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**

TU Delft pwc **Thomas van Velzen –**

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,

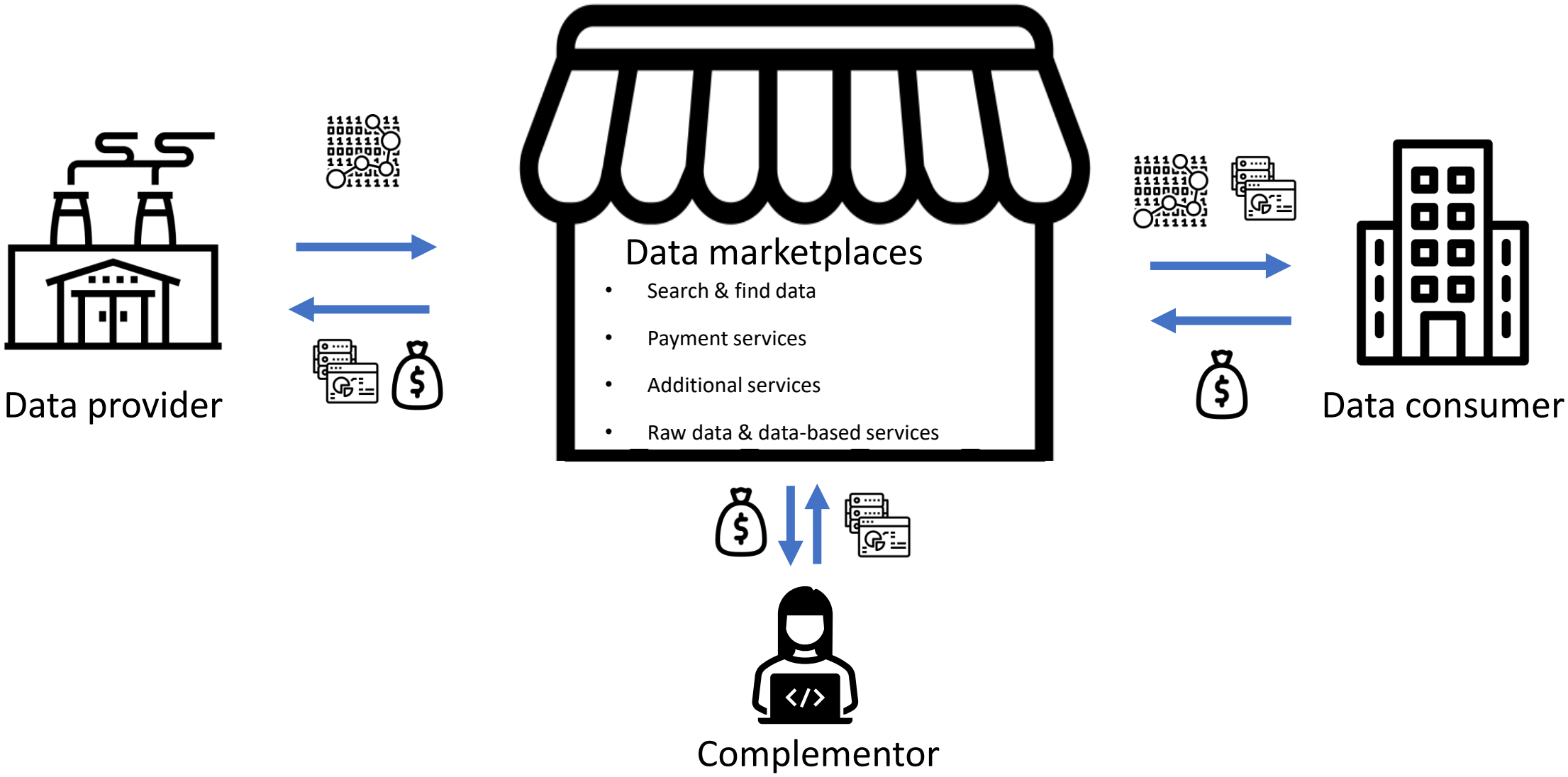
Thomas van Velzen

@student.tudelft.nl / @pwc.com

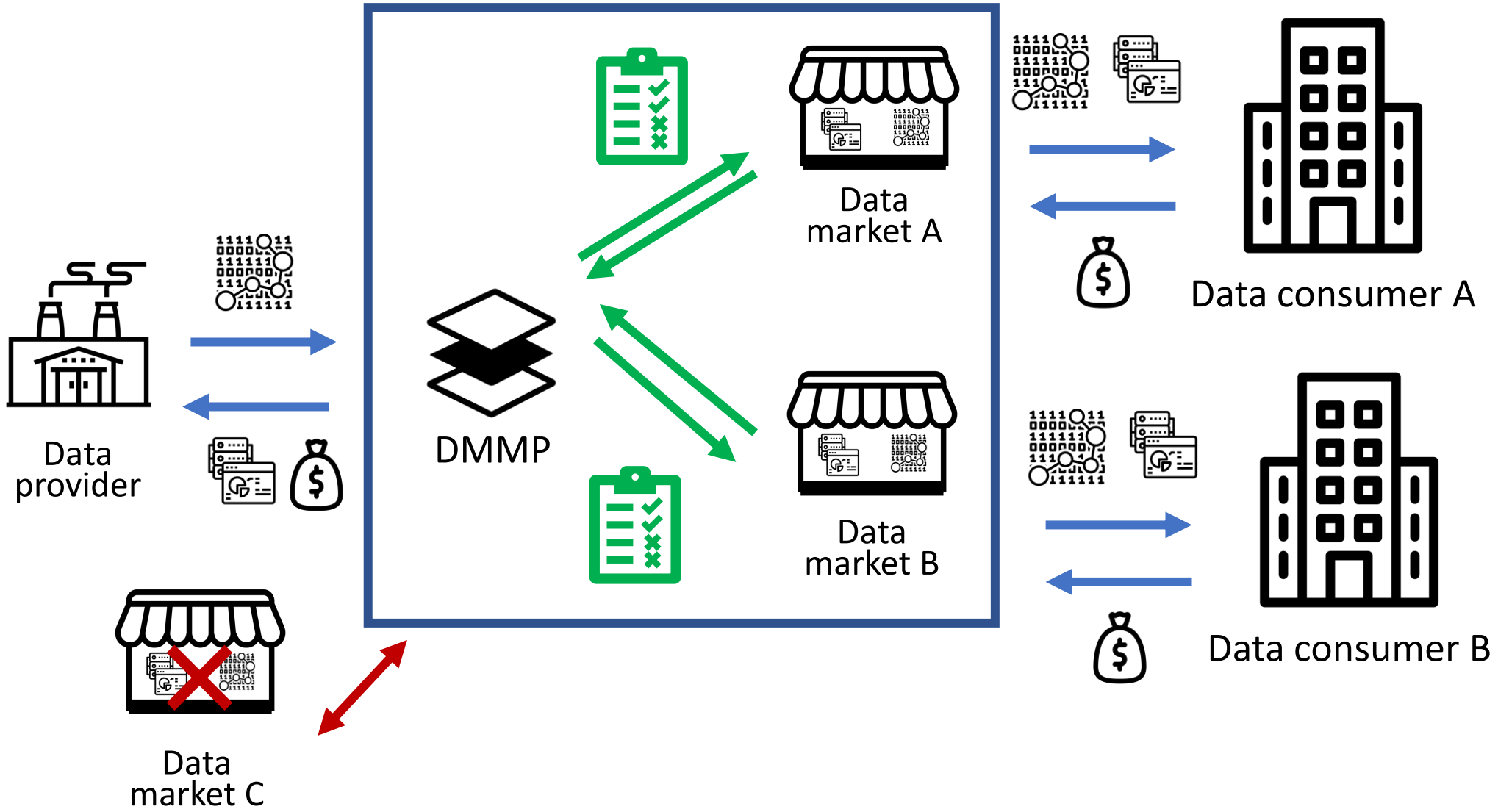
MSc in Management of Technology

Delft University of Technology, Delft, the Netherlands

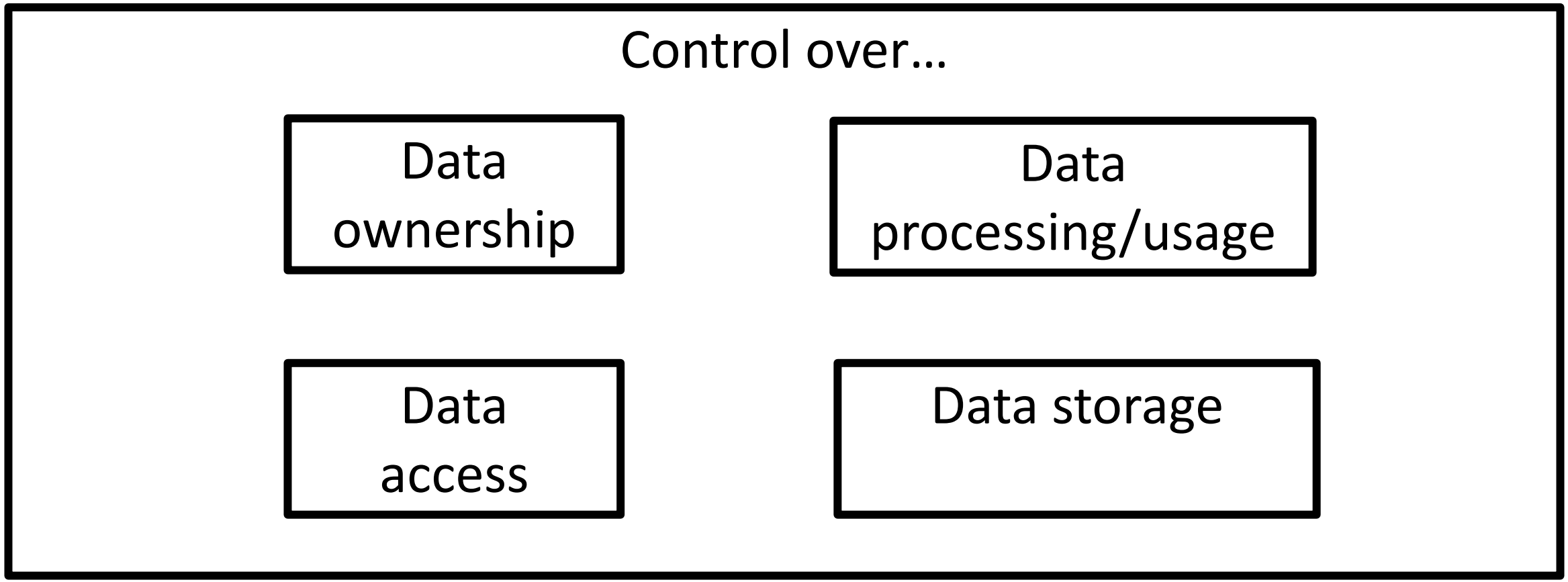
Data marketplace (DM)



Data marketplace meta-platform (DMMP)



Data sovereignty (DS) =



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

Expert reference	**E1**
Professional background	Data sharing and digital identity consultant
Date	3 June 2022
Duration	77 minutes

Introduction

E1 is working for a Dutch consultancy firm with five years of experience in data sharing. This consultancy firm was originally specialized in transactions, but used experience in this domain to specialize in data sharing as well. Furthermore, one of the other specialisations of the firm, digital identity, is very important in the data sharing projects as well. E1 in particular is currently working on an initiative to set-up data sharing within the design, construction and engineering sector. The expert pointed out that it is currently often the case that a collaboration of several organisations within a particular industry sector has the ambition to start sharing more data in a useful manner. E1 is often involved when these data sharing consortia consult the firm of this expert. Lastly, this expert also has personal experience with cross-sectoral data sharing, for example between banks and specific industrial companies.

B2B data marketplaces

After discussing the visual of a B2B data marketplace with E1, he clearly indicated to have experience with these platforms. However, according to E1, these data marketplaces are currently still part of a particular data sharing initiative centred around a particular solution for a specific business problem. E1 gave an example of a data marketplace that is used in a governmental organization for internal use. In that case, the users of the data marketplace are often working for or closely working with the operator of the marketplace.

After this example, E1 was asked about his view on B2B data marketplaces used by a larger group of organisations, so distinct from the example of the internal data marketplace, E1 mentioned the challenges of the two-sided market dynamics that are often playing an important role for digital platforms. He was very clear that for every platform, but for B2B data marketplaces in particular, it is crucial to balance all sides of the platform. He gave two pathways to solve this issue: 1) subsidizing particular market sides, 2) building the platform around a very specific use case. While the former lowers entry-barriers for certain players, the latter makes it more clear and specific for certain players what kind of value the platform can bring.

Data marketplace meta platforms

After the introduction of DMMPs, E1 wondered whether such a platform would focus on the technical exchange or not. E1 asked this as he thinks that the technical challenge within data sharing is rather limited. Next, E1 was asked to consider the viewpoint of data providers of a DMMP and to explain the initial advantages he could think of. His first reply was that he mainly saw a lot of disadvantages. E1 was asked to wait a few minutes to come up with the disadvantages. According to E1, the main advantage would be that all single data marketplaces are integrated, which could be nice for data providers.

Next, E1 elaborated on his perceived disadvantages. E1 indicated that data providers are very afraid what happens with their data. More specifically, E1 warned that this fear will only grow in a large-scale cross-industry DMMP versus the current situation where consortia are taking the first steps to start sharing data with their partners.

