Is data really

The new OIL?

An assessment of the viability of data marketplace business models

TU Delft

ALEXA PETIT

Is data really the new oil?:

An assessment of the viability of data marketplace business models

Master thesis submitted to Delft University of Technology

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in **Complex Systems Engineering and Management (CoSEM)**

Faculty of Technology, Policy and Management

by

Alexa Petit

Student number: 5184150

To be defended in public on August 31, 2021 at 15.00h

Graduation committee

Chairperson	: Dr. ir. G.A., de Reuver, ICT
First Supervisor	: Dr. ir. G.A., de Reuver, ICT
Second Supervisor	: Dr. rer. pol., C.,Werker, ETI
Advisor	: Dr. H.A. Ofe, ICT

Preface

Dear reader,

In front of lies my master thesis titled *"Is data really the new oil? An assessment of the viability of data marketplace business models".* This thesis is my final work for completion of the Complex Systems Engineering and Management (CoSEM) master program at the Technische Universiteit Delft. From the initial research proposal to this final product, it took from March 2021 until August 2021 to complete. Starting the research process I had little knowledge of what a data marketplace is and what value it could have for society. During the past six months, I learned a lot about this topic, but also about doing research and my personal interests. To quote Aristotle, I realized that "*the more you know, the more you know you don't know".*

I would like to express my gratitude to the persons that guided me through the process. First of all, Hosea Ofe, my advisor, thank you for our many meetings. Our discussions and your ideas helped me to proceed when I did not know how to myself anymore. I greatly appreciate your feedback and motivating words throughout the whole process. Also, thank you for involving me in the TRUSTS team of the TU Delft. It was very interesting and valuable to take part in the international meetings and experience how research on data marketplaces is conducted on a European level. Secondly, Mark de Reuver, thank you a lot for your supervision. You were always willing to read through my document and gave me structural and detailed feedback to improve my work. I am thankful that I could always schedule a quick online meeting to run my ideas by you. Thirdly, Claudia Werker, thank you for your valuable feedback and for pushing me to make the most out of it. Additionally, I would like to thank the employees of VETRI, Databroker, and the DX Network for participating in this research. I really enjoyed getting to know your businesses and I truly admire your ambitions.

Furthermore, I would like to thank my family and friends. Thank you for discussing the problems I ran into with me and for endlessly supporting and encouraging me. I feel lucky to be surrounded by so many extremely kind and caring people.

I hope you enjoy reading this thesis!

Alexa Petit

August 2021

Executive summary

Data has often been stated to be “the new oil”. This is said because both oil and data can be very valuable but only after they have been pre-processed. Among other things, a challenge for creating value from data lies in realizing data exchange. This is what data marketplaces aim to do. These marketplaces are therefore assumed to be a prerequisite to live up to the premise that data is as valuable as oil was in the 18th century. The problem is that the viability of the business models of data marketplaces is not yet clear.

The number of shutdowns during the past years implies that the business models are not viable and that there are obstacles that should be overcome to create a data marketplace that is robust in the long run. Therefore, it is useful to get more insight into the obstacles that data marketplaces need to deal with and how these relate to their business models. In this research, we defined viability to consist of two elements (1) the ability to facilitate the main activity of data marketplaces: data exchange between providers and buyers, and (2) being financially stable to keep the activities going in the long run. We believe both items are needed to form a viable business model. An ‘obstacle’ is defined as “something that impedes progress or achievement” (Merriam-Webster.com dictionary, n.d.). We, therefore, defined a data marketplace obstacle as *something that impedes the progress or achievement of data marketplaces.*

The identified research gap we identified was the lack of an explanatory model that relates data marketplace obstacles to business model choices to assess the viability of data marketplace business models. This model would explain why it is difficult to create a viable data marketplace business model based on the obstacles that we can identify. This led us to the following main research question: *“How can we use the relationships between business model choices and obstacles to assess the viability of data marketplace business models?”.*

To answer this question we followed a multiple-case study approach using a framework created by Eisenhardt (1989) that describes the process of building theory from case study research. We started by explorative theorizing of the conceptual background. We found that the obstacle categories for data marketplaces are trust, lack of governance mechanisms, privacy and control, product quality and product description, matching data providers and data buyers, data pricing mechanism, and ‘other’ that do not fall within a specific category. We also found that there are five marketplace requirements for efficient marketplace operation: safeness, no repugnance, provenance information, thickness, and no congestion (Roth 2002,2007). Safeness means that there are no incentives to misrepresent or undertake strategic action which will lead to reduced efficiency. No repugnance means that there are no social norms or legal restrictions that limit the use of pricing as an allocation mechanism. Provenance information means that there is no information asymmetry; there is enough information available to assess the quality of the good traded. A “thick” market means that there are enough opportunities to trade with other participants on the market. No congestion means that there should be no reason that transaction times on the marketplace are slowed down. We assigned the marketplace requirements to the identified obstacle categories. We claimed that trust and governance affect safeness, privacy and control affect no repugnance, product quality and product description affect provenance information, matching and other (unspecific) affect thickness, and pricing affects no congestion. All requirements influence the desired outcome: a viable data marketplace. Because we believed the business models choices to relate to the impact that the obstacles could have, they indirectly affect the viability of a data marketplace business model. Putting all constructs in a framework gave us a theoretical base for the remaining part of the research.

We then selected three data marketplaces by theoretical sampling. This resulted in the selection of Databroker (DAO) as IoT/B2B data marketplace, VETRI as personal data marketplace, and the DX Network as B2B data marketplace. Each case study was subjected to an adjusted version of the Business Model (BM) Stress Test (Bouwman et al., 2018), using obstacles instead of uncertainties and using the business model choices from the taxonomy of van de Ven (2020). The results from the Stress Tests were reflected upon by looking into literature and talking to data marketplace experts. We realized that the relationships between business model choices and data marketplace obstacles were not very comprehensive and we should adopt a broader perspective. The assessment of the true viability of a data marketplace business model thus goes way beyond just the relationships between business model choices and obstacles. It is concerned with how well a data marketplace can live up to the marketplace requirements regarding choices that cannot be directly found in a taxonomy, such as the dynamic capabilities and strategy of a data marketplace. It is also concerned with external factors that the data marketplace cannot even have any influence on, like governmental rules and regulations that concern any form of data trading. So, based on the result of this study we cannot make conclusions about business model choice X, direct impact on obstacle Y, and indirect impact on the viability of a data marketplace. However, we could hypothesize on the importance of the context in which business model choice X is made for overcoming obstacles and assessing and/or increasing viability. Or, in the words of DaSilva & Trkman (2014), how a data marketplace can strategically prepare itself by developing dynamic capabilities based on the business model choices they make. The practical implications of our results include the considerations that a data marketplace owner can take into account when making business model choices. A selection of our findings include that a choice for an enterprise data marketplace can be an opportunity to build trust and allows for more sensitive cases of data exchange, data processing and/or analytics tools in a data marketplace forms an opportunity to increase perceived product quality, a decentralized blockchain-based platform architecture allows for a trust-less data marketplace, a data marketplace is dependent on the capabilities of its providers may the act as their data source, cryptocurrencies can be used to fund a data marketplace but its user interface should remain simple. This is also true for the use of a smart contract, which allows for automatic (micro)payments and can give a warranty on data quality, but is not transparent and can be difficult to understand.

Because the recommendations we can make based on the results of this research go beyond making certain business model choices, we formulated a more general proposition: *"The better the marketplace requirements are safeguarded by considering the business model strategy for a business model choice (from an obstacle perspective) and the influence of the external factors, the higher will be the viability of the data marketplace."* Thus, to increase the viability of data marketplaces we recommend safeguarding the requirements of safeness, no repugnance, provenance information, thickness, and no congestion. To safeguard safeness we recommend creating a trust-less environment for as far as possible. Also, we think it would be beneficial to stress the trustworthiness of a data marketplace in marketing and communication. Furthermore, a data marketplace should anticipate the upcoming regulations and profit from its implementation when possible. To safeguard no repugnance data marketplaces could consider to sell pre-determined insights or provide consulting on the decisions that can be made based on the data. This allows for more control over the data exchange. This same argumentation can for safeguarding provenance information. Product description and product quality will be less of a problem when insight are sold instead of raw data. Another option is the development of an efficient method of metadata generation that also considers the data requirements of a buyer. To safeguard thickness the main point to consider is the pool of participants. Whether working with a narrow pool or big pool of participants for making matches, a data marketplace should ensure that their business model leads

to a high willingness to participate and a high ability to do so among its participants. To safeguard no congestion the only recommendation that we can make is staying up-to-date on the newest developments in the research field of data pricing mechanism.

Nonetheless, data marketplaces still have to overcome many obstacles and deal with many uncertain external factors. If we assume that these marketplaces are a prerequisite to fully profit from the value of data, data cannot be claimed to be the new oil. Not yet.

Content

List of figures

List of tables

CH 1. Introduction

In 2006 Clive Humby stated the following:

"***Data is the new oil.*** *It's valuable, but if unrefined it cannot really be used. It has to be changed into, gas, plastic, chemicals, etc to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value."* – Clive Humby, 2006

Since then, the statement that data is the new oil has often been quoted.

The follow-up part of the quote, saying the value of data is not in the collected data itself, can be connected to the Information Pyramid. This Information Pyramid shows that data can lead to information and knowledge (Rowley, 2007). The pyramid is depicted in figure 1.

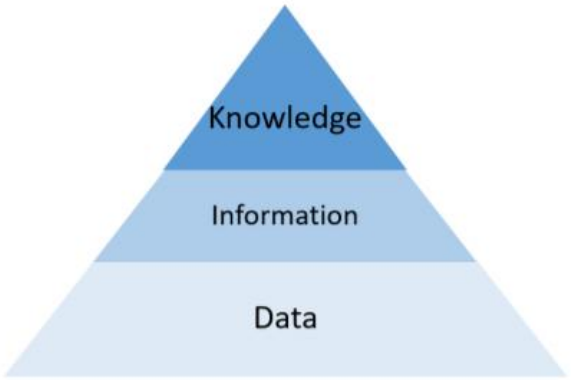

Figure 1: The Information Pyramid (Rowley, 2007)

According to (Ackoff, 1989) the three concepts in this pyramid can be defined as follows:

Data consists of a collection of symbols. They represent the properties of objects.
Information is processed data and can provide answers to 'when', 'what', 'where, and 'who' questions.
Knowledge describes the interpretation and use of information and provides answers to 'how' questions.

Like oil, the data itself can be seen as a commodity. A commodity can be valued on the dimensions of cost, time, and quality and can be exchanged, sold, and traded. In other words: there is a market for this data. Many organizations generate data as an outcome of their activities. They do not process, integrate and use all the data for strategic purposes. An illustrating example is an oil platform containing 30.000 sensors, which uses less than 1 percent of the collected (IoT) data (Manyika et al., 2015). Following this example, 99% of the remaining data that is collected could be used by others. This data – and the information and knowledge generated from it – nowadays can form a source of innovation and can therefore be very valuable.

However, data commodification is a complex sociotechnical practice (Aaltonen et al., 2021). According to Aaltonen et al. (2021), the data that will eventually have to become a commodity will start as a data token, then will need to be processed to a data-based object and then recontextualized to a data commodity. The data tokens result from producing data by encoding incidents or events that occur in an organizational context. To form a data-based object the tokens are aggregated and then validated when industries agree upon certain metrics. To form a data commodity, these objects are aligned to business objectives and market value can be gained.

Data marketplaces are concerned with the process of data commodification and the exchange of data among different parties to achieve this market value. The data can be exchanged through the platform the data marketplaces provide. We, therefore, see these marketplaces as a prerequisite to live up to the premise that data is as valuable as oil was in the 18th century. The problem is that the viability of the business models of data marketplaces is not yet clear. We assume that this is related to the socio-technical complexity of data commodification and data exchange. Therefore, we deem it interesting to explore this further.

In this chapter, we aim to introduce the research area of data marketplaces. We will start by describing the problem identification. We will then identify the knowledge gap, which will lead us to the main research objective and main research question.

1.1 Problem identification

In general, researchers agree on the potential value that data marketplaces have. For example, some even state that *"In the future data marketplaces will be equivalent to other electronic commerce platforms such as Amazon or eBay"* (Lawrenz et al., 2019, p. 95). Or *"data marketplaces are set to expand the breadth and depth of data available today by at least an order of magnitude in the very near future, as Wikipedia did for encyclopedic knowledge"* (J. Smith, 2018a).

However, this value creation has thus far not lived up to its expectations, and its potential can be called unfulfilled (Koutroumpis et al., 2017). A lot of players established in the early 2010s could not live up to it and have failed during the last ten years (Cosgrove & Kuo, 2020). In figure 2 the timeline of the shutdowns of these data marketplaces is depicted. Although it is not unusual that a certain marketplace business model fails (Yoffie et al., 2019), you could say that generally speaking there should be at least one marketplace for a certain commodity that is successful. It is unusual that so many have tried, probably using different approaches, to realize a marketplace for the same commodity (data) but had to shut down and thus have not succeeded. The amount of shutdowns and the current absence of a well-established data marketplace implies that its business models are not viable and that there are obstacles that should be overcome to create a data marketplace that is robust in the long run. The obstacles that are encountered by the data marketplaces possibly have to do with the unique characteristics of data as a commodity (Cosgrove & Kuo, 2020):

- Data is highly adaptable, which makes traceability to its original form extremely hard
- Data is highly substitutable – there are many options for sourcing data
- Data is diverse, which makes it complex to engage with
- Data is often 'raw' – effort should be put in to make it valuable
- Data is digital and easy to copy, which makes stealing easier

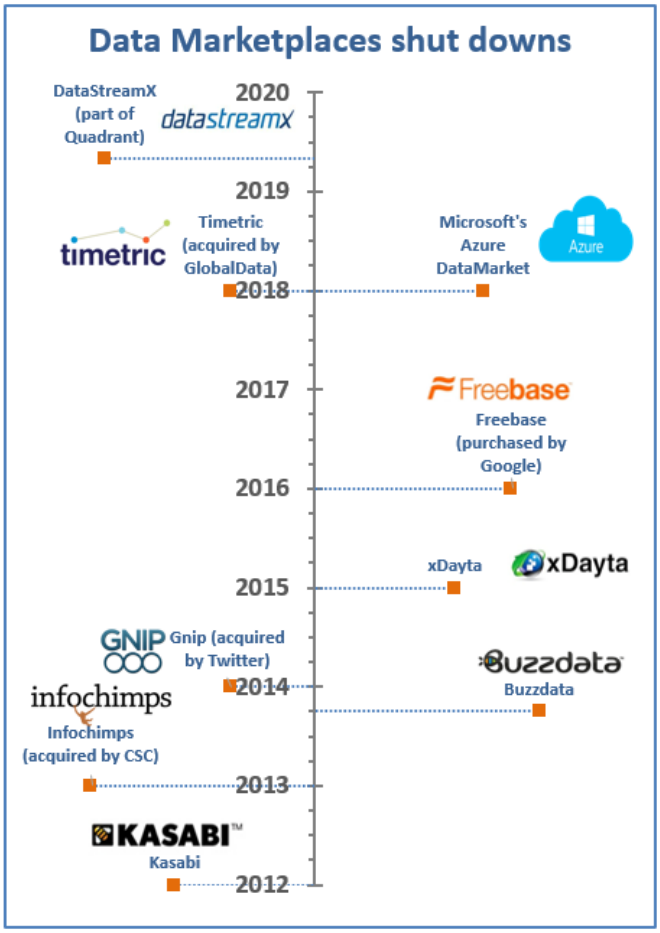

Figure 2: Overview of data marketplaces shutdowns (own illustration)

Marketplaces normally thrive on standardization, a clear value of the goods being sold, and the ease of comparing these marketplace goods. Referring to the characteristics of data described above, marketplaces that focus on data exchange cannot copy the business models that 'normal' e-marketplaces employ

1.2 Research gap

Thus far we have observed two things. The first thing is that data marketplaces could be very valuable for data-driven innovation. The second thing is that a lot of data marketplaces have shut down during the past years. To achieve value creation data marketplaces should not have to shut down – they should have a viable business model that enables them to perform their activities. This gives reason to investigate what is known about the business models of data marketplaces. We are interested to see if the literature on data marketplaces gives reasons for the shutdowns and says something about the viability of the business models or if obstacles are mentioned that hinder this. We will first explore this further. This will lead us to a research objective and the main research question.

1.2.1 Literature search

We collected literature through Google Scholar and Scopus. To find the necessary information, general keys such as "data marketplaces" or "data markets" were combined with concepts such as "business models", “strategies”, “taxonomy”, “case study”, “viability”, and “obstacles”. Some articles were not found directly through hits on the used keywords but through the snowballing effect, using the bibliography of the articles to find other relevant titles on the subject. First, an initial selection of roughly 25 papers was made based on the indicated relevancy in the search engine and the title. From reading the abstract and scanning the heading and content of the articles 16 articles were selected. The articles were selected because they clearly addressed specific knowledge on data marketplace business models or had explicit sections about data marketplace obstacles. The obstacles however are not always explicitly named “obstacles” but were sometimes referred to as “challenges”, “difficulties”, or “barriers”. The final selection of articles used can be found in Appendix A.1.

1.2.2 Literature on data marketplaces

The explorative literature research gives insight into the aspects that are already known about data marketplaces and their business models. There are three broad areas in which research is conducted. The first research area is that of data marketplace business model classification. For example, Schomm et al. (2013) and Stahl et al. (2014) make an overview of the categories and dimensions of data vendors and data marketplaces and identify the data marketplace trends. Koutroumpis et al. (2017) develop a conceptual market design framework and also identify the requirements for data marketplaces. Spiekermann (2019) defines what a data marketplace is, and creates a taxonomy of data marketplaces afterward. Fruhwirth et al. (2020) also create a taxonomy and distinguish four business model archetypes. Van de Ven (2020) extends the existing taxonomies by including a multi-stakeholder perspective.

The second area that is mentioned in the literature concerns the challenges, difficulties, barriers, or obstacles that data marketplaces face. No academic literature provides a complete overview of the obstacles. We only encountered articles where a selection of the obstacles is described in a certain paragraph or a where a certain obstacle is implicitly mentioned throughout the text. For example, Spiekermann (2019) devotes a few paragraphs to challenges and trends in the data marketplace environment. Lawrenz et al. (2019) also touch upon challenges associated with data trading in a data marketplace. G. Smith et al. (2016) identify the open data adoption barriers. Nikander & Elo (2019) discuss the market failure of a data marketplace. For the sake of simplicity, all these concepts will be grouped under obstacles. In the remainder of the report, we will thus only use the concept of obstacle.

