development of these use cases could be done within these communities as well. Lastly, a strong community of DMMP-users (both data providers, data consumers and complementors) could also help to counterbalance DMMP-power and could help to further develop and articulate needs from the user base.

Furthermore, the proprietary data assets that are often most valuable require that data providers can support on mechanisms that offer the increased level of control necessary to retain data sovereignty. For this type of data, DMMPs have to offer a high level of decentralisation, where data is kept at the source of the data provider. Furthermore, the identity and credentials of potential data consumers need high levels of verification, which applies to the data marketplace that the data consumer is using as well. Additional usage of technologies such as MPC (to control the usage) and anonymisation (especially for personal data) will be necessary add an increased level of control over these data. This means, on top of a trusted and secure infrastructure with certified parties which can be considered a boundary condition.

6-6-3 Granting access: identity, proximity and supporting technical enablers

Firstly, based on the expert responses, before granting access, the main key question for data providers is about the identity and credentials of the data consumer. As a result, a DMMP has to include both high-level certification provisions (e.g. to enhance quality and trustworthiness

of new entrants) with a periodic renewal (to retain quality and trustworthiness of existing participants), but also needs to offer fine-grained access control at the level of an individual transaction. This could for example mean that data providers can approve only specific individuals within the consumer's organisation to access.

Additionally, improved interaction between users in the form of a community can address the lack of proximity that is anticipated by some experts. However, due to the potential scale of DMMPs, communities can not be considered as a replacement of the provisions mentioned earlier in this section, it can function as a complementary mechanisms at best to address fears for granting access.

Lastly, the DMMPs could offer several technology-based mechanisms to offer data providers the a technical restriction on data access. For example, the potential solution to bring the algorithm of the consumer to the data at the source of the provider, compared to the other way around. Please note that this provision indirectly impact data storage control as well. Furthermore, anonymisation-techniques do not directly control data access (for example access to the source of data), but because they adapt the actual content of the data, they can hide the specific under-laying data. In this way, they can lower the data access control requirements.

6-6-4 Data usage and storage: inter-twined and in need of enhanced visibility

Firstly, the interviews and review of literature (both from industry and academic sources), provide several potential mechanisms to restrict data usage. In addition to the current legal approach of usage agreements in the form of a contract, new technologies provide more direct usage control as well, for example MPC. Furthermore, bringing the analysis or algorithm to the data can address the challenge that data provider are hesitant to open their data to external storage facilities.

However, the conversations with experts made clear as well that usage and storage control is not always the direct need to control the usage and storage, but to at least know as a data provider what is happening. As a result, the data usage and storage challenge is about monitoring as well. This means that a DMMP has to offer a secure (for all participants) and transparent (at least for the data provider) infrastructure where usage and storage of data can be monitored. The use of data tagging/labelling could be a supporting technology to help implementing these systems.

6-6-5 Data ownership: arrange ex-ante, retain ex-post

During the research, it became clear that data ownership is difficult from a legal perspective, although upcoming proposals by the EC try to provide improved legal basis for data sharing contexts. Furthermore, data are non-tangible assets and highly replicable without additional

marginal costs (Hart, 2002). As a result, even if it would be able to define data ownership, it would be difficult to actually retain it by the party that considers itself the data owner.

However, during the conversations with experts, two distinct perspectives on data ownership became apparent. Firstly, a group of experts considered data ownership as what follows from control over data access, data usage and data storage. Please note that this perspective highly correlates with data sovereignty. Secondly, a group of experts used a more practical view for the specific data sharing context. Experts considered data ownership as a pre-condition to be able to share data as a provider (both for data marketplace- and DMMP-context. This means that when data providers offer data, they are responsible for being the data owner, including potential liability if it turns out that their data contains errors or is belonging to another party. Several experts added a further refinement to this perspective on data-ownership, namely that data owners should be entitled parties. This translates to being the party that is owning the rights to the data and has the ability to grant access rights to others.

Furthermore, the issue of losing data ownership in a DMMP-context was discussed. In short, experts emphasised that licensing of data is a first starting point to arrange the relation between the entitled party and the data consumer. Additional conditions can be included in this license to arrange all aspects of the relation between data provider and data consumer, primarily related to data access and potentially data usage as well. It can also include penalties for misbehaviour on behalf of both parties in the relationship. However, experts also questioned whether licensing alone would be sufficient in itself, considering the large number of transactions that could theoretically take place on a DMMP. As a result, several experts mentioned additional technology-based solutions that could enhance data usage and storage monitoring, primarily to help data providers keep oversight of actual behaviour to compare it with agreements.

This issue also relates to the challenge of disputes mentioned by the experts. Similar to data ownership, a legal basis is a starting point for resolving disputes. However, as legal procedures to resolve disputes often come with legal costs, this can lead to a situation where data providers that are faced with non-compliance by data consumers (or complementors, data marketplaces), have to make a trade-off between the potential recaptured benefits and legal fees. Because this situation can still lead to data providers that decide to not take further steps, several experts mentioned that additional provisions such as data tagging and technical enforce-ability of data contracts could potentially be more effective in practice.

6-7 Chapter conclusion

This chapter started to discuss the data sovereignty related challenges that can arise in a DMMP-context according to the experts. Several challenges were identified at three levels: 1) DMMP-ecosystem, 2) data provider, 3) data transaction. The fears for DMMP-dominance were emphasised by many experts, with vendor lock-in that could lead to a sub-optimal

situation for data providers as primary reason. At the data provider level, several experts mentioned the lack of data sharing capabilities, where larger organisations tend to be more mature. Furthermore, capabilities differ substantively between different industry sectors as well. At the level of the data transaction, discussions with experts indicated that fears to grant access are still a major issue among data providers. More specifically, data providers want manage access control autonomously and at a very fine-grained level. Data usage and data storage appeared to be very inter-twined, and the main challenge regarding these two antecedents is the visibility over them. This includes for example the challenge that data can be easily copied and taken out of sight for data providers. Lastly, due to the large scale and increased complexity of DMMPs, there could arise more disputes, and resolving them via legal enforcement could be difficult.

The second half of this chapter presented the findings from the expert interviews to answer sub-question 4: *"Which potential solutions can address the data sovereignty related challenges identified for the DMMP context?"*

The fears for DMMP-dominance could be addressed by reducing the centralised functionality of the DMMP and/or by adapting the decision making and ownership structure. However, these are architectural solutions that can have a very large impact on other components of DMMPs, and can reduce for example the number of additional service offerings that depend on a centralised structure. Trust was also a very frequently mentioned topic that could address the challenges. However, as there are currently no operational DMMPs yet, there is no proven track record and this requires additional provisions such as certification and intense collaboration with potential users to build up trust. Furthermore, trust has to be sustained over time and this requires that the DMMP has to prove its trustworthiness repetitively over time, for example by certification or audit of the platform itself. Additionally, technology-based solutions have to support solutions at higher levels, for example the usage of data tagging and labelling to improve data usage and storage monitoring. These technology-solutions help to manage the scale of DMMPs.

7

Discussion & Conclusion

This chapter will firstly discuss the findings of the research in section 7-1. Subsequently, section 7-2 will discuss the answers to the sub-questions and main research question. Thirdly, contributions from both a theoretical and practical perspective will be discussed. Thirdly, the limitations of this research project will be discussed in section 7-4 and recommendations for future research in section 7-5.

7-1 Discussion of findings

The main objective of this research was to find mechanisms to enhance data sovereignty for data providers by identifying arising challenges and potential solutions in a DMMP-context. To build further on existing research related to data marketplaces and initiatives related to data spaces, a baseline was established. More specifically, a review of several data sharing initiatives (both data spaces and data marketplaces) further refined with empirical insights from experts. When comparing the current state to DMMPs, challenges at three levels were identified: 1) DMMP-ecosystem, 2) data provider and 3) data transaction. These levels were used to cut through at least a part of the complexity that comes with data sovereignty in a DMMP-context. Whereas the initial premise was that the main challenges would arise at the data transaction level, the empirical findings indicate that this level cannot be seen independently from the both the larger DMMP-ecosystem level and the level of the data provider. At the level of the DMMP-ecosystem, the fears for DMMP-dominance were mentioned frequently, in contrast to what was expected based on existing literature. The lack of data sharing capabilities on the level of data providers that currently leads to difficulties to estimate control requirements and value of data was mainly derived from the empirical part of the research as well.