When asked to compare single data marketplaces with DMMPs and to indicate what the biggest factor for data providers would be to start sharing data or not, E1 indicated that on the positive side DMMPs could be useful to achieve interoperability between single data marketplaces. On the other hand, E1 indicated that it could increase fear of losing of control over data. He also mentioned the difficulty of closed private data compared to open data. The former type is very often also the most valuable data and providers are very conscious to share it with other parties. And even when they do, they are very careful to verify with which party the share.

In the last part of this interview section, E1 as asked if he thought that DMMPs increase the fear of losing control compared to single data marketplaces, he replied:

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."*

Data sovereignty in DMMP-context

The last part of the interview focused on data sovereignty in the context of DMMPs according to the interview protocol. E1 started to ask the interviewer about his understanding of data sovereignty. However, to prevent potential steering of the interviewee, this question was paused for after the interview.

In response to the question about E1's view on staying in control for data providers in a DMMP-context, he pointed out the importance of constraining data to a specific time frame, a specific use case and to limit access to specific individuals within the consumer's organization. Furthermore, E1 added that all these parameters could be made very fine-grained. Additionally, E1 emphasised the level of fine-graininess depends on the type of data as well. He gave an example of geodata or geolocations where this could be less critical.

After showing the visual of data sovereignty, E1 was asked about his view on the four blocks in relation to data sovereignty in a DMMP-context. He started with data ownership and indicated that this was important. However, E1 made a distinction in regarding data ownership. In his view, data ownership does not always mean that you are owner of the data, which would be the case with self-sovereign identify for example. According to E1 it could very often be the case that another party handles the data ownership on behalf of the data owner or data provider. In his view, this could for example be the case when small organisations share data but prefer to use a third party to manage it. According to E1, this use of third parties could also apply to data storage. For example, when third parties offer tools for data providers to control their data more easily or effectively.

While discussing data access, E1 introduced the triple A model of availability, accessibility and application. These three combined enable value creation from data. E1 mentioned that in his view data ownership and data quality are in the availability layer, whereas a DMMP is in the accessibility layer, because it tries to organise access to data. For E1, data sovereignty means control over all the three layers, where a platform could help to control the accessibility layer.

E1 was then asked whether he thought that a DMMP-platform could be useful for the application layer as well. He confirmed this, but also stated that there will always be specialised parties that are better for specific purposes or services. He gave the example that a DMMP could promise to deliver some AI-tools, but there is probably a company specialised in AI that offers better solutions. Additionally, E1 mentioned that data is much more decentralised nowadays than it used to be, which has increased complexity. Next, E1 gave some more background about his personal experiences which made him feel that for the specific industry sector he had been working with, there were several organisations that will never use a DMMP, because they are terrified for external parties dealing with their data. Although, he did not believe in use of a DMMP for that industry, he thought that there can be industries were organisation will use a DMMP. He also pointed at the upcoming Digital Services Act that could pose problems for DMMPs. He gave the example of Whatsapp that is currently very successful in offering all the three layers of the triple-A model, but will be faced with the DSA.

E1 was also asked about his view on data usage. An example was given for the energy sector, where a DMMP could offer dashboards based on inputs from multiple energy companies. E1 responded that cases where new data is created with data, it becomes very difficult.

Regarding data storage, E1 indicated that the capabilities for data storage also depend on the size of organisations. Larger organisations will more often decide to arrange it by themselves, whereas smaller organisations often outsource a lot of their data storage with third parties. For these smaller organisations, a DMMP could be valuable according to E1. Additionally, he pointed out that storage of data also impacts how up-to-date data is. Making copies in a database result in data being out-of-date very frequently. That is why keeping data at the source is a better solution according to E1. For a DMMP, E1 felt that it would be useful to enable parties to access the meta-data that describes the data at the source.

E1 gave a more detailed background about this particular topic from his personal and his firm's work. He emphasised that making agreements about data within a particular sector is currently also very effective. Making mutual agreements about data exchange and access can help parties that do not know each other to share data as they all comply to the agreements. According to E1, these agreements go much further than technical exchange and dealing with metadata, but also about identification, authorisation and fall back scenarios when things go wrong. These agreements lead than more to a data space, and help to create trust. In response to this information, E1 was asked about a project he is currently working on, where there were clear agreements for parties that enter the data sharing

project. E1 was asked if the agreements where used as an entry barrier to enhance quality of participants. E1 confirmed that this was the case, E1: *"Exactly, if you don't comply you will be kicked out."*

Next, E1 was asked if the set of agreements he mentioned where the main thing that creates digital trust. He responded that it is not only the agreements itself, but also that these agreements are enforced. Additionally, also that the content of the agreements are effective. For example, if it is decided to use a particular system for identification, that this system is also reliable and secure.

E1 was also asked in more detail about the necessity of a fallback scenario that he mentioned. He was asked about his view on what happens when it turns out that a provider has sold data that was not mend to be shared or turns out to have errors. E1 indicated that a solution for this type of scenarios should be developed by the industry sector itself, via co-creation because then the parties have both a stake and responsibility. For the particular example, according E1 this could mean that there is a pre-determined fee which the responsible party has to pay to the damaged parties. E1 also emphasised that he had personal experience with a sector where there was a table with liability fees.

Because E1 emphasis the decision making by a sector itself, E1 was asked how this could work out in a cross-industry set-up. E1 believed in coalitions between sectors and to have different degrees in standardization. He gave the example of healthcare, where there is a need for a lot of data sharing. In this sector, identity is now standardized using EIDAS2. This enables both private and public organizations within healthcare to use one type of identity. Fragmentation is decreased. E1 further mentioned that for a different industry this can be different. There will always be balancing between keeping things generic and specific.

Lastly, while asked about his view on governance to improve data sovereignty in a DMMP-context, E1 mentioned that the party behind a DMMP is also important for trust. The DMMP-operator has to be a very reliable player. Additionally, E1 touched on the business and revenue model of a DMMP:

E1: *"And also what conditions lie underneath. I can well imagine that a platform will be created that organizes it very well from a technological perspective, but then starts to exploit participants in the platform. I can well imagine that this will arise, in practice I think it is a lot more difficult."*

Closing

During the closing of the interview, E1 and T discussed the current state of data sharing in practice. E1 emphasised the relatively infant stage that organisations are in currently. He also briefly mentioned the different approaches to data at a global scale:

E1: *"it is a way of thinking that has not yet been introduced. We are very familiar with the central model, the Asian model. And everyone wants to platform, while it's very hard. From the private side a lot comes from the United States, and from the public side it's very much from the Asian model. And Europe is now in a kind of intermediate model, data at the source. And that's just not that easy."*

Expert reference	**E2**
Professional background	Date e-commerce project manager
Date	6 June 2022
Duration	70 minutes

Introduction

E2 is a principal product manager at a Nort-American e-commerce company developing and selling platforms for organisations to monetize their data. More specifically, E2 has been researching the process for data consumers to acquire data to better understand their needs compared to those of data providers. According to E2, there is currently still limited understanding of this process in this state of the data economy. E2 has been working with the firm for six months, but has a longer career of over twenty years in working and managing data in the telecom and financial sector. In his current position, E2 is mainly serving clients in the financial industry that are selling their data related to capital markets.

E2 gave extensive background information about the capital markets industry and explained that this industry is relatively mature in trading and selling data. Additionally, for data consumers the value of data is also clear is most cases. However, even for organisations in that industry, it becomes more complex when data from other sources is added or combined. E2 mentioned that this is what a lot of organisations are learning currently:

E2: *"And that's why you do see a lot of development on the AI and machine learning communities where they're trying to figure out, okay, we have all these data, but what can we do with it? What can we buy as a conclusion of using this or that data? You know, and what is actually being developed by pretty much everybody in the data economy right now is an understanding of how that data can be used."*

Using experience from his research outside the financial industry, E2 gave the example of the automotive industry where many automobile manufacturers are setting up their own data marketplace, which has led to a lot of fragmentation, whereas the value is at combining all these separate streams of data. E2 gave his view on centralizing all these data marketplaces or not and stated that he believes that one central platforms is not the solution. Instead, E2 argued that there will probably emerge a middle ground, potentially fueled by EU-regulations, where some standardisation is developed. Until then, organisations trying to get insights from data will not be able to grasp the full picture.

B2B data marketplaces

Next, B2B data marketplaces were discussed in more detail. E2 was asked about his view on these platforms and E2 described it as mainly a platform that enables a transaction between a data provider and data consumer. Additionally, he pointed out that the process of refining the data for consumption is done along this way. However, he emphasised that this process of refining and processing data can explode in subtility because refining can be done by the provider, the platform, the consumer and the type of refining is also unique for each type of data and application.

E2 explained that the products of the firm he is working for are designed in a way that the data provider gets the tools to refine and package data, but always can decide by themselves which decisions to make. For example, if a data providers wants to sell raw data, that is possible. If the data providers wants to refine or package data before selling, that is possible as well. E2 Additionally, the data providers are offered data contracts as well. However, E2 mentioned that there is still relatively little auditing of the contracts and tools. According to E2, making sure that the tools and contracts are used properly and are adhered to would be something very valuable.

After E2 was asked about what he thinks are the main factors that impact the decision for data providers to start sharing data, he responded:

E2: *"The difficulty that most data providers have is they don't know who actually is interested in their data or how they could transform it to actually feed the right use case."*

This is according to E2 also why artificial intelligence and machine learning are getting much attention during the last years, as these technologies can help industries to understand the value of particular datasets in their models. Furthermore, E2 mentioned the COVID19-pandemic as an example were organisations and societies were faced with an external shock that made the value of data very clear quickly. However, normal markets are much less pressing according to E2, so it can take much more time for organisations to get together in the data economy. The COVID19-pandemic is an example of quick action by many players, but for different situation, maturity levels in the data economy are different:

E2: *"Even if we have marketplaces where they distribute data, it's still very in its infancy actually. We're very far from the self-served. "*

Before discussing DMMPs, E2 was asked on other barriers for data sharing, both within and outside the financial industry. For capital markets, E2 argued that the main barrier is cost, as datasets can be very expensive, up to thousands of dollars. Although this can be less challenging for large financial institutions, it can pose a challenge for smaller players. Outside of capital markets, E2 discussed the privacy-issues of data sharing, especially in healthcare. Furthermore, E2 mentioned the lessons of the COVID19-pandemic for this issue as well: the pandemic has shown that organisation are able to overcome privacy-issues, but only when there is strong commitment across all players.

Lastly, E2 emphasised the importance of trust, both from added regulation, audits and identity. Regarding identity, E2 gave an example of data providers that perform a complete review of data consumers before entering a transaction. All these aspects contribute to trust according to E2, but due to the infant stage industries are currently in, there is still a lot being learned and tried out, which can be harmful for trust. Additionally, cultural context is an important factor according to E2. For example, how organisations deal with data in the United States versus Europe is very different.

Data marketplace meta platforms

At the start of the DMMP-section of the interview, E2 was asked to keep the viewpoint of data providers in mind. While discussing advantages and disadvantages of DMMPs, E2 was very clear regarding his view on DMMPs. E2 stated that he and his firm do not believe in one central platform. However, they believe that there will be companies offering tools and services to help organisations set-up their own data marketplace in a semi-standardised manner. This implies according to E2 that some things can be standardised, but that the tools offered will always enable data providers to offer their data on their own terms and conditions. Furthermore, E2 believed that organisation will always prefer to distribute their data to consumers using their own channels compared to a central platform, except for a scenario were there is a legal obligation. E2 gave an example of the dominant cloud service providers to further elaborate his argument and gave the example of Snowflake, Amazon and Google. Organisations that are using one of those for their cloud needs can determine to use the data marketplaces offered by those platforms, but are then forced to adhere to the specific conditions of these cloud companies. Some organisations are fine with, but others prefer to supply their data to consumers using their own platform with their own terms. This depends, among other things, on the level of control that an organisation prefers over their data. E2 mentioned that market dynamics will have to find out which set-up is preferred in the end. The best provider is the one that offers the best solution for a specific organisation:

E2. *"I'll say something blunt here, […]. We believe that the concept of a full distribution to everywhere from anywhere is possible, but it cannot be imposed. It will develop on its own."*

Next, while being asked about his view on what factors impact data sharing decisions for data providers when single data marketplaces and DMMPs are compared, E2 mentioned that is could also very well be the case that a single data marketplace gradually starts offering more sources and services, and becomes a DMMP on its own. E2 gave an example of a data platform for a particular stock exchange that has gradually grown over the years and now also offers many services based on external data, which consumer would previously have to consume from other marketplaces. Additionally, E2 highlighted that this again shows that all data platforms try to offer a full-picture based on as many inputs as possible. This links back to the earlier statement regarding the need to get the full-picture while working with data.

With this inside in E2's viewpoint on DMMPs, E2 was asked about his view on the development of data marketplaces towards the future. E2 indicated that trust is an important factor. This does not only relate to whether data providers and consumers can trust the data, but that they can also trust the platform. As there are currently so many different data marketplaces, a first step would be to have a common understanding of how trust is build up according to E2. Secondly, democratising and standardising access is important according to E2. Currently, all the major cloud providers such as Azure and Amazon use their own protocols and conditions. E2 compared this to the development of the GSM-standard 25 years ago, which is something we need for data as well in his view. Lastly, E2 mentioned that this understanding of trust and developing standards might especially be relevant, as the current situation of very dominant data companies has damaged the trust of its users. E2 explained that if this trend takes even further,

getting a full-picture using data might be impossible because the users that generate the data distrust these data companies not allowing them access to their data.