The third research area concerns the technological solutions that could deal with those data marketplace obstacles. For example, Sharma et al. (2020) speak about an approach to integrate blockchain technology in a data marketplace to ensure trustworthiness and independence. Azcoitia & Laoutaris (2020) develop an algorithm to overcome the complexity of pricing digital assets. Hayashi & Ohsawa (2020) develop a matching system that visualizes the needs of users, and Perera et al. (2017) review how IoT data can technically be unlocked to sell on data marketplaces. Mišura & Žagar (2016) develop a model of an IoT data marketplace that could create a positive feedback loop. As a final example, Parra-Arnau (2018) examines a mechanism that can give users control over their personal data.

In summary, the literature contains taxonomy developments, descriptions of obstacles, and technological solutions that should help overcome the obstacles.

We lack knowledge about how the research area of taxonomies and obstacles and are connected. A taxonomy describes a data marketplace's possible business model characteristics. However, a taxonomy is only descriptive in the sense that it gives an overview of possible business model characteristics a data marketplace could employ. It does not indicate how the choices that data marketplaces make for their business model, from these characteristics, relate to certain obstacles that should be overcome and contribute to a viable business model. In other words, there is no information on the relationships between business model choices and obstacles.

We established a clear research gap: On the one side, researchers have mapped data marketplaces by developing classifications and taxonomies. On the other side, obstacles and technological solutions for those obstacles are researched. However, no connection between those sides has been made.

1.2.3 Importance of filling the research gap

Now that we have established a research gap, it should be stressed why it is important to fill this gap and make the connection. This is because we want to contribute to diminishing the contrast between data marketplace potentials and the current data marketplace reality:

(+) In principle, data marketplaces have a lot of potentials. Data marketplaces would enhance market efficiency, resource allocation efficiency and improve the match between supply and demand (Koutroumpis et al., 2017). This would generate value for data buyers as well as data sellers. This is because data marketplaces offer a market mechanism to exchange and price the data and match buyers to sellers. In this way, it would become easier to leverage the abundant amount of data for economic benefit (Fruhwirth et al., 2020).
`**(-)** However, practically this potential is not fulfilled as most data marketplaces initiatives have often failed or struggle to scale. Very little data is exchanged via the data marketplaces (Koutroumpis et al., 2020). In few instances, where data is shared, it is often based on a non-profit business model or via bilaterally negotiated contracts (Koutroumpis et al., 2017). While the Internet provides the opportunity to employ innovative business models, data marketplaces do not seem to profit from this. The adopted business models are limited and mainly focussed on secure revenue streams and hierarchical organizational settings limited from the broader exposure that could be tapped from a multi-lateral data marketplace (Stahl et al., 2017; Koutroumpis et al., 2017).

The contrast implies that the business models of data marketplaces are not viable. This is the opposite of what is wanted: a viable data marketplace business model that enables a data marketplace to live up to its potentials.

We define the viability of a data marketplace to consist of two elements: (1) the ability to facilitate the main activity of data marketplaces: data exchange between providers and buyers, and (2) being financially stable to keep the activities going in the long run. Both items are needed to form a viable business model.

The potential lack of viability can be explained by the obstacles that data marketplace providers are facing that prevent them from fulfilling their role as an ideal matchmaker between data buyers and data providers. An 'obstacle' is defined as "something that impedes progress or achievement" (Merriam-Webster.com dictionary, n.d.). A data marketplace obstacle is therefore defined as *something that impedes the progress or achievement of data marketplaces.* As currently very little data is exchanged, and many data marketplaces are shut down it seems that these obstacles could hinder the viability of a data marketplace business model.

A data marketplace's business model choices would need to deal with the obstacles. In other words, making a certain business model choice might affect the impact of a certain obstacle. The "right" business model choice would then in its turn increase the viability of the business model. Therefore, it is useful to get more insight into how business model choices relate to the obstacles. In this way, it can be assessed if, under the existing obstacles and choices that can be made, data marketplace business models are viable and/or certain choices can increase the viability of these business models.

In summary, to close the knowledge gap, this research aims to create an explanatory model that relates data marketplace obstacles to business model choices to assess the viability of data marketplace business models. This model would explain why it is difficult to create a viable data marketplace business model based on the obstacles that are identified. It should also indicate what business model choices might contribute to overcoming the obstacles or at least lower its impact. Because creating such a model has not been done before it will be an add-on to business model theory on data marketplaces.

1.2.4 Research objective & main research question

To make the identified problem and knowledge gap explicit, the **research objective** is formulated as follows.

"Create an explanatory model that relates data marketplace obstacles to business model choices to assess the viability of data marketplace business models."

The formulation of this research objective leads to the following **main research question:**

"How can we use the relationships between business model choices and obstacles to assess the viability of data marketplace business models?"

We realize that "viability" is a broad concept and it will be difficult to model and fully capture by only looking at the relationships between business model choices and obstacles. However, we find that looking at the relationships is only useful when its further purpose for doing this is clear: contributing to knowledge on the viability of data marketplace business models from at least this one perspective.

CH 2. Methodology

In this chapter, we will explain the methodology to answer the research question. We will first describe the general research approach. Afterward, we will describe the implementation of this approach and how this leads to the formulation of the sub-questions.

2.1 Research approach

To conduct this research, we will use an exploratory multiple-case study approach. The cases are represented by different data marketplaces. Each data marketplace will be subjected to a Business Model (BM) Stress Test. We discuss both aspects below.

2.1.1 Multiple-case study approach

To achieve the creation of an explanatory model we will use an exploratory multiple-case study approach. A case study approach is used because research that involves case data can get close to theoretical constructs. It is preferred over broad empirical research because it can form a more persuasive argument in relation to causal forces (Siggelkow, 2007). Each case, which is a real-life data marketplace, will form its own unit of analysis so that it adheres to replication logic (Eisenhardt, 1989).

Eisenhardt (1989) describes the process of building theory for case study research. Gregor & Hevner (2013) distinguish five types of interrelated theory: (1) theory for analyzing, (2) theory for explaining, (3) theory for predicting, (4) theory for explaining and predicting, and (5) theory for design and action. This research aims to explain the relationships between business model choices and data marketplace obstacles. Therefore, we can use the process of building theory for case study research for creating the fourth type of theory: theory for explaining. A theory type of explanation should include what is, how, why, and where. This type of theory does not aim to predict with any kind of precision and will lead to no testable propositions (Gregor & Hevner, 2013). In our case, we choose to embody the theory in an explanatory model. This will enable us to make the what, how, why, and where of data marketplace obstacles and business model choices the clearest.

The framework of Eisenhardt (1989) will structure the steps of this research. The steps and the corresponding activities are described below (Eisenhardt, 1989). To make clear why it is useful for our research objective, a context-specific rationale is given per step of the process of building theory from case study research.

1. Getting started: The first step is to define a research question and specify a priori constructs. The research question will help to shape a research focus and not get overwhelmed by the volume of data. Specification of a priori constructs will make the measurement of the constructs more accurate. In our case, exploring what is already known on data marketplace business models and obstacles will help us identify more accurate relationships. Instead of starting from scratch, the literature available on business models and obstacles will form the basis from which we can derive the relationships. By creating a conceptual background, we will also make it clearer what is already known and can reflect on how we add to the existing knowledge.

2. Selecting cases: The second step is logically one of the main steps of case study research: case selection. The goal when selecting these should not necessarily be to create a representative sample. It should rather be for motivation, inspiration, and/or illustration (Siggelkow, 2007). In this research, the goal is the inspiration for new ideas, as limited theoretical knowledge exists about the phenomenon of data marketplace business models. This makes theoretical sampling of cases an appropriate method of case selection (Eisenhardt, 1989). Theoretical sampling concretely means that cases will be chosen because they fill certain theoretical categories. When single cases are

theoretically sampled it is often because they are extreme examples or unusual cases (Eisenhardt & Graebner, 2007). Multiple-case studies chosen in this way have the advantage over single cases that the grounds they provide for theory building are stronger. The resulting theory should be more robust as the degree of variation in the empirical evidence is higher. This argumentation can also be applicable for creating models. We will select different data marketplaces by theoretical sampling to be the cases that will be examined. This is useful because we can get close to the constructs of interest. For a data marketplace to exist, they must have made certain business model choices. Because they operate in the real world, they must also experience certain obstacles that hinder their operations. These cases thus form an opportunity to identify the relationships between their business model choices and the obstacles that they experience.

3. Crafting instruments and protocols: The third step is crafting instruments and protocols to perform the research. In theory building research normally multiple collection data methods are used. It happens that only one method is used, but this does make the resulting hypotheses weaker because of the lack of triangulation. Researchers can choose to use either qualitative or quantitative tools or a combination of both. For our research, this step concretely means that after we have selected the data marketplaces that form the case studies, we need a protocol to collect the data on their business model choices and obstacles and relate them. Because a BM Stress Test provides an opportunity to identify these relationships, it is appropriate to use this qualitative method. We will describe this method below.

4. Entering the field: The fourth step of the process is entering the field. Frequently, data analysis overlaps with data collection. This is because each case is different and thus has different aspects that can and should be studied. It is therefore legitimate to alter and add data collection during a study. During this phase, it is also important to write down everything that occurs because it cannot be known upfront what will be important for the results and what will not. For us, entering the field means that after we have formulated the correct protocol, we can start collecting the data on relationships between business model choices and obstacles in the field. The "field" in this context is the data marketplaces that we selected as cases.

5. Analyzing Data: The fifth step of the process is concerned with analyzing the data that was aggregated in the field. The data can be analyzed within-case and then coupled with a cross-case search for patterns. For us, analyzing the data will lead us to the first patterns on data marketplace business model choices and obstacles. Patterns can be discovered within a data marketplace or across data marketplaces.

6. Shaping Hypotheses: The sixth step will involve shaping hypotheses based on the analyzed data. From the data that is collected relationships will start to emerge. In this research context, the hypotheses will be the correlation between business model choices and obstacles that we observed in the BM Stress Tests.

7. Enfolding literature: In the seventh step, these hypotheses will be inspected more closely by confronting them with similar as well as contradictory literature. If the literature contradicts the findings, it will form an opportunity for a more *"creative and frame braking mode of thinking"* (Eisenhardt, 1989). If it does not, the confidence in the findings is higher. In our case, we want to form a critical view of the relationships between the data marketplace obstacles and business model choices. The combination of the hypotheses and the critical view will give us better grounds to base our explanatory model on than only the hypotheses

8. Reaching closure: The eighth and final step of the framework is reaching closure. Ideally, the research finds closure when theoretical saturation is reached. This means that adding cases will only lead to incremental learning. However, the considerations for ending research are often pragmatic aspects such as time and money. Therefore, the number of cases is often planned. In this research, three cases will be studied. We want the result of this research to explain why it is difficult to create a viable data marketplace business model based on the obstacles that are identified. It should also indicate what business model choices might contribute to overcoming the obstacles or at least lower its impact.

Figure 3 depicts the adjusted framework by Eisenhardt (1989)

2.1.2 Business Model (BM) Stress Test

The Business Model (BM) Stress Test was originally developed to evaluate a business model design against market scenarios, regulatory uncertainties, or technological uncertainties (Bouwman et al., 2018). The result of this Stress test should be a changed business model that is more robust. The stress test consists of several steps including (1) describing the BM, (2) selecting the uncertainties, (3) mapping the BM to uncertainties, (4) developing a heat signature, and (5) analyzing the results.

In this research, the BM Stress Test will be used to develop a protocol for the data collection, as follows from step 3 of the case study research approach depicted in figure 3. The stress test is used in this way to map the obstacles and their impact on the business model choices. This is needed because the objective of the research is to create an explanatory model on how these aspects relate to each other – meaning the two should be linked. The BM Stress test creates this linkage.

We describe how we intend to implement the steps of the process and how it is exactly intertwined with the BM Stress Test in the section on “implementation of the steps and sub-questions” in the next section.

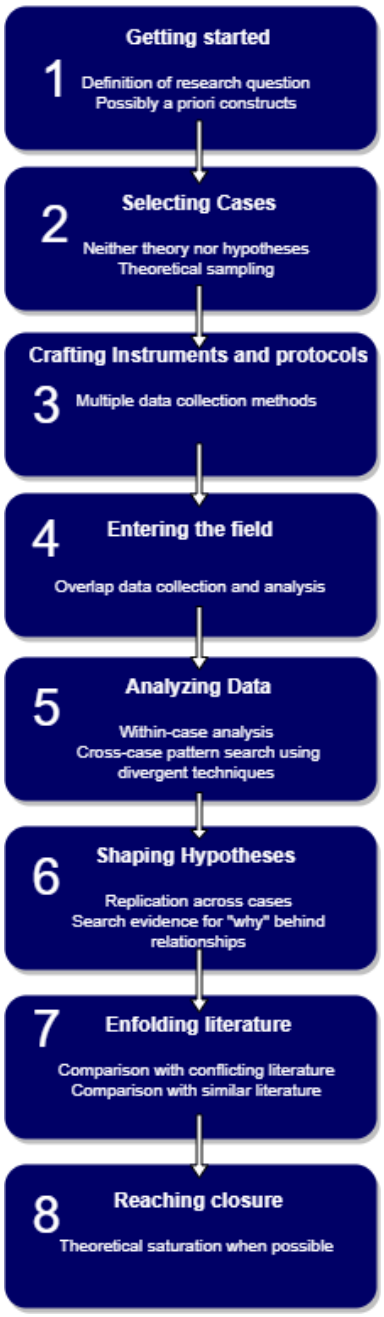

Figure 3: Process of building theory from Case Study research (adjusted from Eisenhardt, 1989)

2.2 Implementation of the steps and sub-questions

As stated, this research aims to create an explanatory model on the relationships between data marketplace obstacles and data marketplace business model choices. This is consistent with the case study research framework by Eisenhardt (1989) described above. The BM Stress Test will serve as an analytical tool for conducting some of these steps. The framework and BM Stress Test steps will structure a flow of sub-questions that will lead to answering the main research question. We will devote a chapter of this research to each step of the framework. However, due to large overlap, we will combine 'entering the field' with 'analyzing data'. The order of chapters thus is: getting started, selecting cases, crafting instruments and protocols, entering the field and analyzing data, enfolding literature, and reaching closure. We will now describe what we will do per step, and thus chapter, of the research.

2.2.1 Getting started

The first step in the framework involves defining a research question and possibly a priori constructs for this research. As this research concerns business models of data marketplaces, we should first further define the concepts of data marketplaces and business models. We will specifically describe what is already known about data marketplace business models and data marketplace obstacles by conducting a literature review. The main research question has already been defined: *"How can we use the relationships between business model choices and obstacles to assess the viability of data marketplace business models?"* Specifically related to the viability part of the research question, we are interested to see if we can define requirements to achieve an efficient market operation.

2.2.2 Selecting cases

The selection of cases will be based on theoretical sampling. The cases that will be chosen will fill three different categories. The categories mentioned are (1) personal data marketplaces – which allow consumers to monetize their data, (2) business-to-business data marketplaces – that allow organizations to exchange data, and (3) IoT/sensor data marketplaces – that allow sensor owners to monetize their devices. We chose the three different types of marketplaces deliberately, as they differ in the goal they try to achieve (e.g. monetizing data, devices, or allow data exchange). The data marketplaces in these categories thus have a different vision of the value that data marketplaces could have. Using theoretical sampling when choosing the case studies has the advantage of creating a view that is as rich and diverse as possible. Furthermore, as data marketplaces do not yet have a proven business model, they might also differ based on the phase of existence they are in. We mentioned earlier that many data marketplaces have already shut down, but there are also data marketplaces that are just starting up. To create a diverse view, it should be interesting to select cases that are in a different phase of their existence. It is assumed that this will lead to different perspectives on the obstacles they perceive. By following this approach, which focuses on the diversity of the cases, we will likely find the most relationships between data marketplace obstacles and choices. Although these types might not fully represent the full scope of options for data marketplaces (e.g. other types may exist), it does reflect the current main areas (Kelly, 2021).

2.2.3 Crafting instruments and protocols

After the cases are selected, instruments and protocols can be involved that will help to analyze the case data. The main tool that will be used for analysis is the **Business Model (BM) Stress Test** (Bouwman et al., 2018). The BM Stress Test will be done for all three cases. This method is qualitative and follows a predefined action plan. The first two steps of this Stress Test correspond to crafting the instruments and protocols before entering the field.

Step 1. Describing BM: The first step of the test is describing the business models of the selected data marketplaces. The BM Stress Test method is independent of BM ontology. During the explorative literature research, we already saw that researchers use taxonomies to embody data marketplace business models. Therefore, we will use a taxonomy to describe the business models of the chosen data marketplaces. By defining the business models of the different data marketplaces as characteristics of the taxonomy, the business model visualization will be structured and comprehensive. The outcome of this step is a visualization of the business model choices that the different marketplaces made, in terms of the characteristics their business model has. This outcome will serve as input for the next step in the process. The sub-question that corresponds to this step is formulated as follows:

> **SQ 1:** *"How can we describe the business models of three different data marketplaces by using an existing taxonomy?"*

Step 2. Selecting obstacles (adjustment from Bouwman et al. (2018)): Normally a BM Stress would use uncertainties to assess the business model choices (Bouwman et al., 2018). However, in this research context, where we want to investigate the relationships between obstacles and choices, obstacles will be used instead of uncertainties. During the 'getting started' phase, we performed a literature review on the obstacles mentioned in the literature. These obstacles can be expanded on by observing a focus group of data marketplaces experts discussing the obstacles for data marketplaces. Based on the conceptual background on data marketplace obstacles and the obstacles mentioned during the focus group a final obstacle selection can be made. These will form the a priori constructs for the interviews with the data marketplaces. During the interviews, it will become clear if the data marketplaces themselves indeed face these obstacles. Step 2 of the BM Stress test can be translated into the following sub-question:

> **SQ 2:** *"What obstacles can we select for performing a BM Stress Test on the chosen data marketplaces?"*

2.2.4 Entering the field

Once the business models are described and the obstacles are selected, data can be collected that directly relates to the objective of this research: identifying the relationships between business model choices and obstacles. Because the previous steps are all only preparation for doing this, the next BM Stress Test steps are grouped under 'Entering the field' of the theory building framework.

Step 3. Mapping BM to obstacles (adjustment from Bouwman et al. (2018)): In this step, the first part of the actual Stress Test will be performed. The obstacles that are identified from the literature and the focus group will form the x-axis of the BM Stress Test matrix. The business model choices of the selected data marketplaces will form the y-axis. Based on the interviews done with employees of these data marketplaces the relationships between the obstacles and their business model choices will be described by facts and issues in terms of characteristics. Step 3 of the BM Stress test can be translated into the following sub-question:

> **SQ 3:** *"How can we relate the identified obstacles to the business model choices?"*

Filling in the matrix will give an initial insight into the existent relationships. However, to formulate an explanatory theory on these relationships it is not only important to find if they exist. It is also of great importance what the impact of a specific obstacle is on a BM choice in this relationship. To assess this, a heat signature is used.