Although there was a preliminary understanding of data sovereignty in a data marketplace-context that was based on analysis of existing literature, the expert interviews resulted in several additional findings. Firstly, the empirical findings indicate that especially control over data access and data usage are both complex, but also critical for data sovereignty. Several potential solutions were identified to address these antecedents, such as processes to negotiate data agreements between data providers and data consumers, and which role each party can have. Additionally, the use of licensing was identified, where the licensing agreement could both be used to formalise agreements regarding data access and data usage, but also to protect data ownership legally. However, the empirical findings also included the repetitive attention for technologies in support of the realisation of higher level agreements. These technologies could be used for direct of control of data access (e.g. bringing the algorithm to the data, anonymisation) and data usage (e.g. MPC), but also to indirectly make the process of monitoring data flows more efficiently at scale (e.g. labelling/tagging of data). Furthermore, the challenge of disputes further supplemented the understanding of data sovereignty in a DMMP-context, and made clear that managing disputes at the scale of a DMMP needs well-designed mechanisms.

7-2 Conclusion

This section recaptures the main research question and sub-questions and provides the answers to them based on the findings from this research projects:

Sub-RQ1: What are the antecedents of data sovereignty in the context of a data marketplace?

Based on the literature review, four antecedents of data sovereignty were identified: data ownership, data access, data usage and data storage. Whereas data marketplaces can create value by connecting market sides and offering those a digital platform to exchange data against a compensation, the good that is traded (i.e. data) is intangible and can be copied with barely marginal costs. As a result, data providers can lose ownership when offering their data at a marketplace. For example, as unauthorised parties can get access to their data, or because the data consumer is using the acquired data for different purposes than agreed on.

Sub-RQ2: What are the current strategies to enhance data sovereignty in data marketplaces and data sharing initiatives?

Experts identified data access and data usage as the most critical antecedents of data sovereignty. Decentralised data sharing, where the raw data is kept at the source and descriptions of the data (metadata) can be kept more centrally, was found to be one of the first solutions to address data sovereignty challenges. Furthermore, one of the findings of the expert interviews was the legal approach that is currently common to keep data providers in control. However,

several experts indicated that technical enforcement of data sharing agreements would become more important and common in the future. Additionally, they shared current solutions to increase control for data providers regarding data access and data usage. Currently, this entails for example that data providers make a relatively specific request, which data providers can accept, reject or negotiate further. Furthermore, much emphasis in current practice is placed on increasing trust among data sharing participants. This does not only entail trust in the other party, but also to the larger data sharing platform and infrastructure. One of the strategies to increase trust is the use of certifications for participating parties, which provides a level of assurance for new and existing participants. These certifications do not only apply to the organisations of participants, but to the technical components that are used to share data as well.

Sub-RQ3: How do the challenges of SQ2 translate into a DMMP-context and what additional data sovereignty related challenges can arise?

Analysis of these expert insights regarding challenges resulted in three areas of challenges: 1) the DMMP-ecosystem level, 2) the data provider level and 3) the data transaction level. Regarding the first level, the main anticipated challenge were the fears for dominance of the platform. This included for example fears for aggressive business models including high fees and vendor lock-in concerns. Additionally, analysis of the expert interviews resulted in several areas of increased complexity, for example the large amount of different data models that is used across industries. At the data provider level, the main anticipated challenges related to a lack of data sharing capabilities at organisations, which can result in difficulties for data providers to set controls effectively. Close related to the lack of data sharing capabilities is the value estimation challenge: remaining sovereign as a data provider is crucial, but only when data providers capture the real value. Lastly, data providers have several different types of data assets, including very sensitive business critical data. At the transaction level, i.e. the relation between data provider and data consumer, granting access is the first challenge. Specifically, the fear to grant access to a party that is not trustworthy, or is a business competitor, was identified by the experts as one of the challenges for a DMMP-context. One of the under-laying problems that results in this challenge, is the decreased proximity in a DMMP-context. Regarding data usage and storage, analysis of the interviews clarified the close relationship these have. Experts frequently linked data storage back to the fears for dominance, where they argued that storage should stay at the data provider. Lastly, the challenge of disputes was identified, especially considering the larger scale of DMMPs, many experts feared that solving disputes (i.e. reacting to breach of agreements or conditions) could become problematic. This challenge is relevant for data providers, as staying in control is not only about the the period before and during a transaction, but also about the scenario when a data consumer breaches agreements that were made.

Sub-RQ4: Which potential solutions can address the data sovereignty related challenges identified for the DMMP-context?

The expert interviews yielded potential solutions to address data sovereignty challenges in a DMMP-context. Firstly, architectural solutions were identified which include limiting the centralised functionality of a DMMP, for example by only bundle metadata centrally or by using DMMPs primarily as a platform were participants (both data providers and data marketplaces) can get certified. Furthermore, decision-making in close collaboration with DMMP users can help to mitigate the fears for dominance. Note that this is a practice which is currently adopted by many data sharing initiatives already. A further step could be to adopt a shared ownership structure where all participants have a stake, instead of the central organisation operating the DMMP.

Furthermore, analysis of the interviews resulted in several solutions that could help to *build* trust as well: 1) data provider autonomy, 2) identity management, 3) certification, 4) reputation-building. Specifically for DMMPs, managing identities will be much more complex, considering the different data marketplaces connected to the platform and data providers from different industry sectors with varying standards. In addition to *building* trust, governance mechanisms to *sustain* trust in the longer term were identified as well:

One of the key components of this category is the use of monitoring. This monitoring (primarily over data usage and storage) could assist data providers in retaining oversight over their data assets over the DMMP. Monitoring could be supplemented by active tagging and labelling of data. This labelling was identified to partly address data ownership concerns as well. The labels can help to keep track of the ultimate owner (i.e. the data provider that originally supplied the data). Regarding data ownership, which is partly addressed by data labelling, many experts indicated the use of data licensing to contractually keep data ownership at the data provider. In a licensing model, the data consumer is not buying data, but is buying access right under pre-agreed conditions. However, even after agreeing on conditions, data providers still have to be sure that these agreements are met. As a result, experts indicated the use of technical enforcement as a potential solution, to supplement legal enforcement.

More specifically, these technical solutions included MPC as a technical enforcement of usage control: the analysis is pre-agreed and as the consumer cannot access the actual data but only the processed output (i.e. the actual insights), risks of disclosing the raw data are mitigated. Anonymisation of data could lower the risk of unwanted disclosure of data, which could potentially lower data access requirements. Furthermore, several experts mentioned the solution to bring the algorithm for the analysis to the data, which relates back to the DMMP architecture as well. Lastly, during the expert interviews, the role of use cases was emphasised frequently. Use cases can both help to make value of data sharing more visible ex-ante, and they can help to make the data sharing application more specific.

RQ: What governance mechanisms can enhance data sovereignty for business-to-business data sharing via DMMPs?