Data sovereignty in DMMP-context

Firstly, E2 was asked about his viewpoint on data sovereignty in the context of DMMPs. E2 replied that in his view this means being in control over your own data, even when using tools from external parties. This means that providers can develop their tailored data supply chain. E2 further mentioned that being in control is very different depending on the type of data that is being traded. He provided an example as well:

E2: *"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 regulatory [constrains] and audits that are done on the data controls. So, you can actually have validation and show people that they 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 and you don't necessarily have the value of the data that actually brings you the lawyers in and say, okay, sue these guys, because they didn't do what they should have done with the data."*

This is were the complexity comes in according to E2 for IoT data for example, which has no clear value up-front in some cases. In this situation, it is much more difficult to determine the fallback scenario, because the damage is not always clear when parties do not adhere to contracts.

To stay in control as a data provider, E2 gave other possibilities such as controlling access, renting data or sharing insights or analysis only without sharing raw data. According to E2, this is where the data sharing community is currently in, to figure out which technologies can help to solve these issues. E2 is also researching zero-copy technology for his work, which could be useful as well to prevent illegal copies of data that is being shared.

Because E2 had mentioned both trust and control (or examples of control) several times during the interview, his view on trust versus control was asked, to which E2 replied:

E2: *"I don't think trust is built with control. I think trust is built with transparency."*

From the level of an individual, he also gave an example of Facebook. According to E2, this social networking company offers all kinds of controls over data for users. However, for users it is actually very difficult if not impossible to verify that Facebook has indeed adhered to the preferences set by users.

After showing the visual of data sovereignty with the four blocks, E2 agreed that all these blocks are important for data providers for data sovereignty. He started to discuss data ownership and argued that it is crucial for data marketplaces that data providers are assumed to be the actual owner. That providers are responsible that they are indeed the data owner, which means the creator of the data. Regarding data processing, E2 mentioned that this could for example mean that data providers determine which part of their data is the most valuable and only sell this part of the dataset. E2 highlighted the factor of time as well. Data can be very dynamics and data sovereignty also means that data providers can decide to change their offerings at any point in time, including pricing.

Regarding data processing and usage, E2 considered the processing to be the activities to develop a sellable data product from the raw data, and the usage to be what the consumer will in the end to with that product. Data usage will also include what the provider themselves do with this processed data product. For E2, while comparing data processing and usage, data processing was important for data sovereignty, whereas usage was not. According to E2, this is because the end product of data processing was already destined to be sold. Lastly, E2 considered the data contract also to be part of data sovereignty. For example, to restrict what a consumer can and can not do with data regarding processing. Re-selling of processed data is something that is typically included in a contract, to prevent a data provider from missed revenue as E2 explained. In licensing of data, this is precisely what is done, to make sure that the provider of the original data also has a stake in the products derived from that original data. Regarding data storage, E2 argued that the platform offers enough flexibility to serve data providers with their preferences, whether this is at their own facility, or by using APIs to send it live to the cloud for example.

To close this part of the interview, E2 was asked about his final view on the four blocks. E2 agreed that all four blocks, except data usage, are part of data sovereignty and that data processing is the most problematic one.

Closing

To close the interview, E2 was asked if he would like to add anything not discussed before. E2 emphasised the current state organisations are in at the moment in the data economy: a lot still has to be figured out and everybody is trying to learn how data can be used in a meaningful manner that benefits society as a whole.

Expert reference	**E3**
Professional background	IT Architect/software developer/data sharing expert
Date	8 June 2022
Duration	51 minutes

Introduction

E3 is working as a senior scientist at a Dutch research institution. In this position, E3 has been involved for over three years with a large pan-European data sharing initiative, both as a software developer and architect. Furthermore, E3 has been working on the deployment of the results of this initiative. Additionally, E3 has experience with an initiative within the Dutch logistics sector to use the results of the European data sharing initiative in practice.

Within the logistics initiative, E3 was involved before data sharing was included in the scope. The initiative started with a group of Dutch companies working in a particular region and industry that wanted to develop a common language/semantics between each other. For example to standardise the format of purchase-to-pay information. One of the boundary conditions was that there were no lock-in effect for the players involved.

B2B data marketplaces

After E3 was asked to tell about his personal experience with B2B data marketplaces, the first aspect which E3 pointed out was that when organisations are charging money for data it becomes a lot more complex, mainly because everything has to be arranged to the finest detail before it can be actually done. E3 could think of a use case were data marketplace enable trading of technical drawings that could be 3D printed locally. Additionally, E3 was aware of privacy preserving technologies such as confidential computing that can help to restrict the usage of data that is traded. However, E3 emphasised that this is still a few steps ahead compared to where organisations are currently.

Next, the visual of the B2B data marketplace was shown and discussed. The first question to E3 was about the considerations of data providers to share data or not. His first thought was about trust. Secondly, E3 questioned whether such a data marketplace is the best solution, or it is more useful to have just a broker for the metadata, as is the case in the European data sharing initiative E3 is involved in. The main difference between these two set-ups is that in the metadata brokering situation the consumer makes a request at the provider, whereas at the full data marketplace set-up data providers provide and consumers can acquire. In the last case, data providers have to really make sure to have indicated possible usage properly according to E3. E3 argued that for now it might be better to have only the metadata brokering set-up as he thinks data providers will otherwise lose control more quickly. He compared a full data marketplace with data lakes that companies often adopt for internal use. However, in the case of data lakes, data is with a cloud provider that is now allowed to use the data, which can be different with a data marketplace scenario according to E3.

Data marketplace meta platforms

The DMMP-section of the interview was opened with the visual to explain the DMMP-configuration used for this research project. E3 understood the idea and also explained that this idea is there in his organisation, to federate data marketplaces. From a trust-perspective, E3 highlighted that technical diversity between single data marketplaces could make it difficult to establish trust among users. Additionally, E3 mentioned:

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. 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, ok you meet a common set of requirements."*

E3 then continued that he definitely sees value in a DMMP, especially for data providers because they do not have to register at all the separate single data marketplaces anymore. However, E3 still questioned whether a DMMP is necessary, or that a lighter solution could solve this issue as well, especially by using more of the existing infrastructure.

When asking E3 about his view on single data marketplaces versus DMMs and the decision for data providers to start sharing data, he explained:

E3: *"I think that's pretty much the same, except that you have to be able to choose your target audience. That you should be able to limit something. You basically have that in a single marketplace too. Then the next question is, if it all works technically, and you can have just as much confidence in your own marketplace as the meta platform, then I*

think that's right in my opinion. You will also see that a data provider in a single marketplace does not know everyone. That is the same for a meta platform."

Data sovereignty in DMMP-context

In line with the interview protocol, the data sovereignty section started with the question what staying in control over data for data providers means according to E3 in a DMMP-context. E3 started his answer that it is currently still utopia, but then continued that staying in control is always important: both when data is kept at the source or kept centrally. E3 then further elaborated on his perspective on control by specifying data for a specific purpose, a specific time period, but E3 added that these options can be endless. Additionally, besides setting conditions by the provider, according to E3 it is important that data providers actually are assured that other parties adhere to their conditions. E3 also added that this includes being sure that there exist no illegal copies of data. He summarised:

E3: *"That, I think, is the greatest point of sovereignty. Whatever you have with that of course, to make it possible in the first place, then you have to have faith in the whole architecture and software stack at all, I can be confident that there is not something in between that is skimming data. All organization and components will also have to be certified. All that together is that piece of sovereignty over that data."*

As E3 mentioned certification, this expert was asked a bit more in-dept about his view on this topic. More specifically, E3 was asked if the thought that certification is the most important component of trust. E3 replied that there are two things you want to have certified: the software code and installation on the one hand and the organization on the other hand. This certification of the organization does not only include the data consumers, according to E3 this also includes the DMMP-operator and additional service providers. In short, al organisations in the network. According to E3, certification can especially help to solve the trust-related problem where data consumers often use acquired data not only for the dedicated purpose, but also for other applications in their back-end systems. This certification can help to give a high-level confidence for data providers. In E3's view, this certification applies to what a DMMP can do to enforce rules technically, combined with legal measures.

Next, E3 was asked to elaborate further on the issue where data goes further into back-end systems of the data provider. E3 was asked if he was referring to the issue of visibility over data for providers. E3 confirmed but explained that is raw data is shared that can be used for analysis, the possibilities for consumers should not be too limited. Furthermore, E3 added that it again all depends on the type of data that is shared. He gave the example of a technical design drawing, for this type of data, which is already a data insight in itself, it is very reasonable that a data provider wants to limit where this piece of data can go. E3 also gave an example of using training of data for algorithms and models. In that case, a data provider and data consumer could for example agree that the training data can only be used for the development of the model, but that the model itself can be used further. In this case, the data provider is assured where the data pipeline is cut off, and the consumer has created value from the data by training the model.

The visual with the four block belonging to data sovereignty was showed next. And E3 was asked about his initial thoughts about the block in relation to data sovereignty. E3 started:

E3: *"The first one is kind of debatable. There is no ownership of data. It is often mentioned, it is a logical term, but not legally, for example. It's true, control over data ownership, in itself, you don't have a lot of control over that. Basically, there is an entity that sort of owns a piece of data. The question is, how much control can you have over that? What is a tricky one, what if you, for example in healthcare, if you have 1000 patients, with only a piece of data that is put together as a dataset, who is the data owner? That is a very difficult one, because if that is the 1000 people, that is possible, in the first step that is still possible. But what if an analysis is done on that data, who owns it? Is it those 1000 patients and the party that does the analysis? That is still a difficult issue."*

With this response, E3 explained his view on difficulties regarding data ownership. For marketplaces, he added:

E3: *"In itself, it is slightly easier with marketplaces, because there in principle the party that offers the dataset should in principle either itself or have received permission from the owners to offer it. In the broad sense, it is a difficult subject. Especially if you're going to combine things."*

In response to this answer, E3 was asked how he sees the combination of inputs from several data providers, which could be the case for data marketplaces and DMMPs. E3 gave a potential solution by comparing all the data use

policies of the different providers and by selecting the most strictest one. He concluded his answer regarding data ownership by stating that it is not always a bit problem, but that there are tricky things to think about.

Next, data processing and usage was discussed. E3 emphasised that this is difficult, remains difficult and will always be difficult. To address this difficulty, E3 proposed the use of both technical and legal enforcement in combination. According to E3, step 1 is to indicating what is allowed with the data and that both parties are aware of these conditions. Looking at the future, according to E3 the challenge is to shift from a combination of technical and legal enforcement towards technically enforceable as much as possible. This means for example the use of confidential computing, remote at station and a stamp from an organisation allowed to certificate. This all will contribute to the confidence for data providers that their data will be processed properly according to E3.

E3 was also asked about his view on smart contracts in this light of technical enforceability of data transactions. E3 knew about smart contracts and blockchain, but E3 stated that he is not a proponent of blockchain personally. E3 explained that it could be a first step to make clear for both the providing and consuming party what is allowed with the data and what not, because currently it happens sometimes that consumers do much more than was agreed on. E3 concluded his view on data usage on processing by stating that it is in general an important first step to make clear for both parties what is allowed and what is not.

Data access was discussed briefly as well. In E3's view, data access is highly related to usage and processing. He also highlighted that it depends on the configuration of the DMMP, i.e. whether data are kept centrally or at the source. For the latter, it is common that a consumer indicates a usage policy up-front and then arranges the exact contract with the consumers once an offer is made by the data consumer.

Lastly, data storage was discussed with E3:

T. *"Do you think it matters to providers how the storage of data and metadata is arranged for their sovereignty?"*

E3: *"In terms of storage sec, not much. I don't think it makes much difference to the provider how it is stored technically. 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 analyzes on this, we want to combine things. Then it is stored in several ways to make retrieving that data easier. I think the main thing is that you have confidence in how that data is stored, the party that does that also does it in a safe and correct way. And probably has it so organized that, for example, you really have a split between the data. That it will not be on one database. That you basically have a separate system or something along those lines for each provider. Something along those lines. Then it stands as a kind of safe at the marketplace. And still partly in control of that provider. That the provider can say, I now want to retrieve that data again. That they have the confidence that it really is over. That kind of certainty is then mainly involved. 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."*

Next, E3 was asked about his view on the blocks and to indicate which one is the most problematic in relation to data sovereignty in a DMMP-context. E3 indicated that the most important thing is confidence in the entire platform, i.e. data consumer, data provider and marketplace. E3 also emphasized the business sense behind data sharing; it is always the questions what the incentives are to start sharing. Additionally, he mentioned that organisations want to prevent being lock-in a particular ecosystem.

When comparing the current state of data sharing and a future DMMP-scenario, E3 summarised that currently data providers know which party for which purposes needs their data, very often because they have an existing relationship and people in the organization know each other. E3 emphasised that this is very different in a marketplace-scenario. E3 was also asked if he missed some blocks or thought some were redundant in relation to data sovereignty, E3 indicated that for marketplaces it is important to deal with metadata as careful as with the data itself. Primarily because the metadata can already reveal a lot about the organization and its data assets.

Lastly, E3 was asked about his final view on governance to improve data sovereignty for data providers of DMMPs. E3 argued that governance models definitely help, but only if the model is well written. Additionally, E3 mentioned that a lot of governance models are written, but are not effective yet, also because of the infant stage of data sharing currently. According to E3, a governance model can be very useful for situations do not go according to plan and there needs to be a fallback scenario. Not only because parties behave illegal deliberately, but also because mistakes always can happen by accident.