Step 4. Heat signature: After the relationships are defined by facts and issues, the possible impact of an obstacle on a certain business model choice and vice versa can be assessed. This is important because not all obstacles harm the business model to the same degree. The heat signature uses a coloring scheme to indicate this impact.

-**Red**: The obstacle is a potential showstopper for the data marketplace. Currently, the data marketplace does not have the means to overcome the obstacle.
-**Orange**: The obstacle requires attention. The existence of this obstacle will likely impede progress or achievement but does not make the business model choice infeasible.
-**Green**: The relationship between the obstacle and the business model choice exists, but because the data marketplace knows how to overcome it, it does not impede progress or achievement.
-**Grey**: No relationship could be identified.

The red- and orange-colored cells of the heat signature are thus the negative relationships between a data marketplace obstacle and a business model choice. The explanatory model that will follow therefore should distinguish between these different degrees of impact. Step 4 of the BM Stress test can be translated into the following sub-question:

SQ 4: "*What is the impact of the obstacles on the business model choices?*"

The output of steps 3 and 4 of the BM Stress Test is a colored matrix that visualizes the relationships, and the impact, between the business model choices of the data marketplaces and the identified obstacles. After the heat signature matrix is developed, the final step will be conducted, corresponding to the next step of the theory-building approach.

2.2.5 Analyzing data

The input for analysis of the data is three BM Model Stress Test heat signatures, one for each of the selected case studies. This heat signature will show what the problematic relationships are between some obstacles and some business model choices.

Step 5. Pattern analysis: The relationships shown in the heat signature can show patterns in the colorings used. Patterns can be recognized within case-analysis or cross-case. Examples of within-case patterns are preferred business model choices, where a choice only leads to 'green relationships' (or no) with all obstacles. Or a potential inconsistency between the choices, where one choice leads to a green relationship with one obstacle, but a red relationship with another. Examples of cross-case patterns can also relate to these inconsistencies. For example, when the same business model choices in one case study have different relationships with certain obstacles than in another case study. It is also possible that the cross-case analysis leads to an obstacle that has the greatest negative impact in all case studies. This step can be translated into sub-question 5:

SQ 5: "*What are the patterns between the business model choices and obstacles?"*

2.2.6 Shaping hypotheses

The heat signature and the analysis of the patterns form the input for shaping some hypotheses on the relationships between business model choices and data marketplace obstacles. This is true for the obstacles that show either green, orange, or red impact. These hypotheses will form the conclusion of the BM Stress Test. This translates into the sub-question 6:

SQ6: *"What are the hypotheses on data marketplace obstacles and business model choice relationships?"*

However, as the obstacles with orange and red impact slow down or even stop the progress of the data marketplace these are interesting to zoom in on.

2.2.7 Enfolding literature

It cannot be derived from the data marketplace case studies if relationships exist as they are claimed by the data marketplace owners. To verify the relationships, it is useful to enfold literature. In the theory-building approach, this step consists of a 'comparison with conflicting literature' and a 'comparison with similar literature. However, there is no literature available that relates data marketplace obstacles to data marketplace business models. There are two ways to still reflect on the relationships. (1) The first way is to find literature on possible solutions for data marketplace obstacles. These solutions can be connected to the negative relationships in the heat signature (the red and orange cells). If a solution connects to a certain red or orange cell of the matrix, it can be assumed that the relationship between the obstacle and the business model choice can change. Upon implementing the solution, the obstacle might be controlled, and the relationship might change of color, and thus impact. (2) The second way is to ask experts for their opinion of the relationships between business model choices and obstacles. In this way, it can also be established if the relationships can change or not but from another perspective.

The aim of this research is not the validate solutions. It is rather to explore the state of the relationships so that this can be included in the explanatory model and for future research directions. This can be translated into sub-question 7:

SQ 7: *"Can we verify the relationships that resulted from the BM Stress Tests?"*

2.2.8 Reaching closure

The end product of this research will be an explanatory model that tries to relate data marketplace obstacles to data marketplace choices. The BM Stress Test approach that will be used to create this model will be the first structured method that combines the identification of obstacles with business model choices. On top of this, it will also provide practical tooling. It will therefore be an addition to Business Model theory, specifically aimed at data marketplaces. The final model will form the theoretical contribution of this research. Based on these models implications will be made about the viability of data marketplace business models. These implications and possibly recommendations will form the practical contribution of this research.

2.3 TRUSTS

This project is related to a greater European-wide initiative called TRUSTS – Trusted Secure Data Sharing Space. This initiative aims to develop a data-sharing platform for secure, trustworthy, and GDPR-compliant data exchange. TRUSTS focuses on technological development, a sound legal and ethical foundation, and viable and feasible business models. Especially the TRUSTS component that concerns the business models for a fully operational and GDPR-compliant European Data Marketplace overlaps with this research and provides the opportunity for interactive value creation.

The TRUSTS resources are accessible and can be used in this research for collecting additional insights. However, the research is not fully dependent on it due to the complexity of the circumstances. For example, data on focus groups of stakeholders of data markets can be collected by participating in the organized workshops. Other involvement will be through bi-weekly meetings on the progress of the initiative.

2.4 Research flow diagram

In figure 4 the research flow diagram for this research is depicted. As described the structure of the chapters is based on the theory-building process as defined by Eisenhardt (1989). The most left column of the figure shows this process and is connected to the chapter structure in the middle column. The column on the right shortly describes what the chapters consist of.

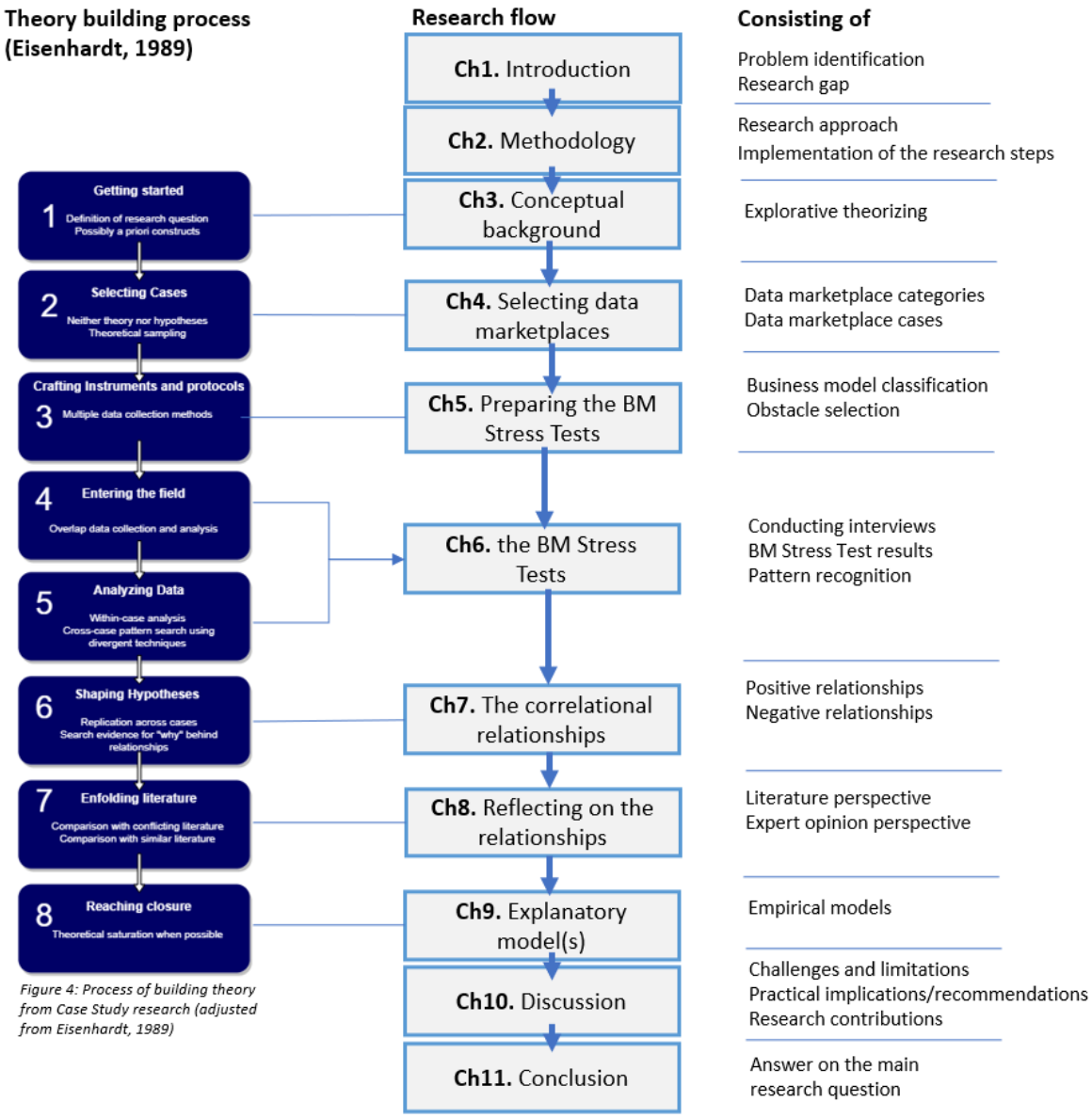

Figure 4: Process of building theory from Case Study research (adjusted from Eisenhardt, 1989)

Figure 4: Research flow diagram of the research (own illustration)

CH 3. Getting started: Conceptual background

This chapter reflects the first step of the theory-building process from case studies: getting started. It focuses on defining the main concepts of this research – the definition of a data marketplace, business models, data marketplace business models, and the requirements for efficient market operation. Describing these concepts will form the basis of the research and will lead to a more thorough understanding of the problem.

3.1 Data marketplaces

The concept of a data marketplace contains and/or overlaps with several other core concepts. First of all, a data marketplace relates to the core concept of a 'market", defined as a meeting place for suppliers and consumers to exchange goods (Spiekermann, 2019). This meeting place and the interaction between supplier and consumer (or: market sides) is not direct and physical but rather facilitated by an intermediary platform. This means another core concept relating to a data marketplace is a 'multisided platform'. According to its definition, such a platform mediates between different user groups, such as suppliers and consumers. The last core concept that describes the definition of a data marketplace is the *data* – the goods that are exchanged on the market. In this case, the good is non-physical and intermediate, meaning that it can be combined and transformed to create other goods (Koutroumpis et al., 2017).

Combining these core concepts, the most simple literal definition of a data marketplace is a *multi-sided platform where data between suppliers and consumers can be exchanged.*

To get a clearer view of the relationships between stakeholders, among others the suppliers and consumers, Spiekermann (2019) created a model of the data marketplace ecosystem. This model is depicted in figure 5.

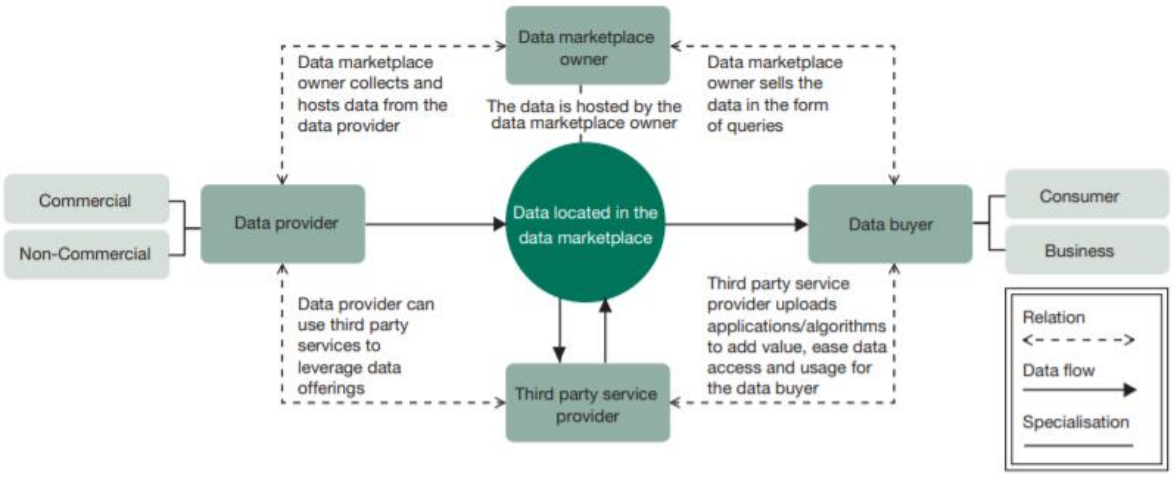

Figure 5: Model of data marketplace ecosystem (Spiekermann, 2019)

As can be seen in figure 5, there are four main stakeholders in a data marketplace environment: the data provider, the data buyer, the data marketplace owner, and the third-party service provider (Spiekermann, 2019). The data marketplace provider is a commercial or non-commercial stakeholder that owns and sells data. The data marketplace owner collects and hosts this data and sells it to the data buyer. This data buyer can either be a consumer or a business. Lastly, the third-party service provider is connected to both the data provider and the data buyer. The third-party service provider enables the data provider to leverage its data offerings and it can upload applications or algorithms that will help the data buyer extracting value from the data, and ease its data access and usage.

Another important distinction that can be made is in the type of data marketplaces that exist. Data marketplaces can be classified using different classification frameworks based on different determinants. For example, Stahl et al. 2016 use the determinants of orientation and ownership. Orientation refers to the market trading structure ranging from hierarchy to market-oriented. Ownership can be private, consortium or, independent (Stahl et al.,2016). Koutroumpis et al. (2017) classify according to the matching mechanism. This results in four types of marketplaces: one-to-one, one-to-many, many-to-one, and many-to-many. 'One' and 'many' refer to the amount of data providers and data buyers connected on such a data marketplace.

However, data marketplaces can also be classified based on whom they are designed for and what the source of the data is. Using this classification scheme there seem to be three broad types of data marketplaces: personal, B2B (business-to-business), and IoT (Internet-Of-Things) data marketplaces (Kelly, 2021). Personal data marketplaces allow consumers to monetize their data. In other words: the consumer using a personal data marketplace agrees to share their personal (and possibly sensitive) data in exchange for some kind of payment. This can also take shape in the form of vouchers or gift cards. Examples of this type of marketplace include, but are not limited to, Datum, SynapseAI, and DataWallet. B2B data marketplaces allow businesses or organizations to exchange data with each other. This category forms the majority of all data marketplaces. Datarade, Oracle, Eagle Alpha, AWS, Axon, and Snowflake are examples of this type. Lastly, IoT marketplaces allow sensor owners to monetize their smart devices. In other words: on this marketplace type, the intelligence that is generated by IoT devices is sold. IOTA Data Market and Streamr form examples for this type.

3.2 Business models

A business model describes how an organization or network of organizations intend(s) to create and capture value from technological innovation (Chesbrough & Rosenbloom, 2003). In other words, it represents a firm's core logic and strategic choices for value creation and capturing (Shafer et al., 2005). It is useful to study these business models because it helps to understand a business domain, makes communication and sharing of understanding of a business model amongst stakeholders easier, and helps design information and communication systems that support the business model (Pateli & Giaglis, 2004).

A term that is often used in relation to business models is 'strategy'. The difference between the two terms is that a firm's strategy goes beyond a firm's business model. According to Seddon et. al (2003) "*A business model is an abstract representation of some aspect of a firm's strategy"* (Seddon et al., 2003, p.237). A strategy includes a highly situated context, such as particular competitors, customers, suppliers, ways of assembling, and delivering to market changes. A business model aims to abstract from these details (Seddon et al., 2003). DaSilva & Trkman (2014) have a similar, but a slightly different view on it. They argue that "*strategy (a long-term perspective) sets up dynamic capabilities (a medium-term perspective) which then constrain business models (present or short-term perspective) to face either upcoming or existing contingencies"* (DaSilva & Trkman, 2014, p.383). According to this perspective, a firm should not need to change its business model to respond to contingencies, but can strategically prepare itself by developing dynamic capabilities.

Different ontologies can be used to make a business model explicit. A few examples of well-known ontologies are the Business Model Canvas (Osterwalder & Pigneur, 2010), the VISOR ontology (El Sawy & Pereira, 2013), and the STOF ontology (Bouwman et al., 2008). The STOF approach is especially important for this research as it currently forms a leading perspective on data marketplace business models (van de Ven, 2020). The STOF framework consists of business model domains: the

service domain (S), the technology domain (T), the organization domain (O), and the finance domain (F). The taxonomy that we will use as a basis for the research is also based on the STOF ontology. We describe this in the following section.

3.3 Data marketplace business models

We have defined a data marketplace as a multi-sided platform where data providers, data buyers, and third-party service providers are connected. We have also seen that to create and capture the value of the data exchange taking place the data marketplace needs a business model. Relatively little research has been conducted specifically on the business models of data marketplaces (Abbas et al., 2021). However, several taxonomies have been created that give insight into the possible business models that can be employed by data marketplaces. A taxonomy reflects the choices that data marketplaces have in certain characteristics when creating their business model. Specifically, a taxonomy lists several business model characteristics in different domains and dimensions, which forms a description of 'kinds' of possible business models (Baden-Fuller & Morgan, 2010).

Currently, two taxonomies are available in the literature (Fruhwirth et al., 2020; Spiekermann, 2019). One taxonomy is based on the Master thesis of van de Ven (2020). The latter aimed to improve the existing two taxonomies by including a bilateral data marketplace perspective in addition to multilateral. The difference is that in a multilateral data marketplace multiple data buyers are matched to multiple data providers, while in a bilateral data marketplace the data trading often goes via negotiated contracts (Koutroumpis et al., 2017). Also, this taxonomy is based on a multi-stakeholder perspective on business models instead of a single stakeholder perspective. Therefore, to live up to the aim of this research of creating a view on the relationships between data marketplace obstacles and business model choices that are as rich and diverse as possible, the taxonomy of van de Ven (2020) is suitable.

We will now explain the taxonomy created by van de Ven (2020) as we will later use it to describe the business models of the chosen data marketplace case studies. The taxonomy developed by van de Ven (2020) consists of the four STOF domains, each related to various dimensions. Each dimension has several options when it comes to the characteristics of a data marketplace business model. The taxonomy is depicted in figure 6. We will discuss the four different STOF domains – service, technology, organization, and finance below. We will further explain the characteristics that are not self-explanatory per domain.