Firstly, this research indicates that DMMPs both amplify challenges that are already existing in non-DMMP contexts, such as data marketplaces and data spaces. For example, the use of legal enforcement that becomes less manageable at a DMMP-scale. Based on the interview findings, DMMPs have to carefully balance their own power with data providers. They need to install mechanisms that enable data providers to be involved with decision-making at the DMMP ecosystem level, but also ensure autonomy for data providers at the transaction level. Compared to data marketplaces, there is more need for DMMPs involve data providers and create a democratic decision making process. However, this requires building of capabilities by data providers. Compared to current data sharing initiatives and data marketplaces where activity is often driven by shared goals in an industry, DMMPs need to actively offer resources for (potential) data providers to help them with data inventory and value estimation. Additionally, a competent user base of data providers will also provide a counter-balance to the central DMMP-operator. Furthermore, due to the scale of DMMPs and the more diverse data assets that can be traded, there is a risk of losing proximity between market sides. This proximity has to be actively enhanced by identity management, which has to account for more different industry players compared to single data marketplaces. Furthermore, DMMPs have to offer their data providers the information they need to make sure that they can actually compare data provider behaviour with agreements and conditions. At the data transaction level, technology-based solutions could be of value both for direct access and usage control (i.e. MPC and anonymisation techniques), but also to enhance data provenance by use of data monitoring and labelling. Lastly, data ownership has to be arranged in a manner that is both feasible, but also ensures data providers that they retain ownership. This requires that before data providers supply data, they declare that they are rightful owners. Next, data licensing can legally retain ownership at data providers, but considering the scale of DMMPs, should be supplemented by tagging and labelling. These technologies can reduce dependence on legal enforcement, and can help to technically monitor data assets for data providers.

7-3 Contributions

This section will discuss the contributions of the research project. Firstly, it will discuss the theoretical contributions to the academic literature. Secondly, practical contributions will be discussed that could be of value for practitioners.

7-3-1 Theoretical

This section will discuss theoretical contributions of this research project on data sovereignty for data providers in a DMMP-context. Firstly, by departing from existing digital platform and platform governance literature, this study has added an understanding of how governance mechanisms can be used when the goal (i.e. enhancing data sovereignty) of the mechanisms is different than increasing power by the platform-operator, a topic which has already received a considerable amounts of attention from academic research. This research has tried to research

governance mechanisms when the goal (i.e. enhancing data sovereignty) is different compared to primarily increasing financial gains for the platform owner, a domain which has received limited attention according to Mukhopadhyay and Bouwman (2019, p. 346).

Additionally, by studying the relatively novel DMMP-context compared to the data marketplace-context which has already received more academic attention, this research has tried to contribute to digital marketplace literature, more specifically by researching a platform that can address fragmentation of the data sharing landscape which can lead to data discovery issues and lack of trust (Simon & Natalia, 2021). Furthermore, the empirical findings related to separating data and metadata and limiting DMMPs to brokering platforms are in line with the work by Lawrenz et al. (2019) that argues for the importance of metadata for data marketplaces. Furthermore, Lauf et al. (2022, p.12) call for future research "on examining the areas of tension in-depth to elaborate more concrete solution approaches to develop data ecosystems perceived as fair from the viewpoint of individuals and companies." Although DMMPs are an instance of a data ecosystem, but not all data ecosystems are DMMPs, this research can still be considered as a contribution in the direction of this call for research, specifically because data sovereignty related challenges and potential solutions were explored.

The use of MPC as an enabling technology to technically enforce usage and access control identified empirically from the expert interviews can be considered as a confirmation of the proposition by Agahari et al. (2021) that MPC can become an important enabler for data sharing via data marketplaces. However, this research mainly identified MPC in a DMMP-context.

7-3-2 Practical

For practitioners, the findings of this research for data sovereignty in a DMMP-context can be used to better understand data provider's needs in this potential future context. However, these insights can also be valuable to improve existing data sharing initiatives. For example to better understand the role of user communities to provide input and feedback for the central organisation of data sharing initiatives. Furthermore, this research can provide insights on the impact of architectural decisions on lower level user possibilities. It can help to formulate answers such as: which functionalities are valuable for data providers to offer centrally? And which ones have to be decentralised? How can new technologies help to better manage my data sharing network scale? What is the impact of offering technical enforcement of data contracts on perceived trust for data providers?

Furthermore, it this research provides an actual perspective by experts on both the current data sharing landscape, and how they see future directions. Additionally, by analysing data sovereignty challenges at the level of the internal organisation of the data provider as well, practitioners can develop strategies to attract new participants more effectively, by actively helping to develop data sharing capabilities initially, but also by including them in active use case development and user feedback process implementation. The analysis based on different

levels of challenges in itself could help to better separate the different actor relations and to implement improvements targeted at specific relations.

Thirdly, the perspectives regarding DMMP-dominance could provide considerations for data sharing initiatives that want to adopt a more centralised model, or that want to implement a commercial business model. Furthermore, the identified challenges and solutions regarding data transactions for the DMMP-context that was studied can provide directions to further enhance trust for existing users as well, especially for data sharing initiatives that are in the scaling phase. Lastly, the centralised DMMP-configuration that was discussed with experts mainly revealed perceived challenges regarding data sovereignty. However, for practitioners, these identified pain points can help to make better decisions when moving towards a more centralised configuration. Contrarily, it can help centralised data sharing initiatives to find pathways where decentralising can be impactful when the goal is to enhance data sovereignty.

7-4 Limitations

However, this research project also comes with several limitations. Firstly, the role of data provider and data consumer in this project was chosen as rather static, whereas expert insights have indicated that these roles can be much more dynamic. Additionally, the findings from this research project where acquired by eleven expert interviews using the same semi-structured research protocol. This means, the eventual findings were not discussed in additional validation sessions or workshops. As a result, it is not possible to say if a different group of experts would validate these, and if they would have indicated at the same data sovereignty related challenges and solutions.

Additionally, the experts that were interviewed are not belonging to organisations that could be data providers themselves. However, as a DMMP-context is already novel compared to data marketplaces, for this research project the priority was given to experts with data sharing, data space and data marketplace experience. As a result, there could be a bias that is more favourable to this context, as interviewees already have personal experience with existing data sharing contexts as most interviewees are working for one of them, or organisations that contribute to them by conducting research. Furthermore, except for two interviewees, all respondents were residing within Europe (among roughly half of them in the Netherlands), this could potentially have led to a more European perspective regarding the perspectives on platform dominance, data practices and the role of governments.

Lastly, the focus on data providers was considered to be justified, as the primary goals of the research project was to enhance data sovereignty that is primary an issue for data providers. Although this single perspective could have led to richer and more in-depth findings regarding the research problem, it could have resulted in solutions that are preferable for data providers, but that could work less well for the larger platform ecosystem.

7-5 Future research

As a result of the research limitations, there are several pathways for future research. Firstly, a research study could be conducted to verify the results of this project. For example by selecting a similar panel of experts and to deduct whether these experts indicate at the same challenges and solutions. Furthermore, research focusing on the perspectives of other parties in the DMMP-ecosystem could contribute to a better understanding of the whole picture. For example, how do data marketplaces respond to DMMPs: do they consider DMMPs to be an opportunity, as extending their functionality could transform them into a meta-platform themself? Are they threatened by DMMPs, because it could enable data providers to multi-home? Furthermore, qualitative research could also lead to a better understanding of the behaviour of complementors in a DMMP-context: do they remain loyal to data marketplaces that they already have a relationship with? Or are complementors aiming at the more universal DMMPs? Additionally, case studies at organisations that are considering to start sharing data using data marketplaces could lead to a better understanding of how these organisations build up their capabilities, and which resources they mobilise. Additionally, the role of data providers in communities could be further studied to see its impact on limiting DMMP-operator dominance. Lastly, more technically-centred research could further study how DMMPs achieve inter-operability with data marketplaces, as this is a boundary condition to make DMMPs work. This of course depends on the view by data marketplaces on DMMPs as well, i.e. will they even be willing to open up to DMMPs in the first place?

7-6 Personal reflection

First of all, when discovering this topic at the list of graduation topics, my interest was sparked immediately. Digital platforms have always interested me, especially their dynamic character and their increasing role in everyday life. Linking these digital platforms to trade a non-tangible, duplicable good called data made me even more curious. Furthermore, I found it interesting to research data sovereignty not only from an academic perspective, but also because it is one of the obstacles that is still challenges in the broader data economy. And by the way, this does not only to the business-to-business setting I researched, but relates in the end back to every individual. Can you think of all the parties that are collecting data about you? And where they store it? And after what period they remove these data? Right.