Closing

The interview was closed and E3 briefly mentioned how his organisation is experiencing development in data sharing and data spaces. He also mentioned that we are now seeing the first steps in practice, but that implementation remains a challenge.

Expert reference	**E4**
Professional background	Experienced IT and project professional
Date	9 June 2022
Duration	62 minutes

Introduction

E4 was working at as a product manager at a software development company for nine months at the time of the interview, but a longer background of over five and a half years within IT and project management. In this period, E4 has worked on several project related to data and data sharing. E4 has for example worked at a company developing products and services to solve the problem of distracted driving. This product included geodata, personal information data and insurance records data among others. This involved sharing a lot of data between several organisations to deliver the service to the end user.

In her current position, E4 is working on software to help organisations share data using a decentralised data sharing space. This includes for example the possibility to attach usage policies to data and metadata. The software of the organisation that E4 is working for never touches on the data itself, as it is focused on connecting users.

B2B data marketplaces

The first question asked regarding B2B data marketplaces was about whether E4 was familiar with this concept. E4 replied that she was, but that she also thinks that currently there is a lot of different understanding of the concepts by different parties. E4 knew personally about data marketplaces where open data and government data is offered in a centralised hub. Regarding data marketplaces where data is actually sold in exchange for a reward, E4 argued that it is currently just in a very infant stage. She mentioned for example that pricing of data and actually building the marketplace is still something parties are trying to figure out. Regarding pricing, E4 specified that some parties consider data being similar as stocks, where prices are determined by the market, whereas others follow the open economy concept more closely which means that everybody is free to price their offerings as they wish. A real working data marketplace with all the regulations, selling and billing is something that E4 is not seeing in the market currently. E4 also mentioned the Data Market Austria, which was running but is not anymore. She thought that this is because organisations are still afraid to share data. All in all, E4 thought that it all comes back to a lack of maturity and literacy regarding data sharing, which is, she emphasised, in the end not an issue at the level of organisations but at the individuals in those organisations. E4 thought that this is where industry should start to solve the problem.

Furthermore, E4 mentioned the view of her organisation that there is both a qualitative and quantitative approach to data sharing. Whereas the former is focused on developing specific use cases for particular data assets, the latter is more about offering data to the broader market without having a specific use case yet. For this former view, value of data is often much more clear in advance, because it is clear which purpose the data serves. E4 emphasised that this is important for organisations as they want to make money, primarily. Additionally, she mentioned that this is where industry should start, with developing use cases and not wait for the most comprehensive policy framework to emerge first for example. The type of data also matters:

E4: *"I also really think we should start with some steps that are maybe not that elaborated. Maybe you don't have the best policy framework ever. But having the basic policies in place and show to each other, it's positive. Maybe also start with not the most critical data. Don't start with personal data. Start with historic data. Makes some use cases to demonstrate this is working and build up trust in the system because thinking in a decentralized system is very hard to people."*

The approach where a use case is not clear in advance, but data is offered to the broader market, E4 emphasised the role and importance of metadata. Metadata could be useful to discover both data assets and data providers.

Next, the B2B data marketplace visual was shown to E4 and discussed. The first question was about her view on factors that determine whether data providers start sharing data or not. E4 started to explain that this is determined by which party can access which data. Additionally, she emphasised that it is important for data providers that a platform is decentralised, meaning that they only share metadata with the platform. E4 gave her view on data sovereignty in a few lines as well:

E4: *"For me, data sovereignty means acting with choice. Basically, all I want to say is I want to act with choice there. So even I offer maybe some sensitive data, which I wouldn't do in a first step. But even if I do, I want to be sure that I fully control who has access to that. I think this is something very important that I would feel safe enough to offer something."*

E4 argued that she would like to differentiate usage policies and pricing dependent on the consumer type as well. For example, data consumers being academic or research institutions should be able to get the data at a reduced price or for free. According to E4, differentiating usage policies could also depend on whether it is a single transaction of raw data or is traded for specific analysis-purposes.

Data marketplace meta platforms

After showing and discussing the DMMP-visual, E4's first question was about the configuration of the DMMP and whether it will be a central entity or just a connector of the different single data marketplaces. E4's first response was that a DMMP-configuration where data consumers use the DMMP (compared to data providers as is the case in this research project).

After asking about her perceived advantages and disadvantages of DMMPs for data providers, E4 mentioned the difficulty of developing over-arching standards in IT. She emphasised that people always try to develop one universal standard to get rid of the 14 existing one, but ultimately just end up with a 15th standard. She then saw value in DMMPs because it can help to overcome to issues in data science: finding the right data and finding data of high quality. However, E4 mentioned that data quality is very dependent on the intended use case of the data. Overall, E4 felt that a DMMP could help to overcome the scattered landscape of data marketplaces.

However, E4 was a bit more worried about the difficulties that do arise by creating a cross-industry data marketplace, for example because data schemes used in a particular industry are completely useless for others. E4 proposed to keep all the data at the individual parties, and just the DMMP-platform to improve discovery and searchability of data. E4 emphasised the use of labelling of data to improve searching data as well. According to E4, a situation where labelling is standardised and a meta-platform is there to search the offerings, that would already be an improvement compared to today's situation.

As E4 primarily mentioned advantages of a DMMP for data consumers to better discover data, E4 was asked if she saw advantages for data providers as well. E4 mentioned that she did see advantages for data providers, mainly to make it less difficult to offer data at more platforms. However, E4 emphasised that it is still very difficult, partly due to the differences in standards between industries. If there would in the future be a DMMP for data providers to offer their data at single data marketplaces, according to E4 it is still key to offer very comprehensive tools to set policies and data access by data providers. E4 mentioned that these policies should be defined by data providers, and not by the platform. In addition to the policies and access restrictions, E4 emphasised the importance of identity and its impact on trust in DMMPs.

Although these measures could help data providers to stay in control over their data, E4 had doubt whether it will be possible to accept all the different conditions of the single data marketplaces when data providers supply the DMMP:

E4: *"You have your conditions how a user can use your platform. But the marketplace A has different conditions than marketplace B. And how do you accept all their terms and conditions? With one click? How can you be aware that exactly is happening what you're now reading if you're connected to all of these marketplaces. And so, yeah, I really think we should, we should start with smaller steps."*

Data sovereignty in DMMP-context

This part of the interview was opened with the question what staying in control for data providers entails according to E4. She replied by giving what is means in the current state of the data economy:

E4: *"What does control currently mean? Currently staying in control means you offer your data, you force someone to agree on your terms and conditions, your policies, your licenses, whatever, so that the actual step of being in control current is someone is ticking a checkbox. And after ticking this checkbox, you just have legal enforcement."*

However, according to E4 this situation is not actual control, as it is just the possibility to keep the consuming party accountable and liable after agreed terms are not met. Using the replicability of data, E4 gave an example where this legal-based approach is not always sufficient to prevent illegal copies of data. To tackle this, a solution that is technical enforceable is preferred and can be technically done according to E4.

As E4 had referred to IDS while giving an example, she was also asked on her perspective on certification of the organisations active on a DMMP. E4 believed that certification can help to improve the system, but in combination

with for example verifiable credentials. Trust is the result of several actions according to E4's view. E4 summarised that this is a whole new kind of trading ecosystem, and just as in the regular economy, even with all the regulation and measures, some parties will always try to betray others.

Next the data sovereignty visual was shown with the four blocks. E4 was asked on her view on these blocks in relation to data sovereignty. E4 started to discuss data storage and mentioned that this could be important in the same sense as geographic sovereignty, i.e. when organisations want to store their data at specific parts of the world. Regarding data access and data usage, E4 emphasised the importance of self-determination by data providers. That they are able to interact on their own terms, and know before the transaction that the consumer has accepted these conditions.

Lastly, data ownership was discussed. E4 asked a bit more clarification and then replied:

E4: *"As you as you said, this part to me it's more legal. Of course, it's very relevant and it's connected to the term data sovereignty. But this part, in my opinion, if we talk about the sovereignty in the context of sharing, should already be clarified, because just as you said, am I the owner of the data because I owned the sensors or used the sensors or am I the company who bought the sensors. So whenever the data gets into a sharing context, data ownership should be already clarified."*

According to E4, having data ownership is a precondition to share data. However, in E4's view, the other three blocks (data access, data usage, data storage) are important for data sovereignty.

As E4 had discussed data access already quite comprehensively earlier in the interview, and gave examples of possible conditions that data providers would like to set, she was asked to tell more about her view on data usage and processing, especially in light of the additional services offered by complementors on a DMMP. According to E4, this could work similarly compared to the data access conditions. Data providers could for example limit the data usage to usage by specific parties, or for a specific application. However, according to E4 restricting the usage too much could also limit innovation.

Data storage was also briefly discussed again. E4 thought that data storage is indeed important in relation to data sovereignty for data providers. In her view, it means giving access to very specific data, which is something different than giving access to the storage. In E4's opinion, the actual storage is always in the data providers ownership and that should never be giving up. Next, E4 was asked about her view on data storage in relation to the creation of illegal copies which can be a risk with trading data. E4 called this control over the sink of the data, which is different data than the original dataset according to her. So data storage is mainly related to storage of the original data, whereas the data transaction which results in new data that should also be managed separately. However, E4 mentioned that this can become problematic, but that there are also solution such as certified connectors.

To wrap up this part of the interview, E4 was asked about her view on the blocks again and to specific which blocks are the most problematic or risky in her view. E4 replied that data access and data usage are the first to figure out. Data storage is something where there are several opportunities in her view.

Closing

At the closing section, E4 emphasised the process of becoming more experienced and mature as an industry again:

E4: *"To me, it's very important that as an industry, and as a data economy system, we should really start making small steps. But the most important thing is to start because most of the time, I feel like we're discussing for discussing and really many working groups, people try to achieve the most elaborated plans you can ever have. Knowing the complexity is so much."*

She closed the interview with highlighting that the issue of data sharing and data sovereignty is always about people. Even with the best systems and processes, it is always about people interacting.

Expert reference	**E5**
Professional background	Experienced professional in financial services and management consulting, currently leading the data practice of a North-American data e-commerce company
Date	9-6-2022
Duration	57 minutes

Introduction

E5 is an experienced professional with a background in the capital markets industry, especially regarding data about these markets. E5 is currently working at a North-American data e-commerce company helping clients to identify data assets that might be interesting for data consumers. Additionally, this company helps clients with the development of their own data marketplace as well. E5 in particular also helps clients to determine the licensing of their content and to become acquainted with market trends regarding capital markets data, for example the emergence of crypto currencies. Before joining his current organization, E5 has worked as a management consultant, has worked at a large US institutional bank and worked prior as a managing director of a financial data provider firm.

Before discussing the later topics of the interviews, E5 gave a very comprehensive overview of the capital markets industry and the state of data sharing and trading in that industry. E5 explained that the capital markets are relatively mature at data trading, and that this practice is happening for many years. E5 explained for example that fifty years ago, global stock exchanges discovered that all the data generated by this exchanges has value beyond the trading fees a stock exchange makes. Companies using the stock exchanges to trade started buying these datasets for their analysis purposes for example. Over time, large data consolidators emerged for the capital markets industry, such as Reuters, Bloomberg and others. E5 has worked at one of these large consolidators personally as well.

E5 described the differences between all the suppliers of capital markets data, for example different data formats. The value created by the data consolidators lays in overcoming this issue, to consolidate all the different data feeds in useful products. E5 also sketched the different type of product delivery channels for these capital markets data products: 1) distribution via a closed environment, 2) a consolidated pipe for use in enterprise systems. With the first set-up, the consolidator knows exactly who is using which data and how. For the second scenario, the data consumer uses the data feed in several application in a manner that is agreed upon by a contract. E5 explained that even with the contract, consumers sometimes used the data beyond intended use or distributed it among more employees than agreed in advance. Over time, the providers and consolidator of the capital markets data acknowledged that they did not capture full value and developed audit practices to make sure that consumers adhered to agreements. The industry as a whole also has become much more mature and sophisticated regarding pricing of data. Whereas it started with creating prices at the level of the enterprise of the consumer, it is now based on usage activity, amount of people and sometimes even location. The pricing also changed because revenues from floor trading activities went down and revenues from trading of data assets become much more important. Additionally, E5 added that a lot of the decisions made by the organisations that were trading data were done retro-actively, because they had to respond to changing dynamics.

B2B data marketplaces

E5 was first asked to give his personal experience and understanding with B2B data marketplaces. E5 explained that data marketplaces differ regarding the level of openness and service. Some data marketplaces are mainly offering users directions where to find particular data, but are not involved in the actual exchange. In this scenario, they are mainly data brokers. Additionally, E5 mentioned that data marketplaces which offer a full-service also present additional problems from a governance perspective, for example dealing with data leakages.

Data marketplace meta platforms

Next, the DMMP-visual was presented to E5 and he immediately asked whether a DMMP would be centralised from a data distribution perspective, or whether it could also be the case that data providers and consumers can also exchange directly between each other. It was confirmed to E5 that this could indeed be the case, as the different DMMP-configurations regarding centralisation are yet to be explored.