	Dimension	Characteristics					
Service domain	Value proposition	Easy data access and/or tooling	Secure data sharing	High quality and unique data	All services in a single platform		
	Enterprise data marketplace	Yes		No			
	Data processing and/or analytics tools	Yes		No			
	Marketplace participants	B2B		C2B		Any	
	Industry domain	Any data	Geo data	Financial & Alternative data	Health & Personal data	Audience data	Sensor & Mobility data
	Geographic scope	Global		Regional		Local	
	Time frame	Static	Up-to-date	(Near) real-time	Multiple		
Technology domain	Platform architecture	Centralized		Decentralized			
	Data access	API	Download	Specialized software	Multiple options		
	Data source	Self-generated	Customer provided data	Acquired data	Multiple sources		
Organization domain	Matching mechanism	One-to-one	One-to-many	Many-to-one	Many-to-Many		
	Platform sponsor	Private		Consortium		Independent	
Finance domain	Revenue model	Commissions	Subscriptions	Usage fees	Asset sales		
	Pricing model	Freemium	Pay-per-use	Flat fee tariff	Package based pricing	Multiple	
	Price discovery	Set by buyers	Negotiation	Set by marketplace provider	Set by external sellers		
	Smart contract	Yes		No			
	Payment currency	Fiat money		Cryptocurrency			

Figure 6: Taxonomy of data marketplace business models (van de Ven, 2020, p.65)

3.3.1 Service domain

Starting from the top of the taxonomy, a **value proposition** indicates how a firm will deliver value to its customers. If a data marketplace is also an **enterprise data marketplace** it also functions as a private marketplace where data can be shared within the company or with external partners on a supply-based model. The **data processing and/or analytics tools** a data marketplace might offer can help stakeholders to perform analytics on their data. For **marketplace participants**, business-to-business (B2B) means that businesses or organizations exchange data amongst each other. Consumer-to-business means that consumers sell their data to businesses or organizations. There are several **industry domains** in which data can be exchanged and the **geographic scope** ranges from global to local. The **time frame** is concerned with whether a dataset needs to be updated frequently to keep the data relevant. Static datasets are not updated, up-to-date datasets are regularly updated and near real-time data is often sourced from IoT devices or online data trackers.

3.3.2 Technology domain

The **platform architecture** of a data marketplace can either be centralized or decentralized. Centralized means that the datasets of the data providers are offered on a predefined central location. Decentralized means that the data is located at the data provider and exchanged with distributed ledger technologies (blockchain). A data marketplace has several options for providing **data access**. For **data sources,** it can gather it from the internet itself, have customers provide it, or acquire it from external data providers.

3.3.3 Organization domain

The **matching mechanism** refers to the amount of data providers and data buyers that are connected through a data marketplace. The amount can be either ‘many’ or ‘one’. For example, one-to-many means that meaning there is a single data provider and multiple data buyers. Many-to-one works vice versa. The **platform sponsor** owns the property rights of a data marketplace.

3.3.4 Finance domain

The **revenue model** of a data marketplace can work in several ways. It can be by commissions, in which the data marketplace receives a fee over each transaction done. It can also be by subscription, in which a recurrent fee for using the data marketplace is paid. Usage fees mean that users pay for certain services offered and asset sales mean that the revenue comes from the sales of own data goods. The **pricing model** can be freemium, where basic functions are free and premium functions require a fee. It can also be pay-per-use, in which the price is proportional to the number of units that are consumed. A flat fee tariff means that full access to data is granted for a standard recurring fee. In package-based pricing, the data is sold in packages. **Price discovery** relates to the process by which a transaction price is reached that both data providers and buyers agree on. A data marketplace either chooses to use a **smart contract** or not. A smart contract is a contractual agreement coded into a script. When the terms in the contract are met this script is automatically executed. The **payment currency** of a data marketplace is typically fiat money or cryptocurrency.

3.4 Requirements for an efficient market operation

Our research aims to relate business model choices to data marketplace obstacles. We do this because we want to contribute to the knowledge on how a viable data marketplace business model can be achieved or how its viability can be increased. The outcome should thus be a viable data marketplace business model. As a data marketplace in essence is a market and employs a market mechanism is it interesting to identify what is needed to efficiently operate a market. By defining these requirements, we can later see how the data marketplace obstacles hinder these requirements. This will allow us to structurally assess the viability of data marketplace business models.

There are five requirements associated with an efficient market operation (Roth, 2002, 2007). According to Koutroumpis et al. (2017), these requirements can also be applied to data marketplaces. The five requirements are:

1. **Thickness** (liquidity) – A "thick" market means that there are enough opportunities to trade with other participants on the market. The positive externalities that result from this are essential for a marketplace to grow.
2. **No congestion** – Congestion is created by popularity and means that transaction time is slowed down. Transactions done on a marketplace should but rapid, but not so rapid that alternatives cannot be evaluated by individuals when considering an offer. This requirement mainly concerns technological choices in payment.
3. **Safeness** – In this context safe is defined as not having incentives to misrepresent or undertake strategic action which will lead to reduced efficiency.
4. **Provenance information** – Not having the information to assess the quality of a good traded on a market will lead to information asymmetry. In this case, the seller has an advantage over the buyer.
5. **No repugnance** – If there are social norms or legal restrictions that limit the use of pricing as an allocation mechanism, the marketplace can be seen as "repugnant". The policy, rules, norms, and cultural expectations can then not be coded into the matching algorithms. In the context of data marketplaces, this is often related to the privacy implications of trading data.

Because we believe that the data marketplace obstacles that we find will hinder these requirements, we will later relate these marketplace requirements to the data marketplace obstacles. We will now first look into these data marketplace obstacles. After that, we will assess how we can assign the marketplace requirements to the obstacles.

3.5 Data marketplace obstacles

Earlier we defined a data marketplace obstacle as '*something that impedes progress or achievement of a data marketplace'.* Therefore, the direct goal is to develop a preliminary overview of the main obstacle categories that withhold data marketplace from progress or achievement.

3.5.1 Obstacle identification approach

From the explorative literature research that led us to the research gap, we have already identified the articles that mention data marketplace obstacles (See Ch 1.2.1 'Literature search'). Therefore, we will now explore these articles in more detail to make a categorization of the main data marketplace obstacles. We will do this by listing the obstacles we can derive per article. This derivation can be explicit, e.g. "this is a challenge/barrier/obstacle/difficulty for data marketplaces" (as we previously mentioned we see these concepts as synonyms and will only refer to them as obstacles). The derivation can also be implicit e.g. "*the main problem towards the data economy is the lack of trust*" (Sharma et al., 2020). This statement does not explicitly mention data marketplaces but does

concern the same context. In this case, trust would also be identified as a potential obstacle for data marketplaces. We will base the clustering of the obstacle categories based on overlap in the mentioned statements.

3.5.2 Obstacle category discussion

The academic literature on data marketplaces does not provide an overview of all the obstacle categories that exist for data marketplaces. However, each article that we selected and read touches upon the subject of data marketplace obstacles. Because we want to create a complete overview of all the main obstacles that currently exist we combine the perspectives that the articles provide. We hereby combine the categories that are already mentioned in the literature (top-down) with categories that we can induce from multiple remarks made in different papers (bottom-up). We find that these remarks on data marketplace obstacles made in the academic articles can be part of greater overarching obstacle categories. To achieve this we made a structured overview of the 16 articles that we found that mention data marketplace obstacles. We noted down the remarks made on certain obstacles and tried to find an overlap in the statements so that we could find the overarching categories that we could place these remarks in. We gave the remarks that we thought overlapped a similar color indication to help us identify the categories (See Appendix A.2).

The first two categories we defined resulted from the top-down approach for our obstacle category identification. Täuscher & Laudien (2018) mention that a "normal" marketplace should provide two primary functions: creating trust and allowing for acceptable price discovery. Using this as a starting point we see that other papers mention why precisely those functions from an issue for data marketplaces. Unlike the marketplace environment where physical products are traded, the environment where data can be traded is not trusted for several reasons (discussed below). This makes 'trust' a good obstacle category for this research. The same goes for price discovery. While in normal marketplaces this might be well-developed throughout the years, remarks in other papers point out that the unique characteristics of data make it difficult to achieve this. These reasons make 'data pricing' a good obstacle category. We conclude that instead of primary functions of "normal" marketplaces, we find that the functions form obstacles for data marketplaces.

Then, we continue using a bottom-up approach trying to discover the similarities between remarks made in several papers and induce the obstacle categories from this. Several articles mention obstacles that relate to describing datasets and guaranteeing data quality. We believe that an accurate data description can lead to higher perceived data quality. Therefore, we put these constructs in one obstacle category. We find our next category based on the mentioning of the word "privacy". Privacy can be defined in several ways, especially as it changed throughout the years with the global digitalization trend. In this research, we define privacy as *"the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others"* (Mišura & Žagar, 2016, p.256). It is important that we explicitly mention that we define privacy in this way, because if we would define it for example as "hiding (personal) information" we could not use it for this research. Data marketplaces revolve around data exchange – and as consequence, the communication of information to others – which is the opposite of hiding information. We find that in literature the concept of privacy is often used in combination with 'trust' and 'control'. We chose to design a different category for trust as this is broader than privacy only. Higher perceived privacy can lead to a higher level of trust (Carlos Roca et al., 2009), but it is not the same. We also chose not to make it part of the trust category, as this might lead to less detailed relationships on the obstacles and business model choices of a data marketplace. We did choose to combine privacy with control. We did this because we find control to be part of the definition of privacy that we use: if a user does not have control over its data it can also not be involved in

determining how it is communicated to others. However, the categories of privacy and control and trust are thus interrelated. The following obstacle category we determined is 'matching data providers and data buyers'. In the selected articles this obstacle is only explicitly mentioned once. However, since it is a crucial concept of a data marketplace, which operations are based on connecting the two market sides, we chose to also make this an obstacle category. In this way, we can later explore it in more detail. Several articles also mention the legal uncertainty of data trading: data cannot be protected by intellectual property rights or copyright. We believe that when creating a category that relates to this, we should not restrict it to legal measurements only. From our point of view, there could be more ways to govern data trading. For example, in addition to rules and regulations prescribed by law, there can also be social agreements and non-textual agreements. Therefore, we chose to devote an obstacle category to the (absence of) governance mechanisms that are involved in the advancement of data marketplaces. Because not all obstacles can be captured in these specific categories, we chose to include an 'other' category. This category contains more general obstacles and can possibly be made more specific after we have conducted empirical research. How we assessed the findings from the selected articles to relate to the different categories can be found in Appendix A.2. Below we describe the categories of data marketplace obstacles we identified. These obstacles and their implications are described per category.

3.5.3 Literature review results

Below we discuss the obstacle categories separately. We describe the issues that we found in the literature that we can relate to a specific obstacle category. We will respectively discuss the obstacle categories of trust, privacy & control, product description & product quality, matching data providers & data buyers, pricing mechanism, governance mechanisms, and other (unspecific) obstacles.

3.5.3.1 Trust

Sharma et al. (2020) claim that data marketplaces should be platforms that have no bias towards either data providers or data buyers. They state that independent data marketplaces at market levels could offer this, in contrast to privately owned or consortia-based data marketplaces. However, it was identified that private parties are generally more trusted for providing reliable services, security, and privacy. A data buyer will be less likely to engage in a financial transaction if there is no trusted relationship (Park et al., 2018). It is even stated that "*the main problem towards establishing a data economy is the lack of trust"* (Sharma et al., 2020, p.41). This can be either in the parties involved or in the system (the data marketplace) itself. According to the authors, trust should be guaranteed both in allowing the participants to trust each other, as well as the system they are using. Another obstacle that relates to the lack of trust (and possibly security) is that data providers fear that competitors (data buyers) could benefit from offering in-house data on a data marketplace in such a way that their own business interests will be harmed (Spiekermann, 2019; Fruhwirth et al.,2020). They might perceive taking part in a data marketplace as a means for their direct competitors to gain a competitive advantage over them which might reduce profits or throw them out of the market (Nikander & Elo, 2019). A high level of trust seems especially important because of the unique characteristics of data. Data has almost zero marginal costs, which means that "*Data is expensive to produce but cheap to reproduce*" (Mišura & Žagar, 2016, p.259). This characteristic enables malicious users to buy data for cheap and then re-sell it for a higher price to make a profit. This of course hurts the owner of the data.

3.5.3.2 Privacy and control

Privacy can be defined in several ways, especially as it changed throughout the years with the global digitalization trend. In this research, privacy will be defined as *"the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others"* (Mišura & Žagar, 2016, p.256). On data marketplaces, general data but also sensitive information can be sold. This information can for instance be collected by IoT devices and sold on IoT data marketplaces, or it can be personal information that can be sold on personal data marketplaces. Following the definition of privacy stated above, the sensitivity of the data does not necessarily need to be a problem. It only means that data sellers should exactly understand what data they are selling and to whom they are selling the data (Mišura & Žagar, 2016). However, this required understanding often poses the problem. Data providers are not always aware of the quality and the legal status of the datasets they are selling (Koutroumpis et al., 2017). For example, assumed anonymized datasets are often not hard to deanonymize. Also, while the importance of metadata was just described, privacy regulations might prevent data providers to disclose this metadata as this might include strategically valuable information. Privacy is also related to trust. Privacy-enhancing technologies are either based on the mechanism of soft privacy or hard privacy (Parra-Arnau, 2018). In soft privacy, the users that share data entrust the entity in being responsible for the protection of their data. In hard privacy, the same entities are mistrusted and therefore not believed to be capable to protect the privacy of the users. In this case, the users only trust themselves and are therefore responsible for protecting their own privacy. Another notion related to privacy is that of control. Users often do not have full control over their data (Perera et al., 2017). For example in the context of IoT devices, the data is managed by different vendors in different silos and cannot be shared or traded without agreements with the vendor. Also, because of the characteristics of the internet, users have little control over where their data might end up (Nikander & Elo, 2019). For example, it should be prevented that the data ends up for trade on the black market (Stahl et al., 2016).

3.5.3.3 Product description & product quality

In principle, a data marketplace is not much different from an e-marketplace except for the goods that are sold. An e-marketplace sells physical products, and a data marketplace sells data. Although this may seem a relatively small difference at first glance, it has some greater implications for the product description and product quality (Sharma et al., 2020; Lawrenz et al., 2019). In physical product trading detailed product descriptions can be given, sometimes even using videos or pictures. In data trading using metadata is the main method for the data provider to give product descriptions. Metadata is data about data (Lawrenz et al., 2019). The data buyer uses it to search for and identify relevant data sets, making it the most important decision factor for this group of actors. Referring to the Information Pyramid: metadata is the *information* about data that is needed to generate *knowledge*. Without the metadata, knowledge cannot be achieved even not by an algorithm because this algorithm cannot be designed. Metadata is thus essential for enabling data trading (Lawrenz et al., 2019). However, it might also be an obstacle for the data provider to give an accurate and correct description of the dataset. Lawrenz et al. (2019) identify the following challenges for metadata:

- The kind of metadata that is required for a data marketplace
- The method of generating the metadata
- The need/possibility for generating metadata automatically or entering it manually
- The form in which metadata should be stored (structured, unstructured, or semi-structured)
- The possibility of metadata to act as verification key for data exchange

In general, the challenge is that big because every data is unique which makes it harder to derive the corresponding metadata (Lawrenz et al., 2019). Apart from the description, product quality is also

important. Correct decisions, in the interest of the companies' revenue, can only be made if the sold data is of high quality (Schomm et al., 2013). However, this concept also forms a challenge. After buying a physical product, the product can be returned if it does not fit the buyer's expectations. However, this is not feasible for datasets as data can easily be copied, which makes the data lose its value (Lawrenz et al., 2019). It could be argued that the quality of data could be measured upfront using objective properties of data (that can be described by metadata) and subjective properties of data, which means that the relevancy is checked upon the buyers' requirements. However, these requirements cannot become known for the data provider as this could lead to fake data provision or revelation of data buyers' business models (Sharma et al., 2020). Stahl et al. 2016 also argue that quality per se is not the correct concept when it comes to buyer requirements. Some customers expect complete, formatted, and reliable data of high quality, but it doesn't necessarily need to be pre-analyzed or specific. In other contexts, for example, when the data is just an add-on product for decision-making, the specificity is way more important than the quality of the data. In general, this difference reflects two scenarios: the wish for a constant standard versus the wish for high individuality (Stahl et al., 2016). The differences between physical product trading and data trading are summarized in figure 7 and figure 8 (Sharma et al., 2020).

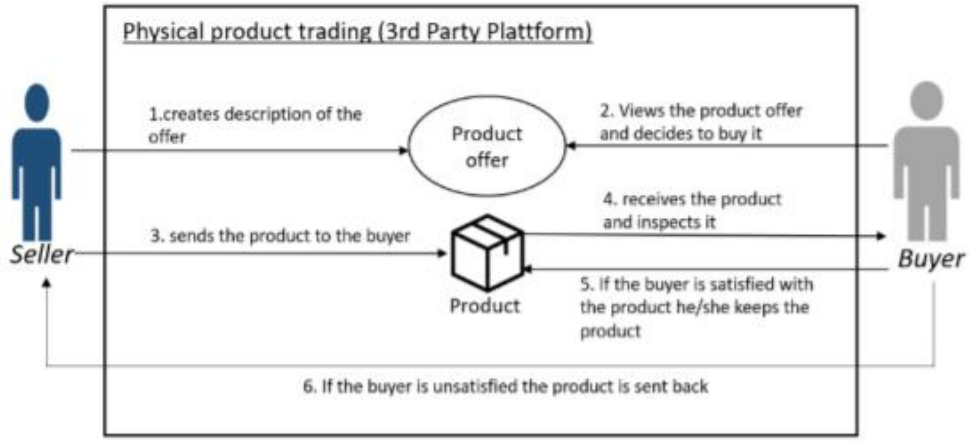

Figure 7: A simplified model of physical product trading (Sharma et al., 2020)

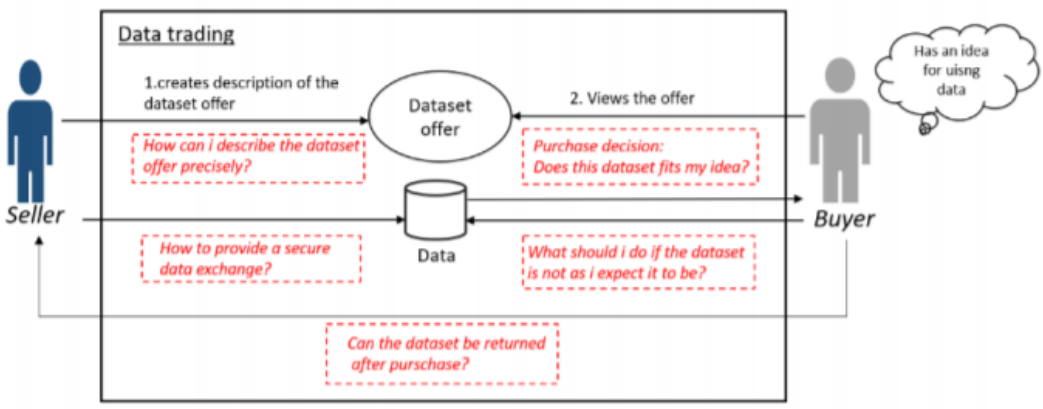

Figure 8: A simplified model of data trading (Sharma et al.,2020)

3.5.3.4 Matching data providers and data buyers

While metadata is the main guidance for data buyers to navigate and can ensure a well-functioning data marketplace from this side of the market, it neglects the information stream from data buyer to data provider. Hayashi & Ohsawa (2020) state that the distribution and trading of data are hindered because there are insufficient means for data providers to learn what kind of data is desired by the data buyers. This could for example be done in the form of a data request by the data buyers.