However, during this research project I also faced some difficulties. First of all, due to the exploratory and inductive research approach, I was not able to develop a structure in advance and had to derive this from the data. In general I prefer to have a structure to adhere to, which made it difficult for me to make progress regarding the report. When this procrastination lead to delays, I was disappointed in my own performance. Secondly, initially it was difficult to find interviewees for my project, which was demotivating. However, I am glad that I have grabbed all opportunities to find participants, which included a spontaneous trip to a German

industry fair. Additionally, I am especially grateful to one of my committee members for the access to his network as well when trying to find interviewees.

All in all, this project has been a tremendous learning period. Not only content-wise, but also regarding the development of interview skills. The conversations with the experts increased my skills to grasp the interesting aspects of particular statements and to ask an effective counter question. These experiences have not only helped me during this project, but will also be valuable for my future career.

7-7 Link to MoT program

The aim of the Management of Technology master program is to educate students to understand technology as a resource that can be used strategically to help fulfilling the goals of an organisation. This requires strategic decision-making by the focal firm, for example regarding development processes of new products and services. As modern organisations are highly inter-twined with partners, this is often a decision with many stakeholders involved. More specifically, there is a substantial people-element involved when organisations want to use or develop technology to achieve business goals. This master thesis project includes several of those elements. Due to increased use of modern equipment with sensors, smart wearables and an increasingly connected world, vast amounts of data is generated by the hour. However, storing and using data solely within a specific organisation is no longer possible in modern business ecosystems. Firm are increasingly dependent on external data sources both for daily operation and development of new products and services.

This master thesis project included several of these elements. Firstly, data sharing is an effort that includes, human decision-making, technical components and different business goals across parties involved. Making such a system work is a delicate balancing act where every participant has to be able to get a share in the value that is created. Furthermore, decision made at an organisational level impact technical possibilities and vice versa. Secondly, considering the current state of data sharing, an operational DMMP is still far away. As a result, there is still a lot of uncertainty. This study has tried to bridge between both technical and non-technical challenges and opportunities and this uncertainty to contribute to a data sovereign data sharing future.

The courses offered in the MoT-program were very valuable while carrying out this research project. Whereas MOT2004 Preparation for Master Thesis brought hands-on experience in conducting and reporting on a literature review to find a research gap, MOT2312 provided a solid basis regarding research methods. Specifically for this research topic, MOT1435 Technology Strategy & Entrepreneurship helped to understand the different processes related to standardisation. Furthermore, this course provided theoretical background related to first-movers and why the outcomes of this strategy can differ. Data sharing is still a new concept in practice and also still lacks a lot of standards and organisations that are hesitant to start

sharing data. Secondly, MOT2421 Emerging and Breakthrough Technologies contributed to the understanding which patterns innovations often show and to relate this to my research context, DMMPs, and current practice which is still relatively immature.

References

Aaronson, S. A. (2021). *Data is disruptive: How data sovereignty is challenging data governance*. Hinrich Foundation.

Abbas, A. E. (2021). Designing data governance mechanisms for data marketplace meta-platforms. *34th Bled eConference Digital Support from Crisis to Progressive Change: Conference Proceedings*, 691–703. https://doi.org/10.18690/978-961-286-485-9.49

Abbas, A. E., Agahari, W., van de Ven, M., Zuiderwijk, A., & de Reuver, M. (2021). Business data sharing through data marketplaces: A systematic literature review. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*, 3321–3339. https://doi.org/10.3390/jtaer16070180

Agahari, W., Dolci, R., & de Reuver, M. (2021). Business model implications of privacy-preserving technologies in data marketplaces: The case of multi-party computation. *29th European Conference on Information Systems (ECIS 2021): Human Values Crisis in a Digitizing World*, 1–16. https://aisel.aisnet.org/ecis2021%7B%5C_%7Drp/59

Asswad, J., & Gómez, J. M. (2021). Data ownership: A survey. *Information*, *12*, 465. https://doi.org/10.3390/info12110465

Baloup, J., Bayamlıoğlu, E., Benmayor, A., Ducuing, C., Dutkiewicz, L., Lalova, T., Miadzvetskaya, Y., & Peeters, B. (2021). White Paper on the Data Governance Act. *SSRN Electronic Journal*, (June). https://doi.org/10.2139/ssrn.3872703

Boudreau, K. J., & Hagiu, A. (2008). Platform rules: Multi-sided platforms as regulators. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1269966

Cennamo, C. (2018). Building the value of next-generation platforms: The paradox of diminishing returns. *Journal of Management*, *44*, 3038–3069. https://doi.org/10.1177/0149206316658350

Chen, L., Tong, T. W., Tang, S., & Han, N. (2022). Governance and design of digital platforms: A review and future research directions on a meta-organization. *Journal of Management*, *48*, 147–184. https://doi.org/10.1177/01492063211045023

Cusumano, M. A., & Gawer, A. (2002). The elements of platform leadership. *MIT Sloan Management Review*, *43*, 51–58.

Data Sharing Coalition. (2021). *Data Sharing Canvas* (tech. rep. April). https://datasharingcoalition.eu/app/uploads/2021/04/data-sharing-canvas-30-04-2021.pdf

de Reuver, M., Sørensen, C., & Basole, R. C. (2018). The digital platform: A research agenda. *Journal of Information Technology*, *33*, 124–135. https://doi.org/10.1057/s41265-016-0033-3

Eaton, B., Elaluf-Calderwood, S., Sørensen, C., & Yoo, Y. (2015). Distributed Tuning of Boundary Resources: The Case of Apple's iOS Service System. *MIS Quarterly*, *39*(1), 217–243. https://doi.org/10.25300/MISQ/2015/39.1.10

Eisenhardt, K. M. (1985). Control: Organizational and Economic Approaches. *Management Science*, *31*(2), 134–149. https://doi.org/10.1287/mnsc.31.2.134

Eisenmann, T., Parker, G., & Alstyne, M. W. V. (2006). Strategies for two-sided markets. *Harvard Business Review*, *84*, 92–101. https://ssrn.com/abstract=2409276

Floetgen, R. J., Strauss, J., Weking, J., Hein, A., Urmetzer, F., Böhm, M., & Krcmar, H. (2021). Introducing platform ecosystem resilience: Leveraging mobility platforms and their ecosystems for the new normal during covid-19. *European Journal of Information Systems*, *30*, 304–321. https://doi.org/10.1080/0960085X.2021.1884009

Ghazawneh, A., & Henfridsson, O. (2013). Balancing platform control and external contribution in third-party development: the boundary resources model. *Information Systems Journal*, *23*(2), 173–192. https://doi.org/10.1111/j.1365-2575.2012.00406.x

Goldbach, T., Benlian, A., & Buxmann, P. (2018). Differential effects of formal and self-control in mobile platform ecosystems: Multi-method findings on third-party developers' continuance intentions and application quality. *Information & Management*, *55*(3), 271–284. https://doi.org/10.1016/j.im.2017.07.003

Hart, D. (2002). Ownership as an Issue in Data and Information Sharing: a philosophically based review. *Australasian Journal of Information Systems*, *10*(1), 23–29. https://doi.org/10.3127/ajis.v10i1.440

Huang, L., Dou, Y., Liu, Y., Wang, J., Chen, G., Zhang, X., & Wang, R. (2021). Toward a research framework to conceptualize data as a factor of production: The data marketplace perspective. *Fundamental Research*, *1*, 586–594. https://doi.org/10.1016/j.fmre.2021.08.006

Hummel, P., Braun, M., & Dabrock, P. (2021). Own data? ethical reflections on data ownership. *Philosophy & Technology*, *34*, 545–572. https://doi.org/10.1007/s13347-020-00404-9

Iansiti, M., & Levien, R. (2004). Strategy as ecology. *Harvard business review*, *82*, 68–78, 126.