Next, E5 was asked on his perspective on DMMPs from the perspective of data providers and their decision to share data via a DMMP or not. Firstly, E5 explained the topic of disintermediation which is important for data providers in his view. Disintermediation, the reduction of intermediary players in the entire chain, played an important role in his personal experience both while working at the financial data provider company and while serving other clients in the capital markets. E5 further elaborated that companies which can be considered data providers prefer to stay as visible as possible, because when they deliver high quality data products this improves the value of their brand for example. If an intermediary party takes the data of providers and re-sells it under their own name, this harms the data

provider. Secondly, according to E5, data providers fear how their data is used by the data consumer and whether the data provider is able to capture any of the value that is created by the use of their data by the data consumer.

As disintermediation was mentioned several times by E5, he was asked whether this also implied that he believes data providers would prefer bilateral agreements outside a platform. E5 responded:

E5: *"Well, you know, if the platform is serving a useful purpose, there's no problem with it being in the middle. The problem becomes that you've completely lost touch with the user [as a provider]]and they don't know that you even exist or that you're a critical part. And again, if you're trying to make a name for yourself for whatever reason, maybe it's the marketing of other services, maybe it's as I said, you know, you want to go public or maybe you want to do a round of financing. You don't want to be anonymous. You want to be known as, you know, the provider of whatever content we want. You want to be able to bang the drum about it. You want to be able to do press releases. You want to be able to make sure the world knows you exist and that you're valuable. And once it gets, let's say, whitewashed almost or white labeled, whatever, you can lose that identity. And that's not helpful."*

After showing and discussing the DMMP-visual, E5 provided his views on the advantages and disadvantages of a DMMP for data providers and data consumers. Regarding data providers, the main advantage according to E5 is the increased number of potential outlets for data providers. Similarly, for data consumers it could lead to more choice in his view. Conversely, E5 also noted a potential counter-result for data providers: if a DMMP enables all data providers to supply more single data marketplaces, it can become difficult to stand out next to competitors. E5 compared this issue with search engine optimization (SEO) performed by many companies, although search engines like Google lead to more visitors to company websites, it is a challenge to stay on the top rankings of the search results for each website. If a DMMP makes it more difficult for data providers to differentiate, it can be a potential risk for these players according to E5, although some data providers could prefer this kind of dynamic as they can market their data more aggressively.

Lastly, E5 was asked to compare a single data marketplace with the DMMP and to give his thoughts. E5 mentioned that a DMMP could make more sense because data providers can distribute their data more broadly across single data marketplaces. Furthermore, E5 added that a lot of principles that apply to single data marketplaces apply to DMMPs similarly, except for the increased challenges to stay visible as a data provider (see above).

Data sovereignty in DMMP-context

Next, E5 was asked about his view on data sovereignty and staying in control over data as a data provider in a DMMP-context. E5 started to discuss selling data versus licensing data and the impact of analysis of raw data and the debate about ownership:

E5: *"This is a tricky one. Because. 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. 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. And that's a problem to this day in the industry, certainly in capital markets."*

E5 further elaborated on this issue and gave an example of exchange tradable funds (ETFs) which are often based on the data of existing market indices. The resulting product (the ETF) is made possible by data created by the market indices, and as a result, the organization offering the ETF has to pay a licensing fee to the owner of the market index as E5 explained. The exact sum of the fee is often determined in different manners, for example based on capital under management or trading volume.

Although E5 argued that in general data licensing can be a viable solution for data providers, it can also lead to disputes when data has errors for example as he explained:

E5: *"You know, most of your data providers will say you're licensing this content, I own this content. But then again, there is a caveat emptor. You know, you buy this data and if I screwed it up, somehow I made an error. Well, buyer beware. You bought bad data. Oops. You know, I didn't buy I didn't sell it to you for any fitness of purpose. You've decided on what you want to use this data for. If you use this data, it had a mistake in it. I'm not liable. Than the argument that a consumer could make is, well, then, if that's the case, I'm not paying you based on the value of your*

data because you're not sitting there beside me with the liability if I provided some bad information to a client. So. I've seen both sides of that discussion."

However, E5 still emphasised that licensing data is always preferable over jus selling data. Something E5 has experienced in the industry is that organisations that created and sold data end up competing with another company that is using their data to compete on products and services. Licensing data as a provider can protect against competing against your own original assets. For a DMMP, E5 stated that tagging of data could be a useful solution to overcome this issue for data providers, as the ultimate owner then remains visible. Additionally, E5 mentioned the use of audits, which is spreading beyond capital markets trading, to for example healthcare.

For personal and personally identifiable data (PII) in particular, E5 mentioned that it is always an issue to make sure that all the subjects in the dataset have given explicit consent before such datasets are traded. E5 was asked which party would be responsible for this type of data that it is traded in compliance with the laws in a DMMP-context. He stated that this should definitely be with the data provider, that the data provider makes sure that the subjects in the dataset have given explicit consent before the data is even listed on the platform. E5 argued that it would be unreasonable to put this responsibility with the data consumer or platform provider.

Next, the data sovereignty visual was shown with the four blocks. E5 discussed data ownership first and was very clear that this, as he had already mentioned earlier in the interview, should remain with the data provider. The issue of data ownership should be cleared before listing data assets on a platform. The licensing of data however could definitely be handled by the DMMP in his view, based on the needs of the provider. However, E5 did mention that data ownership without technical solutions such as tagging and labelling remains a purely contractual thing.

Regarding data access, E5 mainly emphasized that data providers should be able to handle data access on their own terms. In the extreme scenario that a data provider does not care and just wants to open the floodgates, in theory this should be possible. However, according to E5, in most situations there will have to be an appropriate data access agreement In place which has to be agreed upon before access to the data is actually granted. Next, E5 also mentioned that strict data access policies are also impacting the acquisition process for data consumers, which data consumers like to have as frictionless as possible. E5 also compared private company data with open data and indicated that this issue of granting access is probably less of an issue for open data, because it is primarily public data.

According to E5, data usage and processing is also about understanding as a data provider what your data consumers do or want to do with your data. This information can help to identify the licensing structure and fee model for example. Lastly, regarding data storage, E5 felt that storage is mainly about storing data securely, but is not limited to a particular place. According to E5, the data license currently often protects the data from unintended usage by others.

Building further on data storage, E5 was asked about his view on illegal copies of data and the duplicability of data. E5 argued that this means, as with other things not agreed on, a violation of the license for which the data consumer should be held accountable. However, E5 explained that in practice data providers often make an estimation whether the damage fee weighs up against the legal resources it requires to get after the data consumer.

E5 was also asked to summarise his view on the four blocks and to highlight which ones is the most critical for data sovereignty in a DMMP-context:

E5: *"Ownership is the biggest one really. I mean, I think that's the king. All these things follow out of that, right? Because of all these other things, the activities that would come from that have to all stem from data ownership."*

Next, E5 was asked if there were any block missing. E5 highlighted friction for data consumers, which could potentially increase by all of the blocks. E5 gave the example of Bloomberg, the major distributor of other organisation's data (related to capital markets), which is successful in creating a frictionless experience for data consumers be having a very tight and seamless distribution process. They handle also the billing for example. However, this means that third parties delivering data to Bloomberg have to adhere to their conditions. For this scenario this works as Bloomberg is able to deliver enough value for both data providers and data consumers. However, in the scenario that data providers do not agree with the standards of the platform, they can also remain using bilateral agreements according to E5. Furthermore, E5 highlighted the importance of reputation of fair dealing and proper control. This applies to

Bloomberg, but was built over the years. E5 was also asked on his view about how he sees Bloomberg as a platform. E5 responded that they have a high barrier of entry for data providers (due to high buy-in fees) but not for data consumers. The process of entry for consumers is rather easy according to E5. However, E5 emphasised that Bloomberg is a very closed platform, everything is controlled by a central entity, but data providers are still very eager to offer their data there because Bloomberg has a large footprint in the market.

Before closing the interview, E5 also came up with another potential advantage of DMMPs for data providers. According to E5, data consumers of have a difficulty with acquiring new datasets from new players that only sell data directly to others and it can take a lot of time to build up the trade relationship. However, data consumers that acquire data from data providers via Bloomberg can deal much quicker, because Bloomberg has already performed all the checks before that provider entered the platform.

Closing

The interview was closed by thanking E5 for his insights and participation.

Expert reference	**E6**
Professional background	Senior research specialised in trusted data sharing and business ecosystem architecture at a Dutch research institution
Date	10-6-2022
Duration	56 minutes

Introduction

E6 has been working as a Senior Researcher at a Dutch research institution. Furthermore, this expert has a background in both architecture at the enterprise level, but also in a more technical sense as solution architect. E6 is primarily working on trusted data sharing and data spaces in his current position. In his introduction, E6 also mentioned SCSN, a smart industry initiative, as an example of a trusted data sharing initiative that is starting to become successful. Before joining his current organisation, E6 was working at consultancy companies and a technical university.

While asking E6 about his personal experience with data sharing and data sharing initiatives, E6 described an initiative related to smart logistics where suppliers in a larger business network wanted to find a standard to exchange data regarding for example purchase-to-pay information. E6 further emphasised that his organisation is not involved from a purely business or academic perspective, but somewhere in that middle.

B2B data marketplaces

Next, E6 was asked about his view on B2B data marketplaces and his personal experience with them. This expert was aware of the latest developments regarding data marketplaces, but also noted that it sometimes appears that some initiatives are pushed by European governmental bodies. This push by governments also applies to data marketplaces at national levels according to E6. E6 was rather sceptical regarding this push by governments for certain data marketplace initiative, although E6 mentioned that it is always necessary to standardise things. He summarised:

E6: *"You have to facilitate some things as a government, both nationally and in Europe. You have to standardize the minimum in order to scale the maximum."*

Additionally, E6 would also like to see market forces helping to further develop data marketplaces.

After showing and discussing the B2B data marketplace visual, E6 had a question regarding the choice for the data provider's perspective to investigate data sovereignty in a DMMP-context. E6 questioned whether data sovereignty is primarily influencing data providers. Although this is an interesting question, the scope of the project was briefly discussed and decisions for research directions were elaborated. Next, E6 was asked about his view on factors that could influence the decision for data providers to share data or not via data marketplaces.

E6 argued that trading of raw data and data services directly via a data marketplace platform was not something he sees happening. However, E6 did believe more in a platform for data providers, consumers and service providers to find each other but placing data itself on a platform is not something happening according to the expert. Furthermore, E6 saw little use of placing data directly on a data marketplace for data providers and gave also an example of manufacturing companies that have to deal a lot with intellectual property. A similar thing also applies to the logistics sector according to E6, where placing data online, even with username and password protection, is something organisations do not want to do.

For open data, E6 could imagine that a marketplace could work. Furthermore, E6 emphasised the importance of dealing with payments, which could also be done by a party outside the platform, and more importantly: identity. E6 summarised:

E6: *"I think you're going to be a lot more concerned with identity. We get EIDAS2 coming as European citizens. But should it also be done by a central marketplace? Or should it be done differently? You now also see a lot of initiatives in terms of marketplaces that they use a ledger, and they say: we do all the metadata there. But I don't think it's useful to put down all the metadata, even if it's encrypted. Anything that is encrypted can be decrypted again."*

In response to this statement, E6 was asked whether he thinks that a de-centralised data marketplace where only metadata is shared and data is kept at the source is still high risk, which E6 confirmed. E6 then explained that sometimes there are scenarios where providers give up a part of their data sovereignty, but then only in exchange for

service and additionally only when dealing with trusted parties. E6 re-stated that market forces will have to figure out which data marketplace will offer the right level of functionality and standardisation for current industries.

Data marketplace meta platforms

After introducing the DMMP-visual, E6 was asked to elaborate on his view on advantage and/or disadvantages. The expert mentioned a parallel he recognised related to semantic standards, where a new standard tries to integrate the existing 14 standards, but ends up to be the 15th standard. Furthermore, E6 compared a meta-platform with a god mode (a function on Windows desktops that opens up all the settings that are normally not accessible for users). Additionally, E6 was critical about the potential vendor lock-in effect that can arise when organisations can not without the meta-platform anymore:

E6: *"You have to standardize them so that they can be exchanged. I think you should go there, and not another god mode. You have to look at how we can link those data markets together. What is the minimum functionality that you have to standardize, so that you are interoperable? And then it gets fun…"*

To further explain his viewpoint on meta-platforms versus federated data marketplaces, E6 gave an example of GS1 Data Source, which the organisation behind the barcode for products in supermarkets. Additionally, this organisation has a complete standard for all the data fields related to supermarkets, not only the barcode but also planograms for the shelves of the supermarket. The GS1-organisation adapts their standard for each country, but also keeps a lot of the data field the same across several countries. E6 summarised:

E6: *"With such a data marketplace you have a number of elements that are mandatory to be interoperable between different marketplaces, additionally you enter the specific domain. So that could be country or sector. I absolutely do not believe in a god mode or a 15th standard. What I do believe in is standardisation. Then you get more of a flat, federated model."*

Next, a re-cap was made to the question regarding advantages and disadvantages of DMMPs. E6 emphasised the winner-takes-all dynamics that often come into play with digital platforms. According to E6, this is not something we would want to have in Europe, especially considering our democratic principles. E6 also mentioned that this depends on the cultural context and that in other parts of the world other choices are made. Additionally, E6 gave examples of Uber and Just Eat Take Away that initially present themselves as a better digital platform-based alternative to existing services, but end up being very aggressive in their battle to win market share.