3.5.3.5 Pricing mechanism

Another implication of the difference between physical goods and data, which is also related to the previously mentioned product description and product quality, is the involved pricing mechanism (Spiekermann, 2019; Stahl et al., 2016). Unlike physical goods, no established rules exist for pricing datasets (Fruhwirth et al., 2020). According to Spiekermann (2019), the direct obstacle that relates to this is the low willingness to buy data. This results from a phenomenon called the 'Arrow paradox' (Spiekermann, 2019; Fruhwirth et al., 2020). In this context, the paradox expresses itself when data buyers do not recognize the value of a dataset because it cannot be exposed before it is purchased. According to Azcoitia & Laoutaris (2020) the problem of recognizing the value, and thus making the right selection in buying data sets can be broken down into two interrelated subproblems. The first part consists of computing how useful the datasets are going to be for their Artificial Intelligence (AI) or Machine Learning (ML) algorithms. This can be captured in accuracy metrics. The second part is concerned with computing how this accuracy can be translated into monetary gains. However, the second part of the problem is not directly related to the service of the data marketplace and therefore only an indirect problem. The low willingness to buy also partly results from not understanding the costs for data providers for the creation, processing, storage, and distribution of data (Spiekermann, 2019). Guaranteeing data quality is associated with high costs. For data providers pricing data is difficult since the data can be replicated without any costs, and transmitted for free (Azcoitia & Laoutaris, 2020). Also, the monetary value of a dataset depends on the context in which it is used (Nikander & Elo, 2019). Because this context is unknown to the data provider, they cannot evaluate the monetary value and are afraid of setting a price for the data buyer. In the context of personal data marketplaces (or possibly sensitive information from IoT devices) the pricing also consists of a trade-off against privacy (Mišura & Žagar, 2016). In this case, it is difficult to decide for data providers which elements can be offered for mining, which should be protected, and what a good deal is (Parra-Arnau, 2018).

3.5.3.6 Governance mechanisms

Another obstacle for the advancement of data marketplaces seems to be the absence of legal frameworks (Spiekermann, 2019). Without these frameworks, no legal certainty can be given for trading data. No intellectual property rights can be requested for the data, which means that a violation of the rules cannot be 'solved' by liability rules. Creating efficient protection for data seems to be difficult. The structure and organization of a database can theoretically be protected by copyright, but this does not apply to the data in the database itself (Koutroumpis et al., 2017). Limited data rights have been proposed that would balance protection rights and encouragement of innovation using data (Mattioli, 2014). These rights would not affect the distribution of data but do prevent the use of the data for a certain amount of time without authorization. However, this would still be hard to keep track of (Spiekermann, 2019). Because of this difficulty, protection is often offered through contractual means in the shape of lengthy and complicated data license agreements. It should also be considered that when legislation is possible, it is often geographically dependent (G. Smith et al., 2016).

3.5.3.7 Other

It is not possible to handle every obstacle that exists as a separate category. The paragraphs above reflect specific obstacles. These obstacles specifically hinder a certain marketplace requirement. Other obstacles might be less specific but can still affect the viability of a data marketplace business model. For example, Smith, Ofe & Sandberg (2016) identified six types of obstacles for open data (marketplace) adoption: institutional, task complexity, use and participation, legislation, information quality, and technical details. Institutional obstacles are obstacles that refer to an organizational unwillingness to change. The focus on accountability is higher than the focus on entrepreneurial activities. Data providers might not be willing to share data, have insufficient resources for its development, or are unable to respond to the input of users. Task complexity refers to anything that involves difficulty with finding, analyzing, or processing data. For use and participation, the ease and attractiveness of joining a data marketplace are of importance. Related factors are lack of incentives, lack of time, lack of ability, costs, or competition from other marketplaces. The legalization obstacle includes both legal as well as technical issues. An example of such a technical issue is that advanced technological capability leads to advanced ability to process data, e.g. de-anonymizing once anonymized data. Information quality is again concerned with product description (metadata), and technical details refer to the data itself.

3.6 Assigning the requirements to the obstacles

In the previous two sections, we have described the requirements for efficient market operation and the different data marketplace obstacle categories. These sections are correlated: following the requirements leads to a viable business model and the obstacles hinder achieving a viable business model. If we take a step back we can thus say that the obstacles do not directly influence the viability of a data marketplace business model but obstacles hinder complying with the marketplace requirements. To structurally analyze the viability of data marketplace business models we want to connect the potential obstacle categories to the marketplace requirements.

We argue that the lack of trust and its implications could hinder a data marketplace from complying with the ‘safeness’ marketplace requirement. This is because, like trust, safeness is a feeling. We believe that a feeling of safeness can be achieved by creating a trusted environment. In this environment, there should be no incentives to misrepresent or undertake strategic action that leads to reduced efficiency of the data marketplace. On top of trust, we argue that the lack of governance mechanisms also hinders the safeness requirement. When trust cannot be fully be achieved, governance mechanisms could provide (legal) certainty for safety.

We also argue that the obstacle category of privacy and control hinders the ‘no repugnance’ marketplace requirement. In its definition, it says that this requirement is related to the social norms or legal restrictions that limit the use of pricing as an allocation mechanism. In this case, it is privacy and control that limit pricing. If privacy and control are deemed more important than monetary value, and this cannot be guaranteed, it might limit the development of data marketplaces.

In this argumentation, we find a slight overlap with another potential obstacle category and marketplace requirement. Even if privacy and control would lead to no restrictions, the pricing mechanism should be well developed stand-alone. However, we saw that pricing data is difficult for multiple reasons. We argue that the pricing mechanism category hinders the ‘no congestion’ marketplace requirement. In the definition, congestion is said to be created by popularity and means that the transaction time is slowed down. In the context of data marketplaces and obstacles, we interpret this as the interest for buying and selling data being present, but the actual transactions being slowed down because the pricing mechanism is underdeveloped.

Furthermore, we believe that the obstacle category of product description and quality hinders the 'provenance information' marketplace requirement. This follows from the definition of provenance information: if a dataset is or cannot be well described, it will lead to information asymmetry between data providers and data buyers.

Moreover, one of the marketplace requirements is thickness, meaning that there are enough opportunities to trade with other participants on the market. We argue that these opportunities arise when there are (1) enough participants on a marketplace and (2) these participants can be matched to a data provider and/or data seller. We find that involving participants in the marketplace (1) is hindered by the obstacles mentioned in the 'other' category. Things such as unwillingness to change, high task complexity, ease-of-use, lack of incentives/time/ability will influence the creation of positive externalities that are needed for a marketplace to grow. Then, if these participants are involved, they should be matched to make the trading of data happen (2). If the matchmaking is misaligned, the data marketplace thickness' marketplace requirement is not complied with, as it limits the opportunities for efficiently trading data.

We summarize the findings of the literature review and the assignment of the obstacles to the marketplace requirements in *table 1.*

Table 1: Overview of the potential data marketplace obstacle categories

Potential obstacles (categories)	Description	Marketplace requirement
Trust	-Lack of trust in the system itself and/or its participants -Harm of own business interest -Possibly extra important due to unique characteristics of data	Safeness
Privacy and control	-Privacy implications of trading data might limit development -Data providers are not always aware of the legal status and quality of datasets they are selling -Users do not have full control over their data	No repugnance
Product description & quality	-Metadata is needed to describe a product but can be hard to derive -Quality is hard to define: it is based on user requirements	Provenance information
Matching providers & buyers	-There are insufficient means to discover what data is required by data buyers	Thickness
Pricing mechanism/strategy	-Data buyers: Low willingness to buy data (mainly because of the Arrow paradox) -Data sellers: Pricing data is difficult because of o (almost) zero marginal costs o Monetary value is difficult to evaluate o Trade-offs against privacy	No congestion
Governance mechanisms	Absence of legal frameworks; no legal certainty for trading data	Safeness
Other (unspecific)	Unwillingness to change, high task complexity, ease-of-use, lack of incentives/time/ability, competitors, technical issues	Thickness

3.7 Chapter conclusion: framework resulting from explorative theorizing

In this conceptual background chapter, we have discussed the definition of data marketplaces and business models. We also described what a data marketplace business model enfolds and what is already known about this subject. We identified what the requirements for an efficient marketplace operation are and what is already known about the obstacles for data marketplaces. To create a structure for further analysis we assigned the obstacles to the marketplace requirements. Based on the explorative theorizing we performed, we can model a framework as the base for the remainder of this research. This framework is depicted in figure 9.

The relationships in this framework can be explained as follows. The research gap, and therefore the aim of this research, is to find the relationships between the business model choices of data marketplaces and the obstacles that are faced to assess the viability of data marketplace business models. The outcome of the framework is thus a viable data marketplace business model. The business model choices and the obstacles we found are visualized in the first two columns, orange and yellow, of the framework. For the sake of simplicity, only the STOF domains and not all options for the business model choices that are possible according to taxonomy are depicted.

The relationship between the business model choices and obstacles has a double arrow. This is because a business model choice can influence the degree to which an obstacle is perceived as problematic. An obstacle in its turn can influence a data marketplace owner to adapt its business model choices to that obstacle. We argue that the obstacles do not directly influence the viability of a data marketplace business model but do this via the marketplace requirements. These requirements would lead to an efficient market operation (Koutroumpis et al., 2017), and thus presumably also a viable data marketplace business model. The marketplace requirements are visualized in the third green column of the framework.

As we said, the identified obstacles for data marketplaces hinder living up to these requirements. We argued that trust and governance mechanisms affect the safeness requirement, privacy and control the no repugnance requirement, data quality and data description the information provenance requirement, matching and other the thickness requirement, and pricing the no congestion requirement.

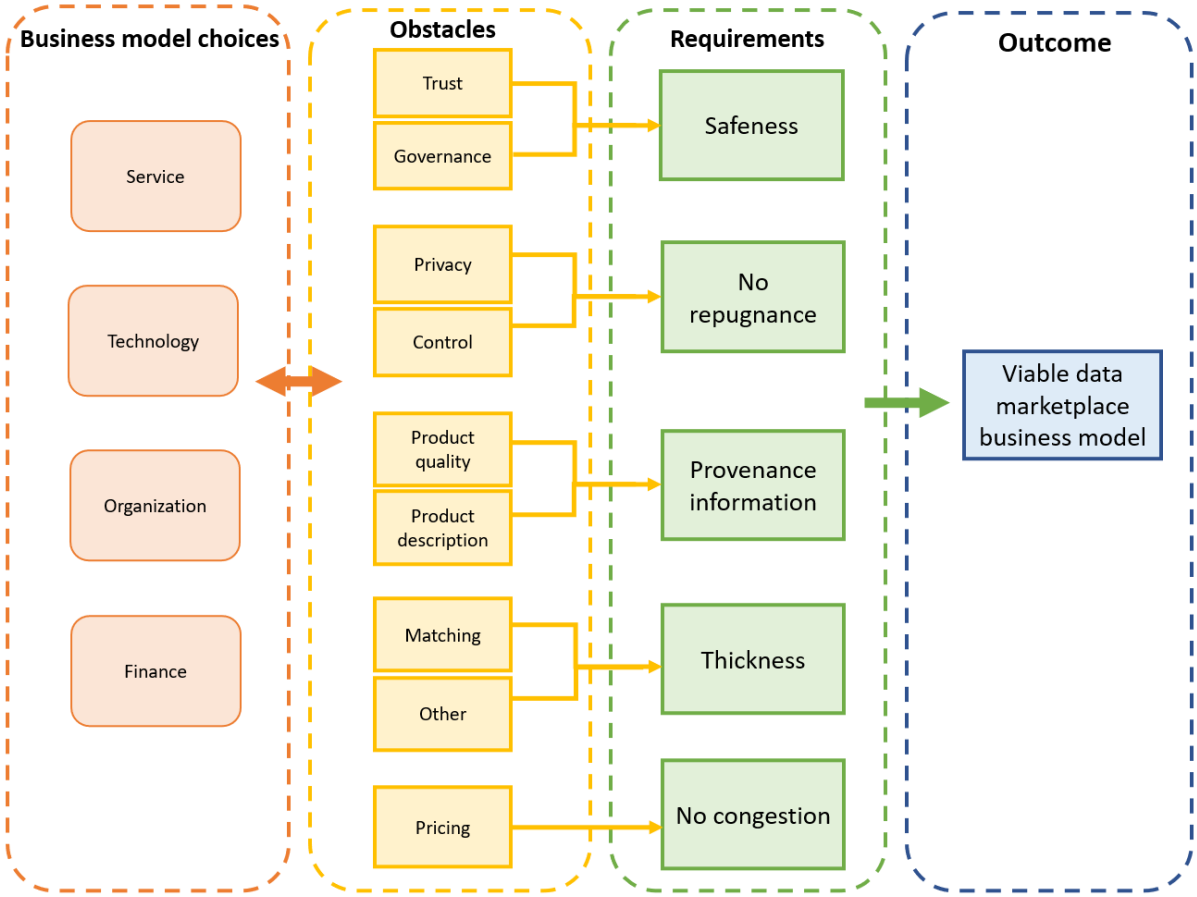

Figure 9: Our research framework resulting from explorative theorizing. The first orange column represents the business model choices that can be made in the four STOF domains. These choices affect the impact of data marketplace obstacles, depicted in the second yellow column. The arrow between these columns is double because the obstacles can influence the business model choice that a data marketplace owner makes. The obstacles hinder complying with the marketplace requirements, depicted in the third green column. Living up to these requirements would lead to the wished outcome: a viable data marketplace business model.

CH 4. Case selection: Selecting data marketplaces

The fourth chapter reflects the second step of the theory-building process from case studies: case selection. As described, the case studies will be selected by using theoretical sampling. This means that the cases will be chosen based on the fact that they fill certain theoretical categories (Eisenhardt, 1989). In this case, these categories are based on who the data marketplaces are designed for, as this will allow for a rich and diverse perspective on data marketplace obstacles. For the first category, the personal data marketplaces, the goal is to allow customers to monetize their personal data. For the second category, B2B data marketplaces, the goal is to allow organizations and businesses to exchange data. For the third category, IoT data marketplaces, the goal is to allow sensor owners to monetize their smart devices. The different categories will first be defined more explicitly, after which the selection of the case studies will be described per category.

4.1 Data marketplaces categories

As described above, there seem to be three main broad types of data marketplaces based upon who they are designed for and where the data is sourced from (Kelly, 2021). To categorize the cases, each type of data marketplace should first be defined in detail.

The first type of data marketplace is a **personal data marketplace.** As mentioned, the goal of this type of data marketplace is to allow customers to monetize their personal data. Personal data is data that can be used to identify, contact or locate an individual person. The main cause for this seems to be that it is believed that giving data away for free is a market failure (Economides & Lianos, 2021). Economides & Lianos (2021) explain this by describing two primary markets for digital services. The first market is that of digital platforms providing a certain service. For example, Google provides the ability to search, and Facebook the ability to build a social network. The second market is that of the sale of personal information. These markets would function separately in a competitive world. However, this is not the case. The platforms collect all sorts of personal information (e.g. locations, interests, activities, political opinions, social interactions) and compensate the user only by providing the service. This is based on a take-it-or-leave-it contract based on an opt-in by default, meaning that personal data can be provided at zero price. Economides & Lianos (2021) state that this results in several harms to users: some users might be excluded from the platform because they are not willing to sell their personal information at zero price, some users would be compensated at a positive price instead of a zero price in a competitive market, asymmetric information leads to unawareness of the value of the data, and the value of privacy might be underestimated by the users. Because the digital platforms have a dominant position, they have a lot of power over the users. The goal of personal data marketplaces is to shift this balance of power between individuals and companies or in other words users and platforms (Parra-Arnau, 2018). A typical personal data marketplace uses APIs to enable users to connect to their apps and services where the data can be sold. A personal data marketplace generally aims to create a win-win situation for both users and companies. The users will receive some kind of monetary compensation (e.g. payments, vouchers, discounts, or gift cards) from the companies. The companies in return will not receive raw data, but rather a profile that gives an overview of a user's data. The users will thus trade their data for something more valuable than "free" services, and the companies will receive higher quality and more accurate data about the users. Apart from this, personal data marketplaces can also be considered as a driving force behind Data for Good (Kelly, 2021). This means that when data exchanged on these marketplace for example concerns insights on human movement and behaviour, it can be used to effectively respond to emergencies and pandemics.

The second type of data marketplace is **a B2B (Business-to-Business) data marketplace.** This type of data marketplace allows businesses and organizations to exchange data to make it easier for companies to achieve their business targets. Typically B2B data marketplaces offer data from a multitude of data providers on one platform. From a data provider side is should be easy to integrate your to-be-sold data on the marketplace, and from a buyer perspective, B2B are generally used to source the right data for the right price (Kelly, 2021). The data exchanged on this marketplace is often used for marketing, sales, or business intelligence purposes. Apart from that, the data can also be used to create new business opportunities and even to develop new data-driven ventures. It thus has a lot of potentials to stimulate innovation. In comparison to personal data marketplaces, the amount of datasets that can be shared is way larger, as businesses often have more technological capabilities to store big data.