Jacobides, M. G., Cennamo, C., & Gawer, A. (2018). Towards a theory of ecosystems. *Strategic Management Journal*, *39*, 2255–2276. https://doi.org/10.1002/smj.2904

Katz, M. L., & Shapiro, C. (1985). Network externalities, competition, and compatibility. *The American Economic Review*, *75*, 424–440. https://www.jstor.org/stable/1814809

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

Koutroumpis, P., Leiponen, A., & Thomas, L. D. W. (2017). The (unfulfilled) potential of data marketplaces, 1–42. http://pub.etla.fi/ETLA-Working-Papers-53.pdf

Koutroumpis, P., Leiponen, A., & Thomas, L. D. W. (2020). Markets for data. *Industrial and Corporate Change*, *29*, 645–660. https://doi.org/10.1093/icc/dtaa002

Kretschmer, T., Leiponen, A., Schilling, M., & Vasudeva, G. (2022). Platform ecosystems as meta-organizations: Implications for platform strategies. *Strategic Management Journal*, *43*, 405–424. https://doi.org/10.1002/smj.3250

Lass, S., & Bender, B. (2021). Dedicated data sovereignty as enabler for platform-based business models. *2nd Conference on Production Systems and Logistics*, 382–393. https://doi.org/10.15488/11299

Lauf, F., Scheider, S., Bartsch, J., Herrmann, P., & Radic, M. (2022). Linking data sovereignty and data economy: Arising areas of tension. *Wirtschaftsinformatik 2022 Proceedings*. https://aisel.aisnet.org/wi2022/it_for_development/it_for_development/19

Lawrenz, S., Sharma, P., & Rausch, A. (2019). The significant role of metadata for data marketplaces. *Proceedings of the International Conference on Dublin Core and Metadata Applications*, 95–101.

Lee, S. U., Zhu, L., & Jeffery, R. (2018). Designing data governance in platform ecosystems. *Proceedings of the 51stHawaii International Conference on System Sciences*, 5014–5023. http://hdl.handle.net/10125/50515

Mineraud, J., Mazhelis, O., Su, X., & Tarkoma, S. (2016). A gap analysis of internet-of-things platforms. *Computer Communications*, *89-90*, 5–16. https://doi.org/10.1016/j.comcom.2016.03.015

Moreau, L., Groth, P., Miles, S., Vazquez-Salceda, J., Ibbotson, J., Jiang, S., Munroe, S., Rana, O., Schreiber, A., Tan, V., & Varga, L. (2008). The provenance of electronic data. *Communications of the ACM*, *51*, 52–58. https://doi.org/10.1145/1330311.1330323

Mosterd, L., Sobota, V. C., van de Kaa, G., Ding, A. Y., & de Reuver, M. (2021). Context dependent trade-offs around platform-to-platform openness: The case of the internet of things. *Technovation*, *108*, 102331. https://doi.org/10.1016/j.technovation.2021.102331

Mukhopadhyay, S., & Bouwman, H. (2019). Orchestration and governance in digital platform ecosystems: A literature review and trends. *Digital Policy, Regulation and Governance*, *21*, 329–351. https://doi.org/10.1108/DPRG-11-2018-0067

Nederlandse AI Coalitie. (2020). *Verantwoord datadelen voor AI* (tech. rep.). Nederlandse AI Coalitie (NL AIC). https://publications.tno.nl/publication/34636511/G0AROl/ette-2020-verantwoord.pdf

Ondrus, J., Gannamaneni, A., & Lyytinen, K. (2015). The Impact of Openness on the Market Potential of Multi-Sided Platforms: A Case Study of Mobile Payment Platforms. *Journal of Information Technology*, *30*(3), 260–275. https://doi.org/10.1057/jit.2015.7

Otto, B., & Jarke, M. (2019). Designing a multi-sided data platform: Findings from the international data spaces case. *Electronic Markets*, *29*, 561–580. https://doi.org/10.1007/s12525-019-00362-x

Ouchi, W. G. (1979). A Conceptual Framework for the Design of Organizational Control Mechanisms. *Management Science*, *25*(9), 833–848. https://doi.org/10.1287/mnsc.25.9.833

Parker, G. G., & Alstyne, M. W. V. (2005). Two-sided network effects: A theory of information product design. *Management Science*, *51*, 1494–1504. https://doi.org/10.1287/mnsc.1050.0400

Peterson, Z. N. J., Gondree, M., & Beverly, R. (2011). A position paper on data sovereignty: The importance of geolocating data in the cloud. *Proceedings of the 3rd USENIX Conference on Hot Topics in Cloud Computing*, 9.

Richter, H., & Slowinski, P. R. (2019). The data sharing economy: On the emergence of new intermediaries. *IIC - International Review of Intellectual Property and Competition Law*, *50*, 4–29. https://doi.org/10.1007/s40319-018-00777-7

Schomm, F., Stahl, F., & Vossen, G. (2013). Marketplaces for data. *ACM SIGMOD Record*, *42*, 15–26. https://doi.org/10.1145/2481528.2481532

Schreieck, M., Hein, A., Wiesche, M., & Krcmar, H. (2018). The challenge of governing digital platform ecosystems. *Digital Marketplaces Unleashed*, 527–538. https://doi.org/10.1007/978-3-662-49275-8_47

Schreieck, M., Wiesche, M., & Krcmar, H. (2016). Design and governance of platform ecosystems - key concepts and issues for future research. *24th European Conference on Information Systems, ECIS 2016*. https://aisel.aisnet.org/ecis2016_rp/76

Sekaran, U., & Bougie, R. (2016). *Research methods for business : A skill-building approach* (7th). John Wiley & Sons Ltd.

Simon, & Natalia. (2021). *TRUSTS Trusted Secure Data Sharing Space* (tech. rep. No. 871481). TRUSTS. https://trusts-data.eu/

Smedlund, A., & Faghankhani, H. (2015). Platform orchestration for efficiency, development, and innovation. *2015 48th Hawaii International Conference on System Sciences*, 1380–1388. https://doi.org/10.1109/HICSS.2015.169

Spiekermann, M. (2019). Data marketplaces: Trends and monetisation of data goods. *Intereconomics*, *54*, 208–216. https://doi.org/10.1007/s10272-019-0826-z

Stahl, F., Schomm, F., Vossen, G., & Vomfell, L. (2016). A classification framework for data marketplaces. *Vietnam Journal of Computer Science*, *3*, 137–143. https://doi.org/10.1007/s40595-016-0064-2

Steinbuss, S., Resetko, A., Menz, N., & Winkel, J. (2019). *IDS Certification explained* (tech. rep. November). International Data Spaces Association. Dortmund. https://doi.org/10.5281/zenodo.5675945

Tallon, P. P. (2013). Corporate governance of big data: Perspectives on value, risk, and cost. *Computer*, *46*, 32–38. https://doi.org/10.1109/MC.2013.155

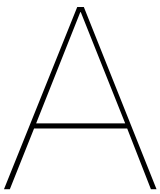
Teece, D. J. (2018). Profiting from innovation in the digital economy: Enabling technologies, standards, and licensing models in the wireless world. *Research Policy*, *47*, 1367–1387. https://doi.org/10.1016/j.respol.2017.01.015

Tiwana, A. (2014). *Platform Ecosystems*. Elsevier. https://doi.org/10.1016/C2012-0-06625-2

Tiwana, A., Konsynski, B., & Bush, A. A. (2010). Research commentary —platform evolution: Coevolution of platform architecture, governance, and environmental dynamics. *Information Systems Research*, *21*, 675–687. https://doi.org/10.1287/isre.1100.0323

van de Ven, M., Abbas, A. E., Kwee, Z., & de Reuver, M. (2021). Creating a taxonomy of business models for data marketplaces. *34th Bled eConference Digital Support from Crisis to Progressive Change: Conference Proceedings*, 309–321. https://doi.org/10.18690/978-961-286-485-9.23

Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). Research commentary —the new organizing logic of digital innovation: An agenda for information systems research. *Information Systems Research*, *21*, 724–735. https://doi.org/10.1287/isre.1100.0322



Interviewee invitation

Dear [**NAME**],

I am approaching you because I saw [**motivation for invitation]**

My name is Thomas van Velzen and I am currently studying platforms that can federate B2B data marketplaces for my master thesis research project. This master thesis project is part of my master program Management of Technology at the Delft University of Technology. I am conducting this research as part of a graduation internship at PwC Netherlands.