While being asked about single data marketplaces versus DMMPs and the main difference for data providers:

E6: *"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. And that can be done via legal enforceability, technical enforceability and ethical commissions and so on. I just don't see any added value where a local company around the corner, or an IT company, would make their datasets or algorithms available there. Those are just yours, the only reason I see such a data marketplace for B2B is that you can find each other. Or for open data, such as the weather or public transport. But those are all openly available APIs. So again it's about connectivity and being able to find each other."*

Data sovereignty in DMMP-context

To open the data sovereignty-specific section of the interview, E6 was asked in more detail about what staying in control means to data providers in his view. E6 started that staying in control is as much about control over metadata as it is as control over the data itself. Additionally, according to E6 staying in control over data also entails non-repudiation and traceability.

Next, the slides with the four data sovereignty block were shown and introduced. E6 indicated that these terms do make sense in relation to data sovereignty in a DMMP-context. More specifically, according to E6, data sovereignty is mainly about access control and usage control, because consumers need data for a specific thing. According to E6, access and usage control can be realised via legal or technical enforcement. The former is used to make sure that parties stick to agreements, because there is for example a penalty otherwise, whereas the latter is about building technical mechanisms that make sure that non-compliant behaviour is simply impossible, or is automatically detected and notified.

E6 sees opportunities for technical enforceability in the future, for example confidential computing, although this mainly applies to usage control in his view. Regarding legal enforceability, E6 argued that this works pretty handy currently, because there is just a party that can be kept liable when things go wrong:

E6: *"The problem is still there, but it has been bought off. That is the practice now, but you would actually like to go further in the future."*

For DMMPs, E6 indicated that when he sees the visual he gets the idea that the DMMP operator becomes a similar trusted third party that manages and oversees the transactions and is liable if something goes wrong. E6 added that similar agreements are made in the textile and logistics industry. When asked why these trusted third parties are trusted, E6 indicated that the main reason is the legal contract below collaborations.

Next, E6 was asked to provide his view on the data sovereignty block, their relation and possibly missing blocks. E6 started with data ownership, as the data owner is the party that grants access according to E6. Additionally, E6 stated that many data owners outsource parts of their data storage needs. Furthermore, E6 mentioned that data owners and data providers can be separate entities, where the data providers store s or manages data on behalf of the data owner. Additionally, E6 also gave a real example of such a situation where data provider and data owner are different related to an organisation working in the Dutch healthcare sector.

When asked to indicate which block would be especially critical for data sovereignty, E6 did not mention one block in particular, but indicated:

E6: *"I think you should especially look, are you interoperable? That is ISO25010-020, that you have to be interoperable and portable. If companies want to work together, the tools have to help, help with that. And you just don't see that very often. That is still a problem. I think you should take a look at that."*

Closing

The expert was thanked for his views and contributions to the research. E6 had no points that he missed during the questions. He ended by emphasising the risk of coming up with a 15th standard when trying to develop over-arching platforms, as he mentioned earlier during the interview.

Expert reference	**E7**
Professional background	Director of pan-European trust and data sovereignty framework, combined with broader experience leading IT- and technology-driven companies.
Date	15 June 2022
Duration	53 minutes

Introduction

E7 is the director of the foundation behind a Dutch initiative centred around the development of trust framework for cross-industry data sharing. Additionally, this framework is currently also being adopted outside of the Netherlands across the globe. This expert explained that this particular initiative started a few years ago in a particular industry sector where organisations acknowledged the value of data, but understood that keeping control over data was crucial to get data sharing starting. E7's trust framework is created to be horizontal, across data spaces and the different data domains. The goal was to establish trust not only legally, but also technically and regarding both access and usage control. E7 in particular has been involved in technology during his career for 25 years currently.

Next, E7 wanted to comment on the opening about the research project, especially regarding data sovereignty for data providers. E7 wanted to add that this already poses a definition discussion, as it is frequently the case in practice that the data provider is a cloud provider, providing data on behalf of the data owner or entitled party. E7 mentioned that this is also where initiatives like IDS and Gaia-X originated from, to improve interoperability between the segments, because it is important to not make new silos again.

E7 was also asked to give an example of personal experience with B2B data sharing projects and gave an example of a project for a national governmental organisation where the trust framework helped this organisation to measure the impact of sustainable energy subsidies by connecting to the smart energy meters after explicit consent by means of authorisation by the citizen or home owner. The authorisation includes the intended use on which users base their permission, and also includes a provision that the data can not be sold for other purposes than the agreed purpose. Although reselling is not possible in this scenario, citizens do authorise this governmental institution to use the data for their annual report for example. This data space with a specific use case does not only reduce hassle, but also ensures up to date data according to E7.

Furthermore, E7 emphasised that the trust framework mainly contributes in the permissions and authorisation of the requesting organisation. Parties first have to sign the terms of use of the network around the trust framework before parties can start sharing data and requesting access.

E7 was then asked to explain a bit more about the role of legal components in the trust framework. E7 clarified that legal and technical go hand in hand. Furthermore, E7 described the trust framework as a network of parties that have all agreed to the same terms of use with a digital way to verify each party. In a next step, when two parties decide to enter a data sharing agreement with each other, unique keys are generated which protect the data transaction. Whereas these unique digital keys can be seen as a technical measure, E7 highlighted that each transaction is also covered by a contract. If one of the parties does not act according to this contract, this party is in breach and liable.

B2B data marketplaces

E7 started this section of the interview with his view on B2B data marketplaces. He explained that there are two entry points looking at data needs: 1) starting with a specific request to communicate with a company and identify which data is available, 2) starting with the need for a specific data asset and trying to get this from X amount of organisations. In the former, the data provider clarifies what is available and data consumers are able to request a transaction, whereas in the second scenario, data consumers have a clear idea what they need and try to find parties that can deliver is. For the second scenario, data marketplaces can help consumers to find the needed data, ideally to get from harmonised supply to the actual endpoints of the APIs. E7 also gave an example in the form of AMdEX where explicit consent is given on closed datasets, but also added that it still remains a puzzle.

After discussing the visual, E7 remarked immediately that the visual made a simplification that is also done very often by others: thinking that data providers and data owners are the same entity. According to E7, as we are living in a cloud world, a data provider is not always the same entity as the data owner. More specifically, E7 prefers to speak of data provider and entitled party, where the data provider acts on behalf of the entitled party. E7 highlighted that this also better aligns with legal views on data ownership.

Next, E7 was asked about his view on data providers (as defined in the visual) and the factor which influences the decision to share data or not. E7 started to answer this question by considering the different types of data:

E7: *"and that's why you actually have to keep in mind that there are different classifications of data, only when you have that in focus, can you see when and which party could start sharing what data. Simply put, you have open data, you have condition-based shareable data, and you have classified proprietary data. Open data, you probably know, is a data set that can be downloaded by anyone. Condition-based is data, let's compare it with a book for example. You buy that book, and you get the data in it if you have met the condition that you have paid a certain amount. And you get it with the condition that it's copyrighted, so it's yours solely and you can't resell it.*
[…]
Proprietary data has a different classification. With condition-based it is quite generic; it doesn't matter who meets those conditions to get that data. That's typically data that's on data marketplaces, just like open data, that's where AMdEX originated. That closed, proprietary data is your business-sensitive dataset. You really need explicit consent for that."

E7 was asked to elaborate on condition-based versus proprietary data. The expert emphasized that proprietary data from the perspective of data providers is more about questions whether the data consumer is one of the organisation's competitors, whether there is any business with sharing such data. Additionally, E7 explained that when this type of data is licensed, the periods are often much shorter and are only limited to the time frame in which the data is really necessary to use. Lastly, when asked what would be the primary reason for data providers to share proprietary data, E7 replied that this is mainly about whether is makes business sense or not. According to E7, this is also why the data spaces that are emerging currently are often focused around specific services where specific types of data fulfil a need. E7 also mentioned that parties need to stay in control over data in line with the EU Data Governance Act.

Data marketplace meta platforms
After the visual of the DMMP was showed and explained, E7 was asked if he saw any advantages of such a platform. E7 indicated that the idea of a meta-platform is also emerging in the Data Sharing Coalition, where they are called proxies which operate between different data spaces. However, E7 stated that he personally does not strongly believe in these proxies currently. Partly because the trust framework he is active for is already helping parties to find marketplaces where data can be found that a data consumer is looking for.

In sum, E7 thought that there are two options: working with proxies (or meta-platforms) or realising organisational interoperability. Regarding the latter, E7 gave an example of context brokers that use data and metadata in combination with linked data to make data spaces from different domains interoperable.

E7 personally believed that in the future there will be more of a federated structure where data spaces decide for themselves which standard to use, but that it is possible to find marketplaces within this structure.

Next, E7 was asked about possible disadvantages of DMMPs he could think of. E7 started by emphasising that the core principle must be that there is always control by the entitled party of the data (i.e. the data owner). For example, where what happens to the data, where it is published according to E7. In line with the Data Governance Act, there must always be explicit consent E7 added.

Data sovereignty in DMMP-context
The first question asked in this section of the interview was about what according to E7 control over data means for data providers:

E7: *"Yes, so data sovereignty is self-determination over data. That means you know which data belongs to you, is labeled to you. You always need that classification for that. This organization has this and that dataset."*

To realise this, E7 mentioned that for data access for example, that entitled parties (i.e. data owners) always know which parties can query their data at the source. And following access control, if someone can query:

E7: *"And if he has asked that question, what can he do with it? Usage control through licenses."*

According to E7, data storage is the third step, as it is follows from data usage. More specifically, E7 indicated that within his organisation's trust framework this is arranged using licenses. These licenses can be enforced legally, but possibly technically as well. However, technical enforceability is still a challenge according to E7.

Next, E7 briefly added some more background on his view regarding access control and the role of the licenses:

E7: *"Yes, in **our initiative** we have the rule that we give access to APIs at the attribute level. That is a very important one, because data sovereignty is about being able to control which data set and which [data] field may be shared for which period. That is the access control, in a very fine-grained manner. And then the license is added, what can you do with it afterwards."*

The process according to E7 must be: 1) determining which data fields to make available, 2) for which time period, and 3) add the license to control the possible usage. E7 added that it is not always a purely bilateral transaction between an entitled party (i.e. data owner) and data consumer at both ends of the transaction. According to E7, data consumers often rely on third parties to fulfil their products and services and they often arrange with the entitled party certain delegation levels, i.e. agreements regarding which other parties where the consumer collaborates with will also be able to use the data. Within the trust framework of E7's organization, there are several levels of this delegation and the choice whether delegation is allowed or not is one of the decisions in itself.

E7 provided more details about the acquisition process as well. In general, all parties within the trust framework operate on the same legal basis, and then arrange the specific agreements for each data transaction. More specifically, the data consumer often makes a relatively detailed request at the entitled party as a first step before the actual transaction. After this request, the entitled party decides to accept, reject or adapt the proposal.

Next, E7 was asked to elaborate on the interplay between legal and technical enforceability of the agreements and licenses and the problems that are still open:

E7: *"There is an insane amount of problems when it comes to understanding how that works. How a federated approach works is still very difficult for everyone. Which also makes sense, because it's a bit about the new world. We are used to concluding a contract with a software supplier, who engineers something, and if that doesn't go well, you know what to do. In this case you have a broader network of parties with whom you work dynamically. And how that game works, we are of course still discovering that together."*

These examples of issues which currently exist in the data economy also raised the question how E7 sees the future and which parties could be drivers. E7 explained that his organization is focusing on collaborations with large institutions, to reach scale. One of these institutions is for example a large governmental institution. Taking this first step helps to roll out the concept among many more companies that have for example a duty to report to this governmental institution.

E7 was also asked about the role of governments in this phase:

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."

Earlier, E7 mentioned the difficulties with technical enforceability of data sharing agreements. After a re-cap to this topic, E7 explained that there might be useful application of multi-party computation in the future, specifically to enforce usage control technically. However, in order to do this, the entitled party still needs to have access control. E7 also explained that especially many SMEs are lacking the technical capabilities to adopt these technologies in the near future. Again, this is why E7 is focusing on larger organisations as adopters of the trust framework, as they are often more mature on these topics.

As all blocks were discussed along the way with some side-steps, E7 was asked if there were any blocks missing in the visual. E7 replied that in his opinion it is especially important to include the need for fine-grained control. He included an example as well:

E7: *"Many application parties say, don't I have access control? Because I can determine whether or not you can log in. Brilliant, but that doesn't help you at all, because then you have no control over what happens to your data. That is a crucial one in everything. And then you only have data sovereignty in this game if you, well, if it is also legally covered. It's not a purely technical game, it's legal and technical together. That one is also often overlooked."*

Regarding the last question whether E7 thinks governance can help to improve data sovereignty for data providers in a DMMP-context, he clearly agreed. And also repeated the goal of his organisation's initiative, to develop a trust framework that can be used among very different data spaces. He also mentioned Gaia-X which is setting up an association or foundation to monitor participants in a non-profit model.
Closing E7 had no further questions and the interview was closed after some after talk related to the trust framework of E7.

Expert reference	**E8**
Professional background	Board member of regional collaborative organisation, specialised in future affairs including digital and data-related topics
Date	20 June 2022
Duration	47 minutes

Introduction

E8 is working for an organisation in one of the largest economic regions in the Netherlands focused on collaborations between governments, companies and knowledge institutions. The goal is to think about developments which are important for the future. In this position, E8 has been involved with a large data sharing initiative since its beginning three years ago.