The third type of data marketplace is an **IoT (Internet-Of-Things) data marketplace.** The Internet of Things can be defined as a *"loosely coupled centralized system of smart objects that are able to sense their environment, process information and network with other devices or users"* (Mišura & Žagar, 2016, p. 255). These smart objects – or devices – generate a data stream, that is (mostly) owned by the device owners. This data could be valuable for third parties. Therefore, the main goal of this type of data marketplace is to allow sensor owners to monetize their smart devices. Smart devices could range from sensors on an oil rig to an airline's jet engines or air-conditioning units in an office tower (Tang et al., 2018). The data that is generated by these devices could respectively be used to create better predictive weather models, train better machine learning algorithms, and gaining a competitive advantage over competitors in the design phase of a new project. These examples are more industrial applications of IoT devices, but individual households currently also possess a number of IoT devices in-house. In general, data from smart household devices such as a smart thermostat, smart plug, smart fridge can be given access to by the data owner to the data consumer. The data consumers could be electricity companies, smart fridge manufacturers, research institutions, interest groups or the government. The data streams can then either be used for micro-level decisions, such efficient electricity schedules, or macro-level decisions such as long term investments for the company. The data owner would be compensated for this access provision by any kind of monetary reward. This example shows that IoT data thus reaches beyond the personal data sold on personal data marketplaces. Especially since the amount of IoT devices is forecasted to triple from 8.74 billion in 2020 to 25.4 billion in 2030 (Holst, 2021), it seems that in theory a lot of revenue could be generated from monetizing the resulting data streams. Completely new business models in different industrial sectors could arise if it is known that there is a market for it. The value of a data marketplace exceeds the value of the data-as-a-service (DaaS) industry, where third-party external data is also acquired to augment internal datasets, as IoT data streams are more difficult to handle (Tang et al., 2018). The devices that generate this data often control critical processes, infrastructures, or sensitive information. Therefore, hackers also have a great interest in attacking IoT devices.

Below, a categorization is visualized of currently active personal, B2B, and IoT data marketplaces. This categorization was done by datarade.ai, a data discovery platform and B2B data marketplace itself. The website contains information about 2000+ data providers across 210 specific data categories. Figure 10 depicts the current top data marketplaces in the three main categories according to datarade.ai.

Personal	B2B	IoT
DATACOUP	OPENPRISE Data Orchestration for Marketing & Sales	DAWEX
datum	crunchbase	Streamr
wibson	DaTaguru.in Powered by SALESHARK	zenodys
PeoplefindersDaaS	Informatica	Datapace
BIGtoken®	The DX Network	oneTRANSPORT Open marketplace for data
Datawallet	Datarade	Mobilitäts Daten Marktplatz
Hu-manity.co	snowflake	SweetData.io BETA
bron.tech	aws marketplace	IOTA
VETRI	Eagle Alpha	TERBINE TAKE THE PULSE OF THE EARTH
OSA decentralized	ocean	databroker dao Powered by SettleMint
		SYNCHRONICITY

Figure 10: Overview of the data marketplaces considered for this research in the three identified categories: Personal, B2B, and IoT (own illustration).

4.2 Data marketplace cases

One data marketplace per category was chosen for the multiple-case study. Choosing the cases via theoretical sampling of categories already contributes to the aim of providing a rich and diverse view of the phenomenon. The diversity can be increased by selecting data marketplaces that are in a different phase of their existence. It is assumed that this will lead to different perspectives on the obstacles they perceive. Using this approach will increase the likelihood that as many relationships between data marketplaces obstacles and choices will be found. Also, we chose to select these cases because (1) out of all considered cases (see figure 10) their websites and whitepapers were most elaborate and gave us the most detailed information, and (2) these data marketplaces were willing to participate in the research. The case selection was not based on specific business model characteristics of the data marketplaces.

For the first category, the personal data marketplaces, VETRI has been chosen as a case study. This platform is still at the beginning of its existence. For the second category, B2B data marketplaces, the DX Network has been chosen as a case study. The DX Network was founded in 2017 but is no longer operating. For the third category, IoT data marketplaces, Databroker DAO is chosen as a case study. Databroker started as an IoT data marketplace but rebranded to an "all kinds of data" marketplace. VETRI has just started the data marketplace 'journey', Databroker has been revisioned – implying that some business model aspects were not catching on –, and the DX Network has stopped operating. Including this one clear failure case might lead to useful insight on which business model elements are real showstoppers.

It should be noted that all descriptions of the case studies are based on their own websites and published whitepapers. Therefore, the case descriptions provide only one perspective and might include biased statements. Triangulation of these cases is very difficult as data marketplaces are in their infancy and therefore there are no articles available that provide any other perspective. The applications of the data marketplace were installed and tried out. This however did not give any insight into the business model behind the marketplace, but only on the customer interface and experience. As this is not the focus of the research this was not deemed useful and is therefore not included in the research. The case descriptions are still valuable as they do provide insight into the intentions of the data marketplaces and how those are, according to them, translated into a certain (viable) business model.

4.2.1 Personal Data Marketplace: VETRI

All information about VETRI is acquired from the website https://vetri.global/ and the white paper VETRI: Value you data published in October 2017 (VETRI, 2017).

VETRI is a personal data marketplace founded in Zürich, Switzerland. The marketplace is still at the beginning of its existence, as it was officially released in January of 2020. Its initial market target consists of the DACH (Germany, Austria, Switzerland) countries, the UK, and the US. Currently, it is only operating in the US. However, soft launches of the platform have taken place in the other mentioned countries. The 'beginning phase' perspective is interesting because the company should have a clear view of what the obstacles for launching/scaling up will be and how their business model is going to deal with that. Its mission is to *"build an ecosystem around our vision for the future of a data economy that benefits everyone and not just a few"* (VETRI, 2017). It wants to empower individuals with control over their personal data and the opportunity to be fairly rewarded for it. It is described that users are currently not in control as due to the growing digitization the sensitivity of the data that is collected by large corporations has increased. The platform aspires to contribute to the principles of the new digital age: full data ownership, full control over the usage of personal data,

adoption of regulatory frameworks and enforcement mechanisms, and continued advancements of enabling technologies. VETRI is available as an application in the Google Play Store and App Store and provides the ecosystem needed for data trading.

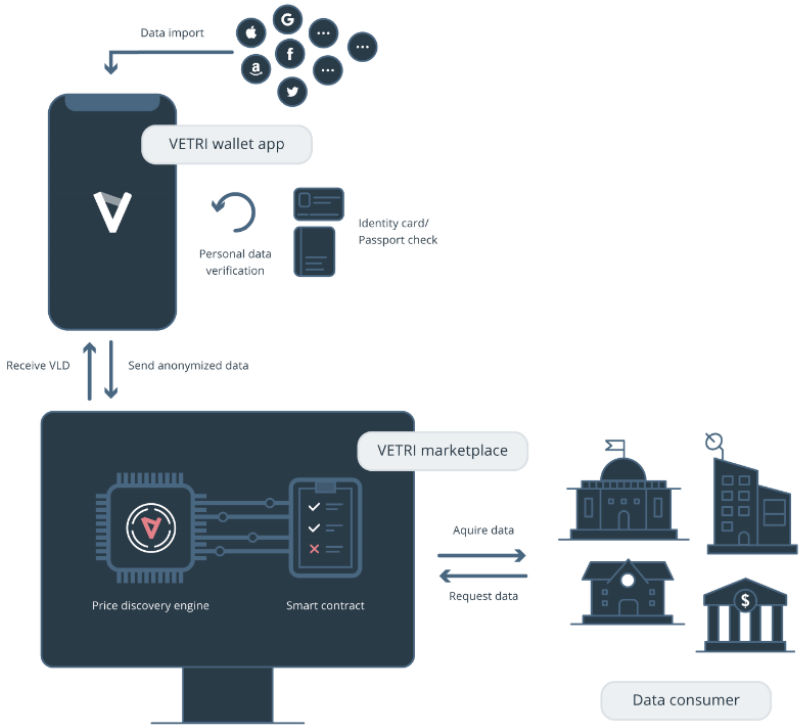

Figure 11: The Vetri ecosystem consisting of a mobile wallet and marketplace (VETRI, 2020)

Figure 11 displays the VETRI platform infrastructure. As can be seen, the VETRI platform consists of two main aspects: the VETRI wallet app and the VETRI marketplace. It also consists of two main stakeholders: the (application) users and the data consumers, or in other words the organizations that want to acquire the data to use for business/research analytics purposes. The wallet is meant for the users to manage their personal data and the marketplace is used by data consumers to access that data. The VETRI wallet is connected to the digital identity (eID) of the consumer for authentication purposes. This is done so that bots, personas, and duplicates can be filtered out and guarantees can be given that a real individual generated the data. However, ultimately the personal data that is connected to this eID should be anonymous, while its integrity and genuineness are guaranteed.

The wallet will in the future also import data streams from all kinds of online platforms like Google, Facebook, Twitter, Amazon, etc. This variety of data could be valuable for different types of research or analytics of data consumers when users opt to share and grant access to share their data. Users themselves can control what they want to share and what they do not want to share. The VETRI app can also be used to just securely store the data, as the data is stored locally, encrypted in the users' phone. VETRI states that this, the privacy and security and not the monetization of the data per se, is their number one priority. On top of the automatically incoming data streams from online platforms, the app can also be used for survey requests by data consumers. For all data and survey requests, users get reimbursed in VLD tokens. This is the cryptocurrency of VETRI itself. VETRI uses blockchain technology and smart contracts to ensure trust in the marketplace and to guarantee smooth transactions between the users and the data consumers. The VLD tokens that are collected by the customers can be traded for gift cards for stores and restaurants. Apart from this function, the token

should also act as a validation/verification mechanism for the agnostic eID module. VETRI is designed as a not-for-profit, peer-to-peer platform funded by the VETRI Foundation. This means that VETRI itself is a non-economic intermediary. This means that the marketplace does not take any commissions or other fees from the transactions that will take place. Because of the foundation base, VETRI does not need these commissions to be sustainable and can pass 100% of the generated revenue on to its users. Figure 12 indicates the possibilities a user has when using VETRI.

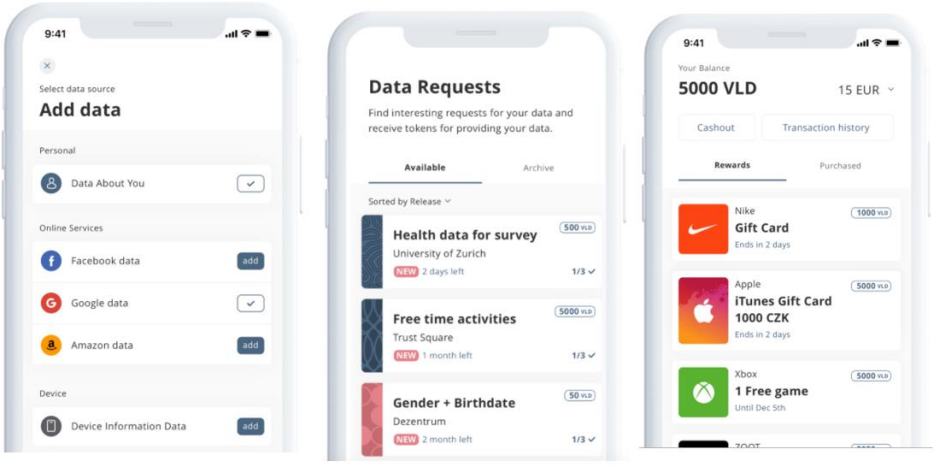

Figure 12: The user interface of the VETRI app. The image shows the ability to add data, show data requests, and exchange VLD tokens for gift cards (VETRI, 2020).

4.2.2 B2B Data Marketplace: the DX Network

The DX Network was a B2B data marketplace and has operated from 2017 to 2020. Their mission was: “Making the Data Economy a reality.” The DX network claimed to be the first real-time marketplace for structured data. The shut-down of this marketplace makes it interesting to study, as it implies that it couldn’t overcome the obstacles that data marketplaces face. We will first describe their ambitions and means to achieve this, so it can afterward become clear why some business model choices did or did not work.

All information about the DX network is acquired from the website and the white paper version 3.6 published by the DX network itself (J. Smith, 2018b).

The businesses that the ‘B2B’ characteristic of the DX network refers to are mainly start-ups. The DX Network saw a great paradox in the rise of the Digital Age: the lack of efficient access to data. Specifically, the structured data that was available about start-ups, which can be seen as the drivers of innovations. Start-ups in this context are defined as young ventures that are founded to develop new technologies or services. This insight was gained when developing another online platform: Startup Tracker (startuptracker.io). This platform aims to make the discovery of start-ups by others easier. When start-up data would be shared in an open, comprehensive up-to-date, and structured manner the rate of innovation would increase for two reasons. The first reason is that the start-ups would be able to more easily reach their target audience. The second reason is that existing companies would be enabled to improve their products by integrating the data acquired from the start-ups. The data that is proposed to be shared between these stakeholders is company data, meaning that the data describes information about a company. According to the DX Network, this differs greatly from personal data concerning privacy and potential conflicts in the supply chain. In contrast to personal data that one often tries to keep private, the goal of companies is usually to reach a wide audience. Also, data protection regulations do often not apply to company data. Figure 13 and figure 14 visualize where the DX Network tried to intervene in the data supply chain. The marketplace should connect the data collectors, which is defined as an individual or organization that has direct access to the structured data about one or more start-ups, and the service and app providers. Having access to these datasets will lower the barrier for entry for these providers, and thus also benefits the end-user who uses the developed application or service. All data offered on the marketplace is public, but the DX network offers in a structured way.

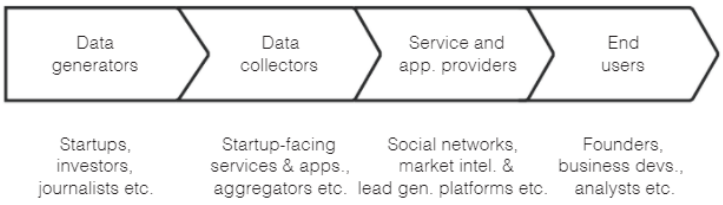

Figure 13: Data supply chain without the involvement of the DX Network (J. Smith, 2018b)

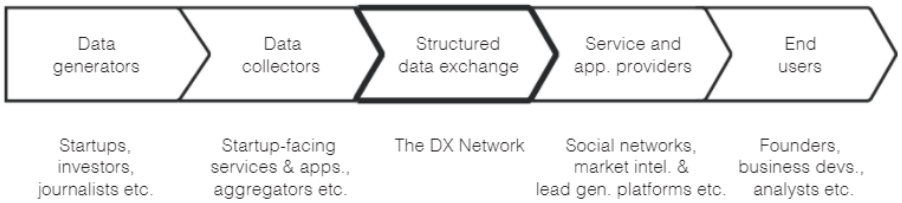

Figure 14: Data supply chain with the involvement of the DX Network (J. Smith, 2018b)

The DX Network marketed itself as having three key advantages for acquiring data. First of all, it is said to have real-time access to a wide range of data sources via a single API. The data that was offered on the platform was listed by the sellers in a standard format. This was defined by the marketplace's data model. Second of all, the DX network made use of semantic queries. To make these queries the DX Network's semantic query language could be used: DX/SPARQL. This can be seen as a simplified version of the standard SPARQL query language. This language would allow for complex user-crafted data queries. To make these queries first a Python library (the DX library) had to be downloaded. An example of the data properties that could be queried for the entity type 'Organization' can be seen in figure 15.

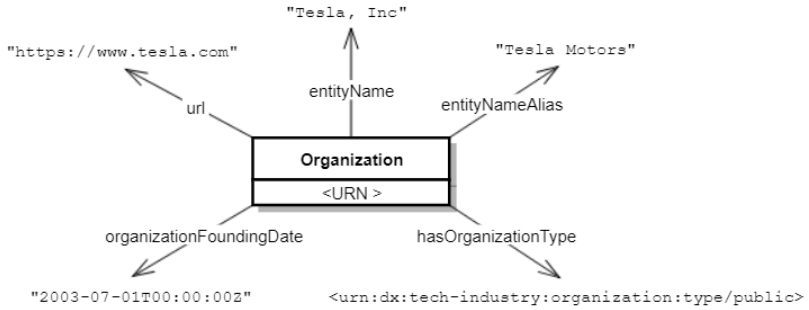

Figure 15: Example of Tesla, Inc for the data properties available for the Organization entity (the DX Network, 2018)

Furthermore, they used a pay-as-you-go and pay-per-datapoint principle. Because of the blockchain characteristics and native cryptocurrency, the DX token, micropayments were made possible. The DX token was built on the Ethereum blockchain. In practice, this meant that every time a data point of a particular seller got queried for, the seller received the correct payment without the involvement of a middleman. The utility of a particular data point was also measured in the number of times this data point was queried for. As for the data collector, the seller of the data, they paid a one-time listing fee to vouch for the quality and utility of the data and to discourage malicious submissions. All transactions included a network fee to operate.

4.2.3 IoT Data Marketplace: Databroker (DAO)

Databroker DAO was a start-up located in Leuven, Belgium. It started as a decentralized marketplace that allowed owners of IoT sensor data to sell their data. Currently, Databroker DAO is rebranded as Databroker and extended its scope from IoT data to all kinds of data. Their message now is: "Databroker is the marketplace for data". Studying this platform is interesting as it includes the transformation of an IoT data marketplace to what seems to be a general data marketplace. The transformation implies that the obstacles as an IoT data marketplace only overshadowed its potential. First, the initial intentions of Databroker will be described so that they can later be compared to their new goals and ambitions. This will give insight into the business model aspects (of an IoT data marketplace specifically) that did not work. The current business model will be used for further analysis and should give insight into their current practices in dealing with obstacles.

All information about Databroker DAO is acquired from the white paper v2.0 published by Databroker DAO itself (van Niekerk & van der Veer, 2018).

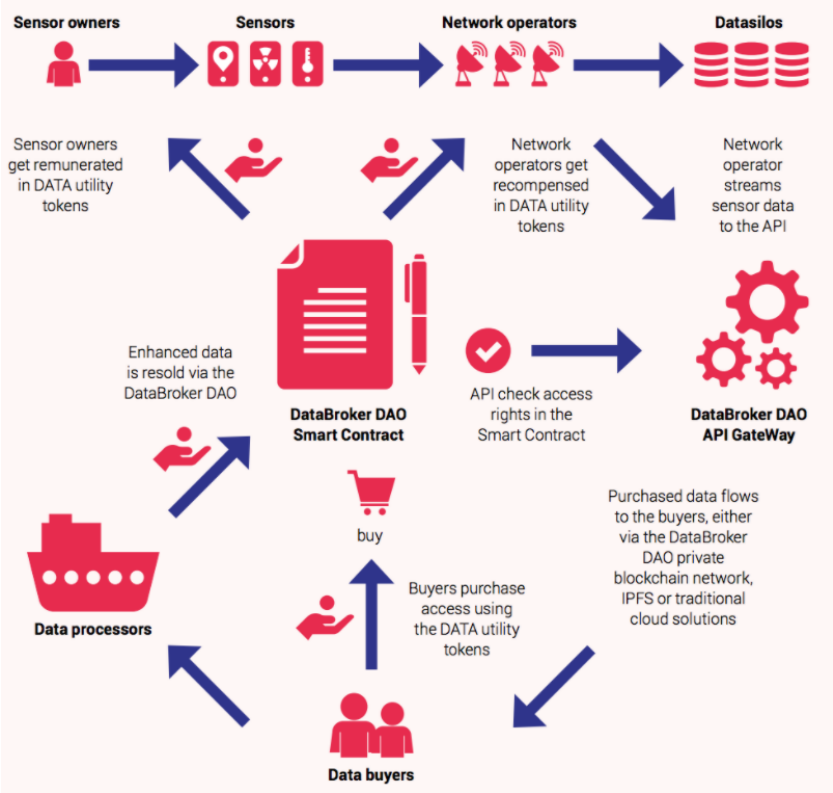

Figure 16: Overview of the Databroker DAO data trading process (van Niekerk & van der Veer, 2018)

The initial Databroker DAO process intention is depicted in figure 16 (van Niekerk & van der Veer, 2018). They indicate four types of stakeholders: the sensor owners, the data buyers, the data processors, and the network operator. The sensors owners have sensor data to sell, and the data buyers want to buy this data. The data processors also want to buy this data but their goal is to enrich it and then resell it to their customers. The network operator should expose the gateway they operate to enable sensor owners to sell their data. Because normally the data that is gathered by IoT devices gets stored in private data silos, an interconnection is needed between the network operator and the Databroker marketplace. This was done by using an API. Both sensor owners and the network operator get compensated in DTX tokens. The data buyers and data processors also buy this

data using DTX tokens. This is the cryptocurrency token of Databroker itself which operates on the Ethereum platform.