The title of my project is: "B2B data sharing via data marketplace meta-platforms: Exploring governance mechanisms to enhance data sovereignty." The goal is to find out which mechanisms can enhance the level of control over data for data providers. As a result, I would like to interview experts on topics such as data sovereignty, digital platforms, data sharing and data marketplaces.

Based on your [**experience**], I would be very interested to schedule an interview with you. This interview would approximately take 40-60 minutes, and I will share my findings with you once the research is finished.

I have included a one-pager with the background of my project, feel free to share it with others as well.

Would you be willing to participate in an interview?

If anything is unclear, feel free to let me know,

Thank you in advance,

Kind regards,

Thomas van Velzen

Figure A-1: Template invitation e-mail

Geachte [**NAAM**],

Ik benader u omdat ik zag dat [**motivatie voor uitnodiging**]

Mijn naam is Thomas van Velzen en op dit moment onderzoek ik platforms die B2B data marktplaatsen federen for mijn scriptie project. Dit master scriptie project is onderdeel van mijn master Management of Technology aan de Technische Universiteit Delft. Ik voer dit onderzoek uit als onderdeel van een afstudeerstage bij PwC Nederland.

Mijn project is getiteld: “B2B data sharing via data marketplace meta-platforms: Exploring governance mechanisms to enhance data sovereignty.” Het doel is om vanuit het perspectief van de partijen die data leveren aan deze platforms te onderzoeken hoe zij controle kunnen houden over deze data. Om dit onderzoek uit te voeren ga ik graag in gesprek met experts op het gebied van data soevereiniteit, digitale platformen, data delen en data marktplaatsen.

Op basis van uw [**ervaring**], zou ik erg graag met u in gesprek gaan. Het interview zal ongeveer 40 à 60 minuten duren, en als u dit wenst kan ik na afronding van het project (een samenvatting van) de scriptie met u delen.

Ik heb een Engelstalige one-pager bijgevoegd met de achtergrond van het project. Uiteraard kunt u deze ook delen met andere personen binnen uw netwerk.

Graag verneem ik van u of u open staat voor een interview.

Als u nog vragen heeft hoor ik dat natuurlijk ook graag,

Bij voorbaat dank,

Met vriendelijke groet,

Thomas van Velzen

Figure A-2: Template Dutch invitation e-mail

B

Thesis project one-pager

B2B data sharing via data marketplace meta-platforms

Exploring governance mechanisms to enhance data sovereignty

Background - Modern organizations are increasingly dependent on data for their operations, business model and new innovations. At the same time, new technology creates opportunities to generate data (for example IoT sensors in smart factories). Additionally, using data from other organizations can unlock new business opportunities as well. The **European Commission** has identified this sharing of business data as one of the key pillars of their Digital Strategy. More specifically, data sharing is one of the key components of the EU Data Act proposal that was released in February 2022.

Data marketplaces are digital platforms that enable the transfer of data as a tradable good between organizations at scale. These platforms bring data providers and data consumers together, sometimes with support of complementors that deliver additional products and services to the platform. However, there still exist **two major problems**:

Problem 1 is about the **scattered landscape of B2B data marketplaces** that exists at the moment. Some focus on specific industries, others on specific regions or countries. This results in data consumers that have difficulty finding the right marketplace and hinders these marketplaces to achieve proper scale.

Data Marketplace Meta-Platforms are platforms that link the separate data marketplaces: they are essentially platforms-of-platforms. This could be the **solution** to the problems of the scattered data marketplace landscape

Problem 2 considers the current **reluctance of organizations to provide data** to these marketplaces because they **fear the risk of losing control over their data**. In sum, there is a lack of data sovereignty. For example, what if a provider wants to withdraw data in the future? Wants to block access for competitors?

Governance mechanisms that enhance data sovereignty for B2B data sharing via Data Marketplace Meta-Platforms could help to overcome this problem of data providers

Research problem – Governance mechanisms for Data Marketplace Meta-Platforms that enhance data sovereignty have not been explored yet. > **Goal** of this thesis project

How? – Exploratory research, literature review to identify key components of data sovereignty, expert interviews to explore which requirements and mechanisms could lower barriers to providing data on these platforms

Who? – Potential interviewees would ideally have expertise regarding one or several of these topics:

- **B2B data sharing**
- **Digital platforms**
- **Data marketplaces**
- **Data sharing barriers**
- **Data sovereignty**

TUDelft pwc **Thomas van Velzen –**

Figure B-1: Thesis project one-pager

C

Informed consent form

Consent form

Study title
Business-to-Business data sharing via data marketplace meta-platforms:
Exploring governance mechanisms to enhance data sovereignty

Introduction
You are being invited to participate in a research study titled "Business-to-Business data sharing via data marketplace meta-platforms: Exploring governance mechanisms to enhance data sovereignty". This study is being done by Thomas van Velzen from the TU Delft. Furthermore, this research project is performed as part of a graduation internship at PwC NL.

Purpose of the study
The goal of this study is to explore governance mechanisms that can enhance data sovereignty for B2B data sharing via over-arching data marketplace platforms, data marketplace meta-platforms. It will take you approximately 60 minutes to complete. As part of this research project, semi-structured exploratory interviews will be conducted. The data will be used for the master thesis of the researcher. Additionally, the findings of this study including the research data can be used for publication in academic journals and conference proceedings for the duration of two years. We will be asking you to comment on the data sovereignty requirements and governance mechanisms that were identified by the literature review earlier in this research.

Processing of Personal Information
As with any online activity, the risk of a breach is always possible. To the best of our ability, your answers in this study will remain confidential. We will minimize any risks by ensuring that your identity will be protected. We will do so by making sure that only the researcher and members of the graduation committee (i.e. Antragama Ewa Abbas MSc., Dr. Geerten van de Kaa, and Dr. Anneke Zuiderwijk) have direct access to your personal information. Your identity will be protected in the final thesis report by anonymizing the participant description. Additionally, after the transcription of the interviews is finished, the recordings will be destroyed. Lastly, the interviews will be transcribed in an anonymized manner.

Rights of the participants
Your participation in this study is entirely voluntary and you can withdraw at any time. You are free to omit any questions. After the transcription of this interview, you will have the opportunity to review the transcription and provide comments or rectify them in case you feel that the transcription does not reflect the actual interview.

Contact details
Researcher: Thomas van Velzen
Telephone: +31(0)6
Email: x
University: TU Delft (data protection officer: privacy-tud@tudelft.nl)
Internship company: PwC NL

Template:
HUMAN RESEARCH ETHICS, INFORMED CONSENT TEMPLATES AND GUIDE,
Delft University of Technology, English version, January 2022

Explicit Consent points

PLEASE TICK THE APPROPRIATE BOXES	Yes	No
A: GENERAL AGREEMENT – RESEARCH GOALS, PARTICPANT TASKS AND VOLUNTARY PARTICIPATION		
1. I have read and understood the study information dated [*DD/MM/YYYY*], or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	☐	☐
2. I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	☐	☐
3. I understand that taking part in the study involves audio-recorded interviews which will be used for text-transcription and that these recordings will be destroyed immediately after transcription is finished.	☐	☐
4. I understand that the student is conducting this master thesis project as part of a graduation internship at PwC Netherlands and that this organisation provides the student with a financial compensation	☐	☐
B: POTENTIAL RISKS OF PARTICIPATING (INCLUDING DATA PROTECTION)		
5. I understand that the study will be used for the researcher's master thesis.	☐	☐
6. I understand that physical interviews come with the risk of COVID-19 and that the researcher will always maintain 1.5-meter distance and conducts a self-test in advance. Lastly, I know that I can opt for a virtual interview at any time.	☐	☐
7. I understand that personal information collected about me that can identify me, such as name, contact details, working experience, will not be shared beyond the graduation team.	☐	☐
8. I understand that after the research study the de-identified information I provide will be used for academic purposes.	☐	☐
9. I understand that for these academic purposes, the data I provide will be stored for the duration of two years after the completion of this master thesis project in the form of anonymized transcripts.	☐	☐
10. I agree that my responses, views or other input can be quoted anonymously in research outputs	☐	☐
C: ADDITIONAL RISKS FOR RESPONDENTS WORKING AT THE INTERNSHIP COMPANY		
11. *(Only applicable to interviewees working for the internship organisation)* I understand that I participate in this study to provide the student with my personal views on the student's work, but that there still is a professional risk (e.g. unintentionally using names of client organisations in my response.	☐	☐
12. *(Only applicable to interviewees working for the internship organisation)* I understand that the student will mitigate the risk mentioned on item 11 of this form by anonymising my personal information, by anonymising my response and by not disclosing the transcript outside the university research team	☐	☐