The organisation realised that if parties want to exchange data but still stay in control, there are innovations needed to facilitate this, legally and technically. E8 explained that when consortia of parties want to share data, some facilities are currently still missing to enable this. This relates to being able to enforce agreements that parties make with each other. Within E8's initiative, data remains at the sources:

E8: *"Additionally, it is not so much about us transporting the data, because data simply remains local with the owners. What matters to us is that we can record and clarify the rules that they mutually make with each other in order to ultimately share data on a large scale in the future. So that they know where they stand and also have a place that if parties do not adhere to it, they can go there. For resolving disputes."*

The approach is driven by specific use cases and E8 elaborated on one of these use cases where the goal is to open up the data generated from sensors and other sources located on a specific terrain within the region of the city. Parties interested in particular data, for example a journalist, can make a request that the providing party can accept. The use case is still in development and E8 remarked that currently, the actual transfer of data is still in the form that an actual package of data is sent. Subsequently, E8 gave examples of other use cases he and his organisation are working on, for example data sharing initiative within healthcare were very sensitive data is exchanged. In general, E8 explained that these use cases help to better understand how to make general rules that work for very specific cases and to learn about other needs that the initiative could fulfil.

Additionally, E8 explained that his organisation also collaborates with a party that is very important for the working of the internet. This partner is governed as an association and this helps E8's organisation to find opportunities for the governance structure of E8's initiative as well. After elaborating on this:

E: *"That makes it decentralized, which is an attractive mode, a similar model could also work for **us**. That is also being explored. Could we implement and use something similar as well? And then you have the business model. Our initiative also has to generate income somewhere, the chance that it will only work from member contributions, I don't know, maybe it can. We should look into that as well, or should there come something from elsewhere as well. And what exactly is the value that you offer there? Because that's important too. What is the value proposition? So there is also a lot of attention at that level. What could that mean."*

B2B data marketplaces

After introducing the B2B data marketplace visual, E8 was asked about his view on factors that could make data providers share data or not on such a platform. E8 opened with the importance of trust and mentioned questions that data providers have regarding what happens with their data, if they have control over the use, identity of consumers and if they are able to restrict the usage and consumers. He then remarked:

E8: *"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."*

The expert further elaborated on this line and shared his view that a data marketplace should be more of a market master and that data providers can also get other things back then a monetary reward. For example, better operations within their own business because they learn based on the data from the consumer and other providers as well. He saw the role of data provider and data consumer as very dynamic and continually changing. He added that this is also very dependent on the set-up of the marketplace, for example if it has very commercial goals or is for example about improving safety in an industry sector.

Data marketplace meta platforms

Next, the DMMP-visual was introduced and explained. E8 was asked about the advantages and disadvantages he could think of for data providers. The first one E8 mentioned was that data providers could set their restrictions and conditions at one place and that their data could only be brought further to other marketplaces if these marketplaces can adhere to these conditions. However, this requires that a DMMP is able to generic solutions that are still workable in more specific data marketplaces and applications as E8 mentioned:

E8: *"You have to be able to say something generic about it. And is it then possible to develop specific rules for different data markets? Whether or not it can be offered."*

E8 also saw potential cost savings due to the scale of a DMMP, for example regarding legal costs. However, according to E8 this requires that a DMMP is able to automate a lot, and must for example be able to resolve disputes highly automatic helped by technology:

E8: *"if you can digitize agreements with each other and have them automatically set up, then there is less and less need for a lawyer and notary."*

For data providers, E8 also saw value in the sense that they have a broader set of single data marketplaces that they can approach for the set-up on the visual.

However, E8 also explained potential issues that could arise:

E8: *"It's just very, very complicated. You know even less from whom and where it is offered, when? To whom? The line is even less transparent. I think it's music of the distant future."*

This expert shared personal experiences that underlined his argument where sometimes it is even difficult for five organisations to agree on agreements for one single data marketplace. E8 thought that organisations should first overcome this, and that in a future scenario data sharing via DMMPs could become viable.

Furthermore, E8 mentioned the issue of transparency and for example audits and accountability becoming more unclear. And in the case of disputes that E8 mentioned earlier, parties that have a dispute need to know where to go: is that the DMMP or the single data marketplace? E8 also stated that, compared to a single data marketplace, for a DMMP is could mean that they have to show and prove everything more heavily.

The DMMP-visual shows data marketplace A and B included to the DMMP whereas data marketplace C was excluded. E8 shared his vision on how this should be determined:

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."*

According to E8's vision, this comes back to what he mentioned regarding developing generic rules, which will be a challenge for DMMPs.

Data sovereignty in DMMP-context

Before showing the visual, E8 was asked about his view on staying in control over data for data providers in a DMMP-context:

E8: *"In my opinion, staying in control is making sure that nothing happens to your data that you don't want, or that nothing happens that is undesirable, or that harms your own interests. That the intention with which you make something available is not fulfilled. Mainly making sure, control, that it can have no adverse effects for yourself, or adverse effects for someone else."*

E8 further specified that these adverse effects could not only be direct, but also indirect, for example missed revenue of credits from the data that a provider supplies. E8 also mentioned that these adverse effects can not only apply to a specific data provider, but also to society as a whole. E8 compared this to the current situation with the large tech companies that have enormous influence on society.

The data sovereignty visual was showed with the four block and E8 was asked to tell about his view on them in a DMMP-context. E8 started that data ownership is always a tricky topic, because it is poorly determined by law. He continued that it often is more about who is responsible for data. In his opinion, this is in principle always at the organisation that provides it to the DMMP. According to E8 data ownership can be a problem, but at least it is often clear who is responsible for data.

Regarding data processing and usage, E8 mentioned that this will take place in the single data marketplace connected to the DMMP. This will be a bit more challenging, most importantly data providers will have to indicate via the DMMP what is allowed and not. E8 also gave some potential solutions to this problem primarily multi-party computation and anonymised data. Additionally, E8 saw bringing the algorithm to the data as a potential solution to overcome the data usage challenges for data sovereignty. The complementors as a third party in a transaction were data is analysed makes it also more quite tricky according to E8.

While discussing data access, E8 remarked that he assumes that data stay with the provider in principle, so not at the DMMP or single data marketplace. When data is not stored at the provider, E8 indicated that it could possibly be stored at a complementor as trusted third party. The complementor then provides a place where the provider can store the data and the consumer can access it. E8 also emphasised that when the complementor is working for the single data marketplace, proximity is even less, compared to when a complementor is linked to the DMMP.

Regarding proximity, while progressing in the conversation, E8 remarked:

E8: *"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."*

One remark E8 made was that he personally is not a data provider, so that is responding based on his experiences during his work with others. When asked to comment on which blocks are the most critical or difficult, E8 responded that data ownership is more of a legal question, i.e. which party is responsible. E8 did not think that that block was necessarily important. In contrast, E8 felt that especially data storage and data processing/usage were critical.

Lastly, E8 was asked about his view on governance in general to achieve data sovereignty in a DMMP-context. He concluded that he feels that it should be arranged in a cooperative model such as an association or foundation. Especially to maintain trust because according to E8 that is were it is all about in the end.

E8: *"But if you want to maintain trust, because that is ultimately what this is about, because you will only participate in it if you know that this is reliable, then it must also be reflected in the way in which you organize it together. I don't think it will work otherwise."*

E8 gave an example when a DMMP-operator is a stock-listed party, their goal will be to increase shareholder value, which is not operating in the interest of data providers automatically. However, E8 also emphasised that it should not be a governmental organisation either.

While asked how E8 sees such a platform emerging, E8 made an analogy with the rise of the internet. He explained that during the development of the internet, the structure was deliberately kept very open with a couple of organisations behind certain standards. In his view, if the internet would have be made by a commercial party it would have never become the success it is today:

E8: *"In my opinion, you should also look at the concept of data marketplaces this way. You are developing a new form of internet for data." … "After all, this [slide] still becomes quite centralized-looking very quickly, everything again within one party there. You should actually have a kind of 3D-slide of this, where you have an infinite number of DMMPs that are also in contact with each other. How do we organize that, and you name it. I think that's a really important one, because the idea of marketplaces sometimes becomes a centralized version again very quickly, and actually it's not."*

Closing

E8 had no further questions and was thanked for his contribution.

Expert reference	**E9**
Professional background	Data management expert at a global professional services firm
Date	20 June 2022
Duration	52 minutes

Introduction

E9 is an experienced professional working for a global professional services firm in a team working on national and European projects to enable better use of data. This relates to better access to data, timeliness of data and data quality. More specifically, E9 helps people within this firm to better understand and use data, but also to help with data protection and security topics. His team works for the Chief Data Officer.

In his working experience E9 has worked on several data sharing initiatives, for example when the firm was acquiring a company and data sharing was part of the take-over. E9 has also been involved with data sharing between different national offices and teams of this firm. In his position, E9 is also experienced with the regulations applying to which data can be shared and which cannot, for example because of professional secrecy laws.

B2B data marketplaces

When asked about his experience with B2B data marketplaces, E9 mentioned data marketplaces by Google and Amazon. However, E9 questioned whether data marketplaces are currently already generating a lot of revenue, or whether some of them are there as test balloons. He mentioned two-sided market problems as well:

E9: *"There was a study recently published and I think 80% of companies said they wanted to have data from competitors and only 10% said they were willing to share data with competitors. So there is a huge mismatch."*

After showing the B2B data marketplace-visual and asking which factors might influence data providers' decision to start sharing data or not. E9 replied with a counterquestion and asked which department of the data consumer would make the decision. E9 wanted to emphasise this because larger organisations are very homogeneous and there is often not one interface when talking about data consumers. E9 argued that before a data marketplace could work for involved parties, internally they need a lot of maturity, such as regarding processes and internal data management.

E9 sketched more background regarding this issue and shared a personal experience when his team was working to acquire external data. During this project, it was very difficult to find not only the data providers, but also the right potential consumers within the own organisation.

Next, E9 discussed the original question about factors influencing data providers:

E9: *"Simply put. You could sell reasonably, data, where the benefit of selling is greater than the risk of exposure would be. Obviously, the more damaging or sensitive data could be to you as an organization, the more expensive it's getting."*

According to E9, to be able to make up the balance, organisations must first start with doing an internal analysis of their data inventory and associated business cases. Only then an organisation can accurately estimate the impact of a potential data sharing action.

Data marketplace meta platforms

The visual of the DMMP was shown and E9 was wondering what the business model could look like of a DMMP and its associated single data marketplaces. A brief description was given with some examples, but this was kept brief because the interviewer did not want to steer the respondent and business models of DMMPs was not the main scope of the research.

Next, E9 was asked about his view on advantages and disadvantages of DMMPs from the perspective of data providers. E9 mentioned that he saw the advantage that data providers do no longer need to cater for all the different single data marketplaces. Additionally, data providers could be able to reach a broader audience of data consumers, both geographically and topically. Furthermore, according to E9 the single point of contact could make customer relationships easier. Lastly, E9 could see added value by added services by a DMMP for data providers, to take over some of the work associated to data sharing of the shoulders of data providers:

E9: *"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 whatever."*

Regarding disadvantages, E9 argued that a DMMP could mean a big vendor lock-in for data providers, because there is one powerful independent platform. Additionally, in the scenario that a DMMP increases supply of data, it could also diminish data value for data providers. However, a DMMP could also provide more unique data according to E9:

E9: *"I believe, spreading an increasing awareness and availability of your data diminishes its value because value is it's partially driven by availability and accessibility, but also by uniqueness. So those two are contradicting each other. One would need to see in practice which one prevails."*

Uniqueness agreements could maybe help to overcome this issue, for example that data providers limit their data to a fixed number of data consumers according to E9.

When asked to compare single data marketplaces and DMMPs, E9 mentioned the winner-takes-all scenario. He thought that the party that offers the most convenient platform with the most aggressive approach will in the end get the data flowing in and out. Regarding single data marketplaces versus DMMPs, he also commented:

E9: *"The biggest one which does not necessarily mean it's the first one or not. And I think it's independent of it being a meta-platform or a marketplace. Because I personally think they could switch between each other. The marketplace could easily become a meta-platform you've seen it with Amazon."*

This platform dynamic could not only happen between single data marketplaces and DMMPs, but also between marketplaces and complementors with an envelopment example:

E9: *"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 trust services and authentication services, whatever. If there's high enough value they can be reintegrated into such a platform."*

In a similar line of thought, E9 also argued that DMMPs could take over single marketplaces and could serve their markets as well. As a response to these examples, E9 was asked if he thought that a particular governance structure could maybe overcome the risk of take-over of other platforms. E9 was sceptical if this could work out, because he had seen many times that in the end everything is coupled to commercial aspects an capitalistic systems, which are inherently non-democratic. He also mentioned blockchain, which promised to be distributed and democratic decoupled from governments, but is in the end not much different than gold.

Data sovereignty in DMMP-context

E9 was asked what in his view control over data means for data providers in a DMMP-context. E9 responded that in his view there is no technology solution for that, but that it is a trust issue:

E9. *"Being in control means you have trust in a system that its participants and constituents will be using the thing, the data, in only acceptable ways, like pre-approved ways or pre-agreed upon ways."*

He also discussed the issue that data is highly replicable so that there can be made unlimited copies. Even with control over the input and output, E9 still questioned whether that would be enough, as he explained that there will always be ways around. According to E9, trust in governance systems means ensuring trust in every participant. Furthermore, E9 emphasised the role of transparency and being clear to establish trust. For platforms, this could for example meaning to open up the technology and showing how it works. This could work similarly for the decision making as he explained:

E9: *"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. 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."*

In his answer, E9 also discussed smart contracts and they promise to do everything automatically, but that the human element is missing in that solution. According to E9, this human element is crucial in developing trust relationships.