As stated, Databroker has been revisioned, rebranded, and retooled. All the information about Databroker has been acquired from the databroker.global website and the white paper 'Databroker: The Next-Generation Data Marketplace' (Databroker.global, 2020). The scope of the platform is no longer limited to IoT sensor data, but can now be used as a marketplace for all kinds of data. This includes categories such as environment, geographics, agriculture, economy, transport, supply chain, energy, and people. According to Databroker, the scope was extended because *"the evolution of the data landscape and valuable market feedback revealed an even greater potential: to become a true peer-to-peer marketplace for all data.* (Databroker.global, 2020) It can be said that Databroker is now both a B2B data marketplace and a sensor data marketplace. To realize this, the marketplace uses a new architecture that is depicted in figure 17.

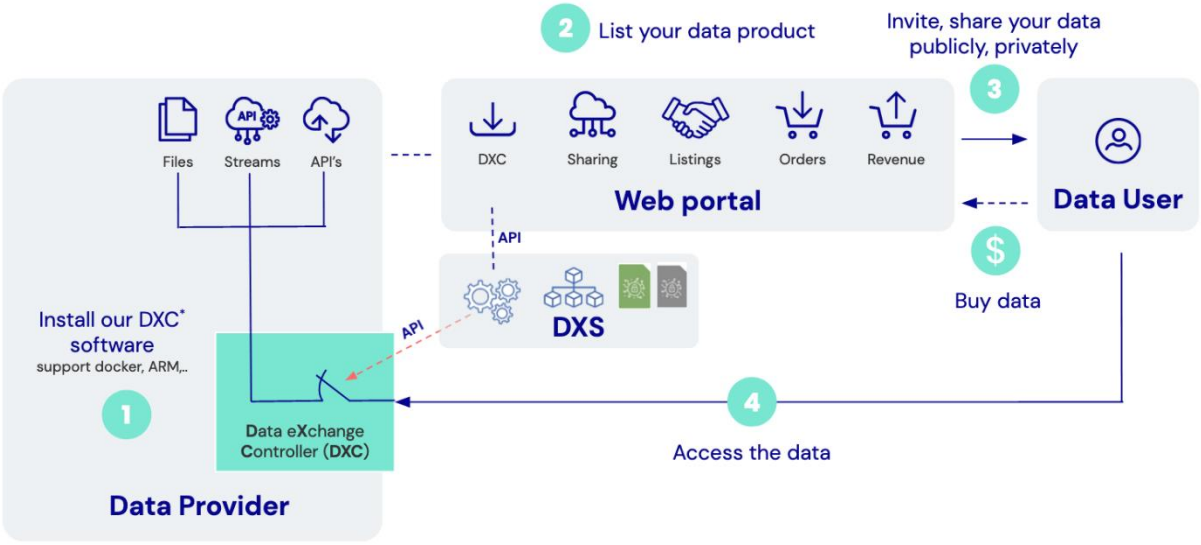

Figure 17: The Databroker platform architecture (Databroker.global, 2021)

The data exchanged on Databroker is not stored on the platform itself. It is transferred directly from data provider to data buyer via the Data eXchange Controller (DXC), which is a piece of specialized software developed by Databroker. The DXC supports three data formats: files, streams, and APIs. It forms the gateway between the data provider and the data users. The DXC is in sync with the smart contracts written by the DXS (Data eXchange Smart Contract) that govern the access to the data. When a data user pays for a data product it effectively sends DTX tokens to a smart contract, which are then sent to the data provider after a certain amount of time.

However, the data buyer itself pays in fiat currency via credit or debit card. This amount is transferred to DTX tokens and stored in the Data provider's wallet. From here it can be converted back to fiat currency to cash out. Data can be sold either publicly or privately. The latter meaning that the data provider and data buyer will directly negotiate a purchase agreement. In all cases, the data provider drafts a license which states the rights to use the exchanged data. The data marketplace also includes a "DataMatch" service. This service helps identify and match data providers and data buyers. The data buyer can request a DataMatch provider for specific datasets, who will then contact the data providers. Apart from the commercial appliance of the Databroker marketplace, the company also offers a white-labeled Platform-as-a-Service. This means that Databroker's technology can also be sold as a PaaS for data-rich organizations that need a more secure and personalized method for data exchange.

CH 5. Crafting instruments and protocols: Preparing the BM Stress Tests

The fifth chapter reflects the third step of the theory-building process from case studies: crafting instruments and protocols. As explained, the instrument that will be used is the Business Model (BM) Stress Test. This Stress Test is a matrix with business model choices based on the taxonomy as the y-axis and data marketplace obstacles as the x-axis. To prepare this Stress Test the business models of the selected data marketplace therefore first need to be classified. This will give us the y-axis. After this, we will try to expand on the conceptual background on data marketplace obstacles (Chapter 3). This will give us the x-axis.

5.1 Business model classification: BM Stress Test Y-axis

Based on the description of the chosen data marketplaces, we can now classify their business models into the taxonomy of van de Ven (2020). This is the first step of the BM Stress Test (Bouwman et al., 2018) and therefore an instrument to perform this test in a later stage of the research. This sub-chapter answers sub-question 1 of the research: *"How can we describe the business models of three different data marketplaces using the existing taxonomy?"* The following section will show the classification and will explain the reasoning behind it.

5.1.1 VETRI's business model

The classification of VETRI's business model is depicted in figure 18. The section below will describe why certain aspects are modeled as they are.

Value proposition – VETRI's number one priority is to offer users a secure way to store and manage their personal data. This data can however also be precisely requested by a data consumer (data buyer), who targets specific attributes. Therefore, VETRI also offers high-quality data. Therefore, the value proposition is that it offers all services in a single platform.
Enterprise data marketplace – VETRI does not offer any additional services for organizations to share data within the company or with external partners and is thus not classified as an enterprise data marketplace.
Data processing and/or analytics tools – VETRI does not offer any data processing and/or analytics tools.
Marketplace participants – The marketplace participants of VETRI are consumers as well as businesses (C2B). Users using the VETRI application and integrated wallet are connected to the data consumers (the businesses) through the VETRI marketplace.
Industry domain – VETRI is focused on exchanging personal data from users with the data consumers.
Geographic scope – VETRI is currently focussing on some initial markets (DACH, UK & US), and aims to afterward expand into new ones. Its scope is therefore currently regional.
Time frame – The data that is offered by the users of VETRI is mainly up-to-date data.
Platform architecture – VETRI uses a decentralized platform architecture making use of blockchain distributed ledger technology.
Data access – The data will be accessed via specialized software.
Data source – The data on VETRI can either be imported from different online platforms or can be directly provided by the user (e.g. by filling in data categories in the app or filling out surveys from data consumers). Currently, VETRI is also cooperating with data aggregators for data import. Therefore, there are multiple data sources used.
Matching mechanism – The matching mechanism in place is one-to-many. One user profile, with its corresponding data attributes, can be sold to many data consumers.

Platform sponsor – The platform sponsor is the private VETRI foundation.
Revenue model – VETRI is designed as a non-profit platform and therefore does not generate revenue on the data that is sold via the platform. However, VETRI does possess a certain amount of VLD tokens that will increase in value in the future. We model the sale of these tokens as asset sales.
Pricing model – VETRI chose a pay-per-use pricing model. In the future, they might transfer to package-based pricing if this would turn out to be efficient for dealing with datasets of different sensitivities.
Price discovery – The price of the data that is sold is based on the desirability as perceived by the data consumer. This is also a function of the nature and reliability of the data and the privacy settings and sharing preferences defined by the user.
Smart contract – VETRI uses a smart contract to enable individuals to control their personal data within the marketplace.
Payment currency – Cryptocurrency. VETRI uses the VLD token that can later be redeemed by the users in gift cards for restaurants and stores.

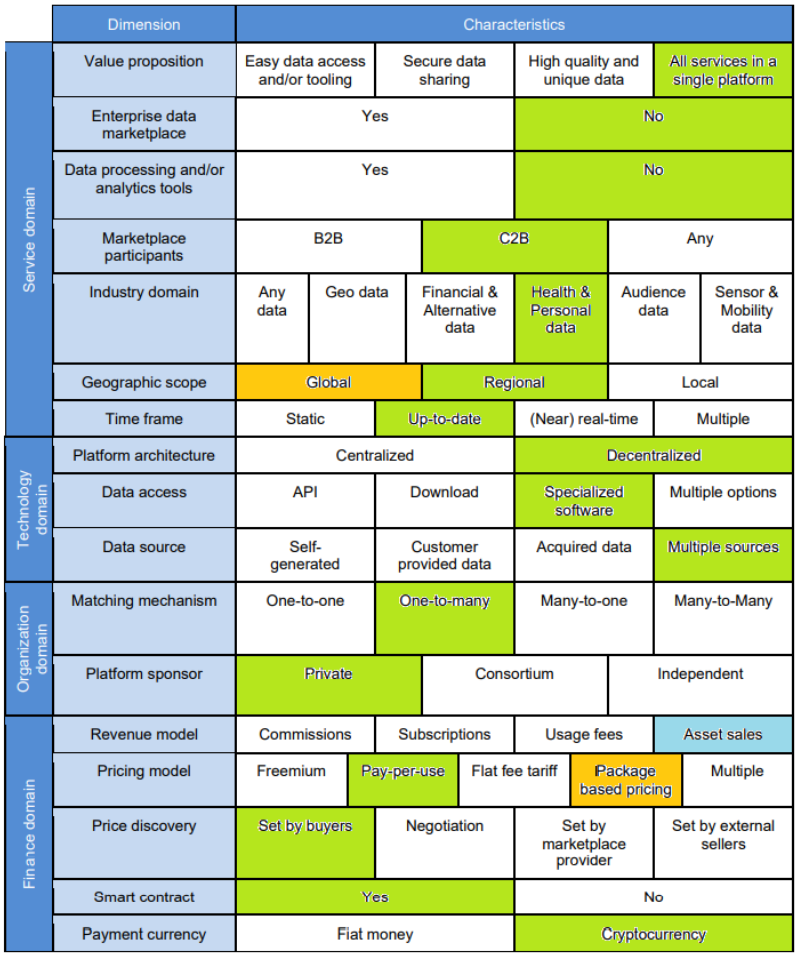

Figure 18: Classification of the VETRI business model. The green elements reflect what VETRI is currently doing, the orange objects reflect what they (possibly) aspire for in the future. The blue color reflects that the revenue model is not conventional, as VETRI is a not-for-profit organization.

5.1.2 the DX Network's business model

The classification of the DX Network's business model is depicted in figure 19. The section below will describe why certain aspects are modeled as they are.

Value proposition – The DX Network describes its main objective as making access to data efficient. It is described that it would be a positive-sum game for the start-up industry and the data industry that underlies it.
Enterprise data marketplace – The DX Network does not offer any additional services for organizations to share data within the company or with external partners and is thus not classified as an enterprise data marketplace.
Data processing and/or analytics tools – The DX network does not offer any data processing and/or analytics tools. It merely allows marketplace participants to buy and sell structured data in real-time via an API.
Marketplace participants – The DX Network is a B2B marketplace.
Industry domain – The data offered on the DX Network was structured start-up company data. We classify this as 'alternative data'.
Geographic scope – Since the DX Network was based on Startup Tracker, which allows searching for start-ups from all over the world, the geographic scope is also classified as global.
Time frame – As it is described that the DX Network can be thought of as an exchange or real-time marketplace, the time frame is classified as (near) real-time.
Platform architecture – The DX Network is said to be a product of crypto-economics. This is the science of designing decentralized systems for realizing desired properties. It uses cryptography for the validity of the system state and economic incentives so that the properties are economically sustainable.
Data access – All data sources could be accessed through a single API.
Data source – The DX Network allowed data collectors to put structured data about companies up for sale and let data consumers buy it. The data source is therefore the data collectors, which we group under customer-provided data.
Matching mechanism – the DX Network enables many data collectors to sell the same data to an unlimited number of buyers. We classify this as a many-to-many matching mechanism.
Platform sponsor – The DX Network had an independent platform sponsor.
Revenue model – The DX network was non-profit. It argued that it should be to optimize functionality rather than profitability. However, the aim was to also be self-sustainable, meaning that the service should not have to rely on external financing past its initial funding. It aimed to fund the service by the DX token sale and by including a small network fee for each transaction to support the operation.
Pricing model – The DX Network redistributed a query fee paid by the data consumers to the data collectors.
Price discovery – The price was set by the data collectors.
Smart contract – The DX Network made use of a smart contract to ensure payment between the parties involved.
Payment currency – Participants of the DX Network paid in the marketplace's own cryptocurrency: the DX token.

	Dimension	Characteristics					
Service domain	Value proposition	Easy data access and/or tooling	Secure data sharing	High quality and unique data	All services in a single platform		
	Enterprise data marketplace	Yes		No			
	Data processing and/or analytics tools	Yes		No			
	Marketplace participants	B2B	C2B	Any			
	Industry domain	Any data	Geo data	Financial & Alternative data	Health & Personal data	Audience data	Sensor & Mobility data
	Geographic scope	Global	Regional	Local			
	Time frame	Static	Up-to-date	(Near) real-time	Multiple		
Technology domain	Platform architecture	Centralized		Decentralized			
	Data access	API	Download	Specialized software	Multiple options		
	Data source	Self-generated	Customer provided data	Acquired data	Multiple sources		
Organization domain	Matching mechanism	One-to-one	One-to-many	Many-to-one	Many-to-Many		
	Platform sponsor	Private	Consortium	Independent			
Finance domain	Revenue model	Commissions	Subscriptions	Usage fees	Asset sales		
	Pricing model	Freemium	Pay-per-use	Flat fee tariff	Package based pricing	Multiple	
	Price discovery	Set by buyers	Negotiation	Set by marketplace provider	Set by external sellers		
	Smart contract	Yes		No			
	Payment currency	Fiat money		Cryptocurrency			

Figure 19: Classification of the DX Network business model. The blue color reflects that the revenue model is not conventional, as the DX Network was a not-for-profit organization. The project was supposed to be funded by its token sale.

5.1.3 Databroker (DAO)'s business model

The classification of Databroker's business model is depicted in figure 20. The classification is divided into two parts. The former business model (belonging to Databroker DAO – until 2019) is visualized in yellow. The business model choices that noticeably changed when the extended scope version Databroker was launched in 2020 are visualized in green. If a certain dimension only contains a green square it means that the choice has not changed. The section below will describe why certain aspects are modeled as they are.

Value proposition – Databroker DAO is described as a marketplace that enables IoT sensor data valorization, as it is now locked into private silos. They stress that there are many use cases for data that is existent, but just inaccessible. Databroker is a B2B and sensor data marketplace, with an intuitive, user-friendly interface. This makes it easy for data sellers and buyers to find the right match. We interpret this as a value proposition of easy access to data.
Enterprise data marketplace – Databroker also offers a White-labelled Platform-as-a-Service (PaaS). Data-rich organizations can use this solution to operate their own data exchange platforms for public and private data sharing. Databroker can thus be classified as an enterprise data marketplace.
Data processing and/or analytics tools – Databroker does not offer any data processing and/or analytics tools.

Marketplace participants – The mission of Databroker is to help businesses thrive by adopting the data economy. The marketplace participants are thus B2B.
Industry domain – Databroker DAO focussed on selling IoT sensor data. On Databroker all kinds of data can be exchanged.
Geographic scope – Databroker is a global data marketplace.
Time frame – Databroker supports three data source formats: data files, data streams, and API flows. Data files are static and can even be subscribed to for monthly file updates. Data streams are data that is "pushed" directly from the data provider to the data user's server. API flows provide a bi-directional service and have more function in comparison to files. Because of these functionalities, they reduce the amount of data to be exchanged and stored on multiple servers.
Platform architecture – Databroker (DAO) was and is a blockchain-supported decentralized marketplace.
Data access – API flows are accessed via an API. Other sold data is traded via the Data eXchange Controller (DXC), which is specialized software that should be installed by the data provider. The DXC acts like a proxy server and offers a secure connection to the sold data.
Data source – The data on the Databroker marketplace is provided by customers.
Matching mechanism – Databroker makes use of a team of expert DataMatch advisors. These advisors match buyers' needs with the available data assets of the providers. The DataMatch advisors can also contact the providers to see if they have the exact datasets that data buyers request.
Platform sponsor – Databroker has an independent platform sponsor.
Revenue model – Databroker charges a transaction fee of 10% of the sale price.
Pricing model – Databroker uses a pay-per-use pricing model. The data buyers pay a price that is proportional to the dataset(s) that are acquired.
Price discovery – The price of the data can either be set by the data provider itself or be based on a (negotiation) bidding process. It can choose to set a fixed price in EUR, and do or do not also allow bidding. It can also choose to not set a price and only allow for bidding. Data providers are advised to take a look at similar products on the marketplace or on the internet to understand how to set their own prices.
Smart contract – The whole process surrounding Databroker DAO was based on using a smart contract. The DXC is also integrated with a smart contract to automate the distribution of payments.
Payment currency – Data buyers can pay for the data using a fiat currency (EUR). However, when a data product is purchased the amount is converted from EUR to the digital currency Data eXchange (DTX). The DTX amount is transferred to the Databroker wallet of the data provider. This can then again be transferred to the bank account of the provider in a fiat equivalent amount.