Signatures

_____	_____	_____
Name of participant	Signature	Date

I, as researcher, have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

_____	_____	_____
Thomas van Velzen	Signature	Date

Study contact details for further information:

Thomas van Velzen, x@student.tudelft.nl, +31(0)6 xxxx xxxx

D

Interview protocol

Interview protocol

1. Introduction (±10 minutes)

[Goal: study background, goals of project, informed consent recap]

Thank you for participating in this interview, which is part of my master thesis research project. My research is about over-arching data marketplace platforms that can enable businesses to share data with other businesses. More specifically, the research aims to identify which governance mechanisms can enable data providers to stay in control over their data. In other words, how could governance of these platforms enhance data sovereignty. This focus on data sovereignty from the perspective of data providers was chosen because one of the current barriers for business-to-business sharing at scale via data sharing platforms is the risk that data providers face to lose control over their data.

On the other hand, sharing of business data can fuel innovation, enable new business models and unlock revenue streams that are currently untapped. Additionally, the EU Data Strategy has data sharing as one of its pillars. As a result, it is reasonable that business data sharing will increase in the future. In this light, it could be useful for organisations as well to explore how they can stay in control over their data.

I will further introduce and elaborate on the key concepts during this interview.

A few days ago, I informed you about the informed consent form which explains the goal of this research project, potential risks that come with participating, and steps that will be performed to mitigate these risks.

To arrange this interview and to analyse the data, I need to collect and process personal data of you (such as name, e-mail address, voice recording). One of the risks is that your identity is exposed because these personal data is exposed unintentionally to others, or because interview responses could indirectly lead to your identity. This risk is mitigated by secured data storage, anonymization of the transcriptions, and summaries of the interviews. Furthermore, the recordings are only accessible by me (Thomas van Velzen) and my TU Delft graduation committee and they will be destroyed two years after this research project is finished. *Lastly, for respondents working for the internship company, there is an additional professional risk, for example unintentionally mentioning names of (former) clients. The researcher will mitigate this risk by removing all names of internship company clients and former clients from his documentation.*

Q0: Based on the form, do you have any questions?

If anything is unclear or you have a question, feel free to interrupt me. Before we start, I would like to inform you that in case it is necessary for time reasons, I might interrupt you to make sure we can finish all the questions. The interview takes approximately 60 minutes.

First of all, I would like to know briefly more about the company you work for and your background:

[Goal: getting background info about interviewee, also to comfort interviewee]

Q1: Could you tell me about your current position?

Q2: How long have you been working in this position?

I will further introduce the concept of business-to-business data sharing soon, but before we continue:

Q3: Do you have experience with business-to-business data sharing in your current or former positions?

Q3.1: Could you think of data within your organization that could be shared?

2. Business-to-Business data sharing via data marketplaces (±10 minutes)

[Zoom in further on B2B data sharing via data marketplaces, introduce concepts]

Next, I want to further introduce business-to-business data sharing.

Business data sharing can be performed via several arrangements. For example via bilateral arrangements or via data sharing portals owned by the company itself. However, I would like to zoom in on data marketplaces: digital platforms that enable business-to-business data sharing:

Q4: Are you familiar with data marketplaces where businesses can share data with each other?

If yes, let interviewee explain If no, introduce data marketplace directly

Next, **show visual of data marketplace** to make sure that interviewee understands

For the coming question, I would like you to assume that you and your organisation (*for internship company respondents: your clients*) are a (potential) data provider:

Q5: Which factors could influence your decision to share or not to share business data using a data marketplace?

Thank you for your answers and insights so far, I would now like to continue to the next part of this interview, which is about data marketplace meta-platforms. These types of platforms are the core of my project.

3. Data marketplace meta-platforms (±20 minutes)

[Introducing DMMPs, discussing specific factors for DMMPs]

Currently, a lot of data marketplaces have emerged. Some focus on specific regions, others on specific industries, for example, the telecommunication or automotive industry. I investigate the idea of developing an over-arching platform for these data marketplaces: a data marketplace meta-platform. This is a platform of platforms.

Show visual of data marketplace meta-platform > ask if concept is clear

Q6: Again, let's assume that you are a data provider. Considering data marketplace meta-platforms, could you explain to me what are the key advantages of these data marketplace meta-platforms in your opinion?

Q7: And, in your opinion, what would be the disadvantages of those data marketplace meta-platforms?

Q8: Now, compare the "single" data marketplaces we discussed earlier with these data marketplace meta-platforms. Can you tell me how meta-platforms would change the decision to share data compared to these single data marketplaces?

4. Data sovereignty (20 minutes)

[Open discussion about data sovereignty in the context of DMMPs from the perspective of data providers]

The next, and last, part of this interview is about the concept of data sovereignty. Data sovereignty means that organizations that create or generate the data stay in control over these data, even after sharing it with other organizations over the meta-platform.

Q9: Could you describe me what staying in control over your data entails when you consider the context of data marketplace meta-platforms?

I would now like to show you a visual of several concepts that are related to data sovereignty and I would like to ask you to discuss your first impressions.

Show visual of data sovereignty antecedents

Q10: From the perspective of data providers that want to stay in control over their data, what thoughts do you have regarding the different concepts?

Q12a: Data ownership?

Q12b: Data access?

Q12c: Data processing/usage?

Q12d: Data storage?

Q12e: Data control?

Q11: And how do you feel that they relate to each other?

Q12: For you as a data provider, would one or several of these blocks (i.e. data sovereignty aspects) influence your decision to share business data more than other blocks?

Q13: To what extent do you feel that the governance of data marketplace meta-platforms could be used to improve your level of control over business data sharing as a data provider?

If yes, for which block or blocks specifically?

6. Closing (5 minutes)

[Wrapping up, asking if there is anything the interviewee would like to add]

This interview now comes to an end. The information gained from this interview will be utilized to enhance our understanding of data sovereignty in the context of the data marketplace meta-platform. Your knowledge is extremely valuable.

Q14: Do you have any closing questions? Is there anything you would like to add or haven't discussed during the interview?

Q15: Do you want to receive the final output of this study?

Thank you so much for taking the time to participate in this interview.