Next, the four blocks were discussed on the visual for data sovereignty. E9 responded that data storage and access are closely related. Storage of data, which could very well be in a cloud environment, means that there are already access controls, but that it could be wise for data providers to implement additional access controls for this scenario. E9 also saw value in limiting and controlling the processing systems:

E9: *"So you would eventually not give access to the raw data, but to the distilled insights or to the aggregated whatever. So that the risk of losing control of the actual underlying data is reduced."*

However, E9 also warned for the risk of re-discovery of underlaying data of distilled insights and gave examples of research that has proven that outputs can often be tracked to the original data. Regarding data ownership, for E9 it follows from all the other three blocks. When asked which block would impact the decision of data providers regarding data sharing, E9 responded that it would be data access in his opinion.

E9 also briefly discussed data storage again and explained that it could be very critical for very specific organisations, for example because of ISO certifications or legal obligations.

Lastly, E9 was asked about whether he thinks that governance can help to enhance data sovereignty for data providers in a DMMP-context. E9 thought that it could help, but argued that it has not been figured out yet why we should even use data marketplaces at all, mainly because it is not known yet which data has to be shared for which purpose:

E9: *"And we can have like endless discussions about the pitfalls and the opportunities of such platforms and ecosystems. Really doesn't matter, because why should we do such platforms has not been answered. And in the meantime, it stays an academic question. And don't get me wrong, it's a totally interesting one. Just from a practical perspective, it's not relevant. I mean, it will be relevant eventually. But it's the third and the fourth step in a journey we haven't practically begun"*

Closing

After the last question of the protocol was discussed, there was a short after talk and some zooming in on E9's last comment above. E9 mentioned that currently organisations are still very slow to adopt these kinds of platforms, because they simply do not have to yet. For the future challenges such as climate change and a feeding growing world population use of data could be needed and valuable. However, according to E9, currently a lot of organisations are still focused more on the short term to grow profit and shareholder interest. E9 also discussed that attention from politics and public debate could change behaviour of individual companies to move the use of data from increasing profits to solving societal problems and better serving employees.

Expert reference	**E10**
Professional background	Data expert and research engineer at a German research institution
Date	21 June 2022
Duration	37 minutes

Introduction

E10 is working as a research engineer at a German research institution focused on testing and modelling of materials for a range of strain rates. In this organization, E10 is working to improve collaborations with other organisations regarding data sharing. In the materials science, data is very heterogeneous and sometimes scarce as E10 explained. To improve the exchange and availability of data, E10 is currently working to set up a materials data space. In total, E10 has three years of experience with data spaces of which one year with the actual implementation of the materials data space.

B2B data marketplaces

E10 was asked about is personal experience with data marketplaces. In response, E10 elaborated on the materials data space and compared it to a marketplace model. Next, the B2B data marketplace visual was introduced and E10 was asked which factors would influence the decision of data providers to actually share data using a data marketplace.

First of all, E10 mentioned that sharing data via a marketplace could be done to contribute to the community. Additionally, E10 stated that data providers could also be motived by economic value that could be captured by trading data. However, E10 thought that providers could also be scared to share because they might be giving away company secrets or critical business information. Additionally, E10 mentioned that a factor for the decision to share could also be that it means that providers can also consume data, so that the data providers could also be data consumers.

Furthermore, E10 believed that large-scale data marketplaces could lead to new emerging business models. Additionally, E10 gave the example of Wikipedia which is completely different than old encyclopaedia publishers. Looking into the future, E10 believed that it might be necessary for organisations to participate in fairly open data marketplaces to stay relevant in the future. He sketched a scenario:

E10: “So I think data marketplaces open new business models. Fairly open data marketplaces is the business model of the future. And if you don't, as a data provider, if you don't follow this business model in the future, you won't be able to make a lot of money with your company.”

However, E10 was also asked about potential factors that bring risks or challenges. E10 replied that data providers might sell data to a consumer but might be unsure if this consumer is actually doing what is promised up-front. For example:

E10: *“And on the other hand, I couldn't, maybe that's a small aspect, I couldn't control, what he is using the data for? For example, if I provide the data not for military purposes or something, I cannot really control what he's using it for.”*

Data marketplace meta platforms

The DMMP visual was introduced and E10 asked why data consumers would not access the DMMP, for example to find the cheapest data marketplace. E10 was thanked for this question and it was explained that this could in theory very well be the case, but that for this specific project a set-up was chosen were data providers are the providing via the DMMP and consumers stay customers of the single data marketplaces.

When asked for potential advantages for data providers, E10 pointed out that it could mean that data providers can more easily provide to many different data marketplaces, decreasing the workload. According to E10, a DMMP could potentially also help to discover new (single) data marketplaces for data providers. A larger audience of potential consumers could also mean more financial revenue for data providers according to E10.

When asked about potential disadvantages of DMMPs for data providers:

E10. *“. It would be even more difficult to track down the use of my data. Yeah. Basically, the data sovereignty aspects.” … ” There's another layer which has to track the whole provenance of the data usage. “*

Additionally, E10 mentioned the additionally technical complexity of DMMPs and the difficulty to technically implement it and connect all the data marketplaces.

Next, E10 was asked to compare single data marketplaces and DMMPs and to give the main changes for data providers. E10 noted that is he was a data provider, he would prefer a DMMP because it is more convenient and easier to distribute data to more consumers. However, according to E10, using a DMMP would also come with concerns about where the data actually goes and whether a provider is able to track the usage of it.

After E10 introduced his view on challenges and benefits of DMMPs, he was asked if he could think of any solutions to his concerns. E10 mentioned that a solution could be to not let the raw data go to consumers, but to bring the algorithm to the data and only share the outputs with the actual data consumer.

When asked about the governance, E10 explained that in his view there are two configurations: either there is a centralised instance that controls what is happening, or there is a more de-centralised solution with for example blockchain technology to log all the data.

Data sovereignty in DMMP-context

In this part of the interview, E10 was first asked what staying in control over data means for data providers in a DMMP-context. E10 mentioned that defining usage policies are important for data providers to define how their data can be used and for which purpose. E10 elaborated on what he called boundary conditions:

E10: *"If I want to provide my data only for a week so I can control it, you can use it only for this week. And the same if I say you can use it one hour per day. So I control."*

Next, the visual with the four blocks related to data sovereignty was showed and E10 was asked to share his view on them related to potential risks and opportunities in a DMMP context. E10 replied that it depends very much on the configuration of the DMMP, and even the single data marketplaces as well. For example, when DMMP only enable transfer of metadata, it is very different compared to when the raw data itself goes through the platform as well according to E10. E10 emphasised that it is about a combination of rules and technical implementations.

The influence of the configuration of the DMMP applies to data access as well according to E10:

E10: *"The data access also really depends on what is meta-platform is able to, what the principles are here to access the data, either really provide the data and the data consumer can download it somehow or if the data access is really like I mentioned via algorithm to data."*

Regarding data processing, that should not be that much different compared to data access according to E10.
Data ownership was discussed as well and E10 highlighted that the DMMP should always be aware who the owner of the data is, and that this also remains traceable when data is being transferred from DMMP to data marketplace to data consumer.

Next, E10 was asked if there was any block which was the most critical for data sovereignty in a DMMP-context, or if there was something missing. According to E10, if data usage and processing control was figured out, he would not be too scared to share data with a DMMP. E10 would then even trust the DMMP enough to be willing to store the data at the DMMP as well.

Lastly, E10 was asked if he in the end thinks that governance of DMMPs can help to improve data sovereignty for data providers. E10 replied that it would help if there is a separate institution that controls a certification system, to make sure that not just every organization can enter the platform. According to E10, this would create trust. Additionally, E10 mentioned that a de-centralised solution where data stays at the provider would be better than just uploading everything to the cloud, although a de-centralised configuration could impact efficiency for analysis and other additional services on the platform.

Closing

E10 was thanked for participating in the interview and his contribution to the research. As there were no further questions, the interview was ended.

Expert reference	**E11**
Professional background	Developer and semantic web expert, data sharing initiatives expert, data engineer
Date	24 June 2022
Duration	47 minutes

Introduction

E11 has a background in information and communication technology and is specialised in topics around the semantic web, data ontologies and has contributed to several European data sharing and data marketplace initiatives. E11 is currently involved as data engineer to develop an app store for a large pan-European data sharing initiative. Additionally, E11 works on the development of a federated data catalogue for another data sharing initiative. At the moment of the interview, E11 was working at a large German research institution.

When asked about his experience with B2B data sharing, E11 provided an example of mobility-related data that is shared in a data space in Germany. In this example, car companies and the telecom industry to share and use data related to mobility.

B2B data marketplaces

E11 was asked to elaborate on his experience with B2B data marketplaces and explained the use and purpose of the data marketplace that is included in a large pan-European data sharing initiative he is currently working on. E11 also mentioned that this includes the use of privacy-preserving technologies and methods for data providers. Currently, a lot of data sharing projects are driven by a particular use case or problem in a specific industry sector as E11 explained. One of the current challenges is getting industry parties on board of data sharing initiatives:

E11: *"As we said, the problem, for the data market at least, for data trust we have difficulties to link it with the industry since those people say: yes but like what is the benefit there? So we need to find a way to assure that people say yes we have a great benefit with it and this will help you and will help the people."*

This expert noted that the data marketplace which is part of the larger initiative might also help to make the potential more visible for parties that are currently no involved.

Next, the B2B data marketplace visual was introduced and discussed. E11 was then asked to share which factors he thinks are influencing the decision of data providers to share data or not. E11 responded that the main reason is that parties don't want to lose control over their data. Additionally, potential data providers often have their own methodologies that do not align to the data marketplace. On the data consumer side, there are privacy concerns as well according to E11.

E11 proposed a few potential solutions as well, for example methods where the raw data itself is not shared, but just the analysed output or just facts. Another possibility could be the use of encrypted or anonymised data according to E11. For the first solution, it could be that the data consumer indicates which information need there is, and the platform only returns the answer to that particular request, without sharing the underlaying data. To achieve this, E11 emphasised the role of complementors and complementary services, as these services will do the transformation of raw data into insights.

Data marketplace meta platforms

For the second part of the interview, the DMMP-visual was introduced and explained. E11 asked if the DMMP was an additional layer between data provider and existing data marketplaces and this was confirmed.

E11 thought that a DMMP could be beneficial for data providers, because it could become an additional protection layer to control their data on the separate single data marketplaces. He also mentioned the role of metadata for DMMPs, as this can also help to control the data usage. For example, to restrict data usage to particular geographical areas. E11 also mentioned the role of semantic web ontologies or vocabularies. This expert also gave the example of ODRL where ontologies are used for privacy control. All in all, E11 thought that an additional layer could be beneficial for data providers to control data usage.

In response to his answer about semantic web ontologies, E11 was asked to elaborate further on this topic. E11 mentioned that metadata can help to find out which parties to avoid when sharing data. Additionally, this language is machine-readable as well, which could help to automate things, for example to automatically determine whether a transaction is allowed based on the machine-readable rules.

Next, E11 was asked about his view on advantages for data providers considering the DMMP. E11 indicated that it could help to automate transactions when the DMMP uses a common semantic language, and improve inter-operability as well.

Data sovereignty in DMMP-context

Next, E11 was asked what he thinks staying in control over data means in a DMMP-context for data providers. E11 started that organisation are always the most afraid of sharing raw data. To overcome this, E11 mentioned a technique where small bits of noise are added to the data to protect the raw data, which is a privacy preserving methods. For personal data, E11 explained that it is sometimes necessary to leave out particular sections of the data to protect the subjects in the dataset.

According to E11 this could be combined with the solutions he mentioned earlier, such as only sharing insights or facts, or giving an answer on a very specific request. Furthermore, E11 also noted the use of applications where users access the data in a protected environment. These application could apply both to data providers as well as data consumer according to E11.

Before asking the next question, the data sovereignty visual was showed including the four blocks and E11 was asked to give his view on them in relation to data sovereignty for data providers in a DMMP-context. Regarding data ownership, E11 mentioned that the problem is often that it is sometimes unclear which parties have owned the data. However, for the DMMP-context in particular, E11 thought that data ownership is not the biggest issue, as this is often considered to be the providing party.

Next, E11 emphasised that for data sovereignty a high level of detail is necessary for access and usage control. For access control this could for example be only access to particular parts of the data or for example a fixed number of times a consumer can access. According to E11, data access, data processing/usage and data storage are important for data sovereignty.

Regarding data storage:

E11: *"But also, if we think about the data storage, the location of the data, how we storage, how we storage, where we store it. It is sometimes like as you can see, that we have a problem in Europe to store data outside of Europe."*

When E11 was asked which block or blocks were the most critical, E11 started with data storage, because it is key that storage is secure and that it cannot leave its supposed location. He then mentioned that data access and data usage are not less important, but that data storage is the starting point.

E11 also mentioned governments and law and that this brings requirements for how the relation with data consumers is documented. When there are disputes, the governments can help in the sense that there is a possibility to go to court. In response to this answer, E11 was also asked again about his perspective on rules of the platform. He explained that these can certainly help to stay in control as data provider, as it can help to stick data consumers to the agreements. However, E11 also explained that data providers will always have more control over data stored on premise compared to stored at the DMMP.

Closing

Lastly, E11 was asked if there was anything he would like to add before ending the interview. The expert mentioned the IDS- and Gaia-X-initiative to take a look at, in particular for the certification of connectors and how the data space is used within those initiatives.