Regards,

Thomas van Velzen

@student.tudelft.nl / @pwc.com

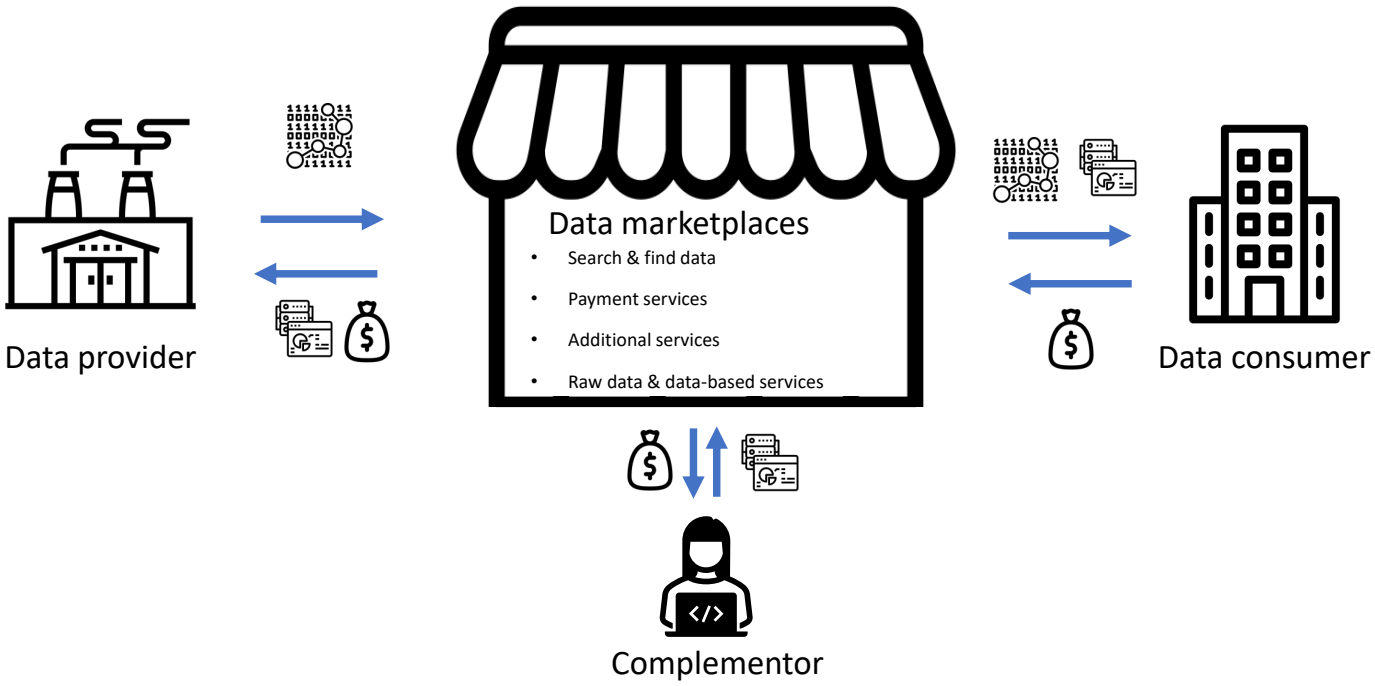
MSc in Management of Technology

Delft University of Technology, Delft, the Netherlands

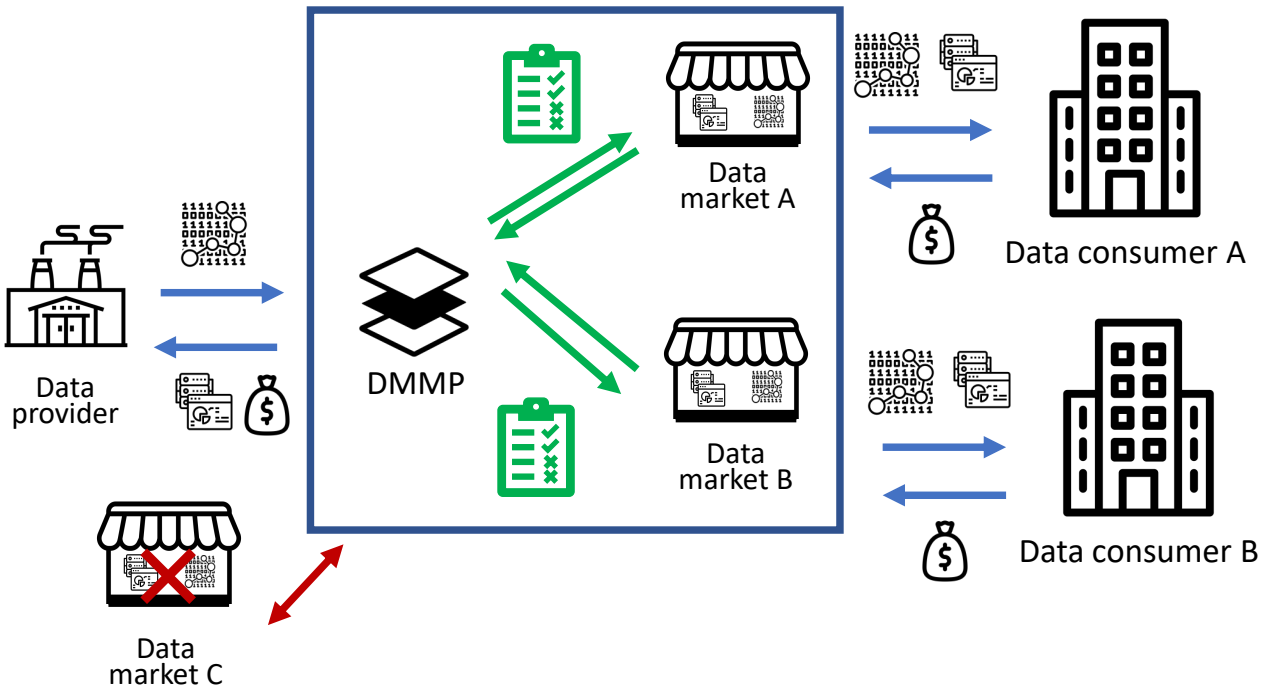
E

Interview slides

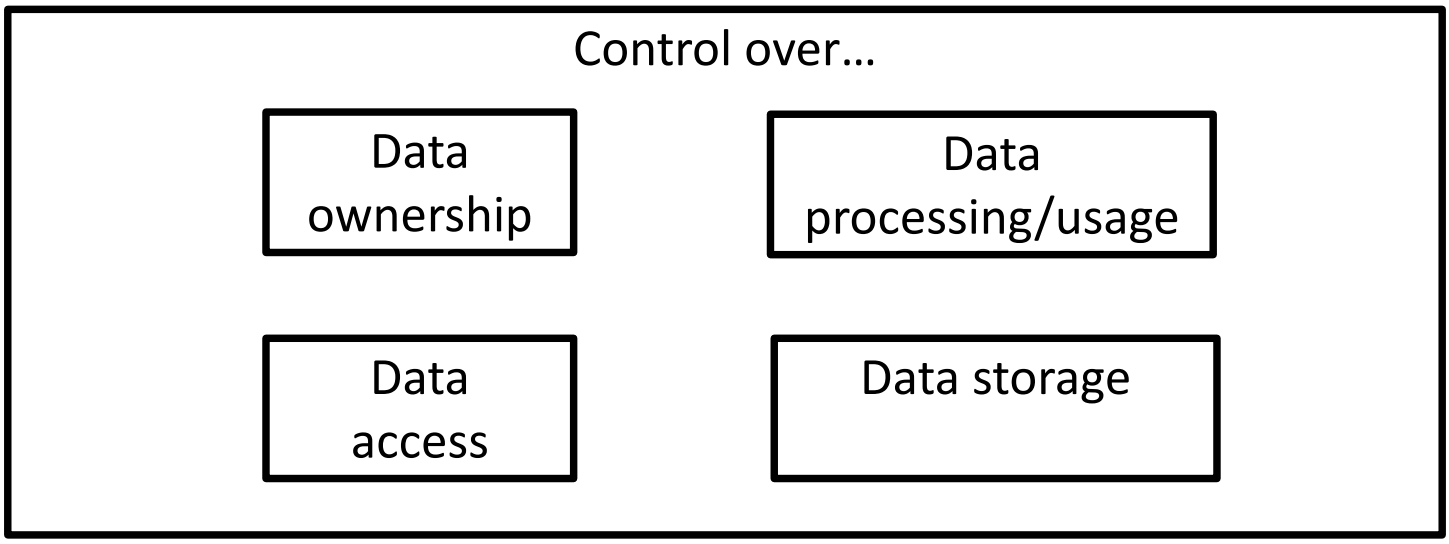
Data marketplace (DM)



Data marketplace meta-platform (DMMP)



Data sovereignty (DS) =



F

Interview summaries

In accordance with the informed consent form, the full anonymised transcripts of the experts interviews will not be made publicly available. However, summaries of these anonymised transcripts are available at 4TU Research Data[1]. The file with these summaries can be retrieved at:

https://figshare.com/s/94fb0763e33e989addf8

[1] https://data.4tu.nl/