	Dimension	Characteristics					
Service domain	Value proposition	Easy data access and/or tooling	Secure data sharing	High quality and unique data	All services in a single platform		
	Enterprise data marketplace	Yes		No			
	Data processing and/or analytics tools	Yes		No			
	Marketplace participants	B2B	C2B	Any			
	Industry domain	Any data	Geo data	Financial & Alternative data	Health & Personal data	Audience data	Sensor & Mobility data
	Geographic scope	Global	Regional	Local			
	Time frame	Static	Up-to-date	(Near) real-time	Multiple		
Technology domain	Platform architecture	Centralized		Decentralized			
	Data access	API	Download	Specialized software	Multiple options		
	Data source	Self-generated	Customer provided data	Acquired data	Multiple sources		
Organization domain	Matching mechanism	One-to-one	One-to-many	Many-to-one	Many-to-Many		
	Platform sponsor	Private	Consortium	Independent			
Finance domain	Revenue model	Commissions	Subscriptions	Usage fees	Asset sales		
	Pricing model	Freemium	Pay-per-use	Flat fee tariff	Package based pricing	Multiple	
	Price discovery	Set by buyers	Negotiation	Set by marketplace provider	Set by external sellers		
	Smart contract	Yes		No			
	Payment currency	Fiat money		Cryptocurrency			

Figure 20: Classification of the Databroker (DAO) business model. The yellow color reflects the business model choices of Databroker DAO, when the platform was still only an IoT/sensor data marketplace. The green color reflects the current business model of Databroker.

5.2 Obstacle selection: BM Stress Test X-axis

In Chapter 3 'Getting started': Conceptual background, we have already identified seven obstacle categories: trust, privacy and control, product description and product quality, matching data providers and data buyers, pricing mechanism, governance mechanisms, and others. The obstacle categories that have been identified from literature are extensive, but not necessarily complete. To make the overview as complete as possible it and see if possibly new obstacle categories needed to be added it was valuable to participate in the TRUST World Café workshop on data marketplace business models.

5.2.1 TRUSTS World Café workshop on data marketplace business models

The TRUSTS World Café workshop was organized by the TRUSTS project. The H2020 project TRUSTS "Trusted Secure Data Sharing Space" aims at creating a secure and trustworthy European Data market for personal and industrial use (https://www.trusts-data.eu/). The World Café workshop on "EU and World-Wide Data Market Trends" was organized by TRUSTS as an equivalent to a focus group of stakeholders. The goal of this workshop was to gather real-world experience from stakeholders that are familiar with the operation, usage, or research in the area of data marketplaces. TRUSTS invited the participants based on their previous experiences and because of this considered them to be experts knowledgeable on the topic. In total, around 40 individuals participated in the workshop. The participants were employed at around 30 different companies or organizations. Because the participants were diverse, experienced, and knowledgeable on the topic of data marketplaces they are representative of want we want to achieve: a rich and diverse view on data marketplace obstacles.

During the World Café, a structured approach of brainstorming took place where four virtual tables were set up. Each table discussed one topic. As we mentioned, all stakeholders were employees of companies that are concerned with data practices and already knew the definition of a data marketplace. Data marketplaces were explicitly referred to as "An online marketplace that facilitates data trading between data providers and data buyers". The participants discussed a certain topic for 15 min and then switched tables, and thus topics. One of the topics up for discussion was "business models for data markets". At this table, the stakeholders were asked for perspective on the obstacles that data marketplaces face. Because this topic directly concerns our research area, we chose to observe all sessions at this table and did not switch tables. During the sessions, we made notes on the obstacles that the participants mentioned. The sessions were not recorded, and we did not use the notes from other tables for further analysis. We only used our notes from the table that discussed data marketplace obstacles. We used a structural approach to write down these notes. We wrote down the obstacles mentioned by the participants and wrote the descriptions of the obstacle next to this. These notes are written down in the first two columns of table 3. To make these notes more useful for this research we want to see if the mentioned obstacles are different from the obstacles categories that we already identified in Chapter 3 and should be added, or if they fall in these specific categories.

The first thing that was mentioned during the workshop is that data marketplaces currently have unclear value propositions and end-user needs. The participant explained that all stakeholders need to benefit from participating in the marketplace. If the value proposition is not clear it is very hard to come up with a sustainable business model. This would fall under 'other' and relates to the thickness marketplace requirement. This is also the case for the mentioned obstacle of network effects – the phenomenon that describes how an increased number of users leads to a higher value of the service – and incentivization to participate in the data marketplace. It was even questioned if there is a market for data marketplaces. Many users appear to not see the added value of it. For trust two

explicit notions were made. The first notion was on the trust relationship between data providers and data buyers. The trust is not only dependent on the data marketplace itself but also on the stakeholder that is being interacted with. The other notion on trust was on the unwillingness of data providers to share sensitive data. Data pricing was indicated to be a key challenge. Product description and quality were also mentioned as it is often not well-structured and differs from the rules of physical goods. It was mentioned that the possibilities for bundling and aggregating data are endless and it is difficult to predict how users would want and use it. This corresponds to the obstacle of matching. Lastly, it was stressed that public data would be less difficult to develop a business model for as this data is less privacy-sensitive. This corresponds to the obstacle category of privacy and control.

The last column of Table 2 shows the links to the obstacle categories identified during the literature review of Chapter 2. In this column, we marked a few mentioned obstacles as 'other'. This means that it could be either material for new obstacles categories, or could specify some concepts for the 'other' category. This category was first created because we saw that some obstacles were quite general and not specific enough to form a separate data marketplace obstacle category (e.g. the 'unwillingness to change' is way broader than 'no generally accepted pricing mechanisms for data exist') We now believe that the descriptions given during the workshop could help make this 'other' category more specific. For data marketplaces to exist, creating network effects is vital. Also, incentivizing stakeholders to participate in the data marketplace is difficult.

In conclusion, the workshop led to some new insight regarding the 'other' category and gave some new interpretations of the more specific categories. However, no completely new specific categories have to be added.

Table 2: Summary of the obstacles mentioned by the stakeholders participating in the TRUSTS World Café workshop on "EU and World-Wide Data Market Trends"

"*What obstacles do data marketplace face?*"	**Description by stakeholder**	**Related to (obstacles from literature)**
Unclear value propositions and end-user needs for data marketplaces	-All stakeholders need to benefit from participating in the marketplace. If the value proposition is not clear it is very hard to come up with a sustainable business model. -Value could be in the facilitation of data sharing, e.g. infrastructures, match-making mechanisms, speed of acquiring the right data, anonymization of personal data, encryption of sensitive data	Other
Unwillingness from data providers to share (sensitive) data	-Data providers want to keep their data private for their competitive advantage or are afraid of hacks -Some organizations have higher risks when sharing data (e.g. a bank has a higher risk than an agricultural organization)	Trust
Building trust between data providers and data buyers	Trust is also related to your information position. E.g. people might not want to go to the police but do go to the bank when they lost money	Trust
Data pricing	Data pricing is a key challenge, therefore the business models do not seem to be viable	Pricing mechanism
It is hard for data buyers to add value to the data	Data is not always organized and well-structured, which makes it harder to recognize its value	Product description & quality
Network effects	If no one joins a data marketplace, the platforms will fail. Network effects are therefore vital	Other
Trying to come up with an incentive model to stimulate people to participate	E.g. an incentive could be that data marketplaces make it easy for a data buyer to avoid the contracts to get a data set (this could reduce coordination and search costs)	Other
Data as a product	-We are stuck in thinking in physical goods – with data the same rules do not apply -How can it be assured that the right data is being sold? It is hard to find the right use case for the data; it is sometimes perceived as a lottery game.	Product description & product quality
Data aggregation	The possibilities for bundling and aggregating data are endless: you cannot know upfront how data buyers want to use it. It is thus risky to put time and effort into this.	Matching data buyers & providers
Privacy/sensitivity of data	Public data (e.g. satellite data, agriculture data, machine-generated data from sensor industries) has fewer barriers to develop a business model for, as this is less privacy-sensitive	Privacy and control
Is there even a market for data marketplaces?	No real evidence for the need for a data marketplace by stakeholders: it is rather a technology push than a technology pull	Other

5.2.2. Obstacle structure for the BM Stress Test

The research steps on obstacles we have taken so far are conducting a literature review on obstacles and observe the discussion on obstacles by data marketplace experts in a focus group. We now want to structure these findings to make a set-up for performing the BM Stress Test. To make this set up we combine what we discovered in the literature with what we heard during the focus group. The first obstacle category from literature was **trust**. In the literature as well as in the focus groups, we heard two perspectives on trust: (1) the lack of trust in the system and/or its participants and (2) the harm of own business interests. We chose to put these in the BM Stress Test separately to see if that can lead to more accurate relationships with the BM choices. This is also the case for **privacy and control**. Because the concepts are interrelated, which makes them hard to question separately, they form one obstacle category. However, as they are not the same concept, we assume we can identify more accurate relationships if they are separated in the BM Stress Test. We chose to put the obstacle categories of **product description and quality, matching data providers & buyers, pricing mechanism, and governance mechanism** on the BM Stress Test x-axis as they are. We see no value in assessing the product description and quality separately as the description is only a means to achieve quality (in comparison to privacy and control: privacy is not only a means for control and control is not only a means for privacy, which is why they are treated separately). For the **other** category, we had to implement some specifications to make this category measurable (the relationship between a BM choice and 'other' cannot be assessed unless 'other' is specified). We could have made many choices to fill in this category, but we chose the four that seemed the most important: (1) network effects – we chose this because its importance was stressed during the TRUSTS focus group, (2) incentivizing to participate – we also chose this because its importance was stressed during the TRUSTS focus group, (3) competitors – we chose this because it was mentioned by G. Smith et al. (2016) as obstacle and it represents the link to other data-driven organizations that act in this field, (4) technical issues – we chose this because it was mentioned by G. Smith et al. (2016) as obstacle and it represents that BM choices are not only social constructs. A BM choice might a good idea in theory, but possibly cannot be realized due to technical issues. The x-axis of the BM Stress Test is depicted in figure 21.

Trust		**Privacy and control**		**Product description and quality**	**Matching providers & buyers**
Lack of trust in the system and/or its participants	*Harm of own business interests*	*Privacy*	*Control*		

→

Pricing mechanism	**Governance mechanisms**	**Other**			
		Network effects	*Incentivizing to participate*	*Competitors*	*Technical issues*

Figure 21: The x-axis of the BM Stress Test. The arrow indicates that the concepts below are a continuation of the concepts mentioned above. In the actual BM Stress Tests, these concepts are all aligned in one row.

CH 6. Entering the field & Analyzing data: the BM Stress Tests

The sixth chapter reflects the fourth and fifth steps of the theory-building process from case studies: entering the field and analyzing data. Now that the business model choices of the selected data marketplaces and the obstacle categories have been identified the Business Model Stress Tests can be performed. In these Stress Tests, the obstacles will be related to the business model choices of the data marketplaces. These relationships will be identified via interviews with data marketplace owners. This will answer sub-question 3 of the research: *"How can the identified obstacles be related to the business model choices?"* The final BM model Stress Test results also include a heat signature resulting from an impact assessment. Therefore this chapter will also answer sub-question 4 of the research: "*What is the impact of the obstacles on the business model choices?*" The final step of the BM Stress Test is pattern recognition. Therefore, in this chapter, we also want to see if there are patterns that can be identified from the BM Stress Tests. This would answer sub-question 5 of the research: "*What are the patterns between the business model choices and obstacles?"*. First, the interview protocol will be explained. The results of the BM Stress Test will be discussed per case afterward. Subsequently, we will discuss the patterns that we have identified.

6.1 Conducting the interviews

As stated, the obstacles identified in the literature research will form the a priori constructs on which the interview protocol will be based. This is valuable since these constructs can then be measured more accurately (Eisenhardt, 1989). As there are seven categories of obstacles, these categories will be used to structure the conversation. The identification of these categories allows for a semi-structured interview approach. In this approach, the interview is guided by identified themes in a consistent and systematic matter using prepared questions (Qu & Dumay, 2011). The focus is to learn more about these themes by having the participants reveal real-life experiences and include their perspectives on the complex social reality. Qu & Dumay (2011) state that this way of interviewing is especially valuable if researchers want to understand how the social world under study is perceived by the interviewees. In this case, the interviewees will be employees of the selected data marketplaces, as they should have direct insight into the current business environment and the progression of data marketplaces. During the interview we will pay special attention to (1) maintaining the flow of the interviewee's story, (2) maintaining a positive relationship with the interviewee; and (3) avoiding interviewer bias (Qu & Dumay, 2011). This will maintain the quality of the interview. Avoiding bias is especially important since the theoretical findings on certain obstacles are not necessarily practical obstacles.

Based on the literature review on data marketplace obstacles the framework of questions is as follows:

1. Could you explain if establishing **trust** is an obstacle for [name data marketplace]? If yes: how do you currently account for this in your business model?
2. Could you explain if **privacy and control** of data are obstacles for [name data marketplace]? If yes: how do you currently account for this in your business model?
3. Could you explain if **product description and product quality** of data are obstacles for [name data marketplace]? If yes: how do you currently account for this in your business model?
4. Could you explain if **matching data sellers and data buyers** is an obstacle for [name data marketplace]? If yes: how do you currently account for this in your business model?
5. Could you explain if the **pricing** of data is an obstacle for [name data marketplace]? If yes: how do you currently account for this in your business model?

6. Could you explain if the **lack of governance mechanisms** of data is an obstacle for [name data marketplace]? If yes: how do you currently account for this in your business model?
7. Could you explain if there are any **other** obstacles for [name data marketplace]? If yes: how do you currently account for this in your business model?

When preparing the interviews with the data marketplaces, it became clear that it would be more valuable to specify the questions for each marketplace. Therefore, we integrated the information found on the company websites and whitepapers into the questions. In this way, we could gain more specific insights on the relation between the obstacles that data marketplaces experience and the business model choices that they employ to cope with those obstacles.

For example, specifying the question on **matching data sellers and data buyers** for Databroker (DAO) led to the following formulation of the interview question:

"It was described in the academic literature that there are insufficient means to discover what data is required by data buyers. You provide a DataMatch service. How do you experience the use of this service?"

In total three interviews were held, one for each selected data marketplace. The interviews were held with employees that fulfilled a high function within the organization (e.g. either chiefs, (co)-founders, or members of the board of directors). Because start-ups are small and therefore often have aligned perspectives towards a common goal, we did not deem it necessary to interview multiple persons within the same business. Employees that fulfill a high function in the business were expected to be most compatible for the interviews as they would have the most experience and the best overview of the whole business including operations and (long term) visions. We transcribed the interviews and the statements that we could include in the BM Stress Test were marked. Before including their answers into the BM Stress Test matrix we checked if the statements did not contradict any of the statements made in the whitepapers or on the website. We did this to ensure that the results were based on the overall experience rather than the issues of the day.

The preparation for all interviews can be found in Appendix B.1, C.1, and D.1.

6.2 The BM Stress Test results

In this section, we will discuss the results of the interviews in the form of a Business Model Stress Test and heat signature matrix. This stress test has been done for each marketplace based on the answers acquired during the interviews. The test was done as follows. Firstly, all business model elements were aligned at the y-axis of the matrix. Secondly, all obstacles were aligned on the x-axis of the same matrix. These obstacles include trust (in the system and/or participants and separately harm of own business, privacy, control, product description, and quality, matching providers & buyers, pricing mechanism, governance mechanisms, and other (network effects, incentivizing to participate, competitors, and technical issues).

We then used the interview to fill the individual cells of the matrix. This intermediate result showed the relationships between the business model choices of the data marketplace and the identified obstacles. After this, an impact assessment needed to be conducted. The impact assessment was based on the context that surrounded the answer to a question regarding an obstacle-business model choice relationship. The cells were color-coded to reflect the expected amount of impact an obstacle has or will have on a certain business model element. A red color indicates that a certain obstacle is a possible showstopper. This means that if this is either not handled correctly, or cannot be influenced in a way that is beneficial for the operations of a data marketplace, the obstacle will make it very hard for a data marketplace to have a viable business model. An orange color indicates

that this aspect requires attention. The existence of an obstacle might not be crucial to have a viable data marketplace business model, but should be dealt with to achieve full efficiency. A green color reflects that there are no direct negative effects or in some cases even positive effects of certain obstacles. A grey color reflects that no direct relationship could be identified from either the information on the website, the whitepaper or the interview data.

Figure 22 shows an example for each color of the final heat signature. In the first row, we see a BM choice of C2B marketplace participants. We asked the employee of the data marketplace owner about the obstacle of trust, specifically on trust in the system or trust between participants. The data marketplace owner explained that a technological solution such as blockchain is very promising to establish trust – it helps to create a trust-less environment. However, the problem the data marketplace owner saw is that he believes the customers should first be explained what blockchain does. In that way, they will understand what it does and can then trust it. This led us to the following hypothesis on the BM choice/obstacle relationship: if the marketplace participants are C2B they perceive trust as an obstacle because they do not understand the technology that is used and need to be educated on it first. We marked this relationship orange because we believe that if the relationship should be true, it requires attention but will be no showstopper as education is something that relatively easy can be achieved. In the second row, we see the BM choice for a certain value proposition and the obstacle of privacy and control. We asked the data marketplace employee of VETRI how privacy and control is an issue for VETRI. The employee explained that control is a problem because they aim to let their customers monetize their data but also allow them full control over their personal data by storing everything (only) in the VETRI wallet. This is not yet possible because they are not powerful enough to compete with big-techs such as Google, Facebook, and “your bank” (see interview context) to have them stop gathering the data. So this led us to the following hypothesis on the BM choice/obstacle relationship: the value proposition of a data marketplace determines how great the impact of the obstacle of control is. In this specific case, the impact is great because VETRI can only work with a duplicate system at the moment (e.g. the personal data is stored at the big-techs as well as in the VETRI wallet). We marked this relationship as red because it is a showstopper: the full value proposition cannot be fulfilled due to the existence of the obstacle. Lastly, in the third row, we again see a BM choice of C2B marketplace participants and the obstacle of product description and product quality. We asked the data marketplace employee if they perceived product description and quality as an obstacle. The employee explained that because they work with customer user profiles they can easily send requests to their users to find the attributes that data providers are looking for. Because they do not use raw datasets, describing the data product and guaranteeing its quality are less of a problem. This led us to the following hypothesis on the BM choice/obstacle relationship: if the marketplace participants are C2B user profiles can be created that allow for specific data requests, which makes product description and quality less of a problem. We marked this relationship as green because the effect of the BM choice is positive.

BM choice	Obstacle	Relationship	Interview context	Colour (impact)
Marketplace participants: C2B	Trust (system and/or participants)	Customers need education on the technology used	*"Blockchain is actually a complexity. People don't understand blockchain today. So we need to explain what blockchain does and how they can trust the technology."*	Orange
Value proposition: All services in a single platform	Privacy and control (control)	Working with a duplicate system	*"The market strategy have now is not the full vision. It is to get value from the data rather than control. Once we have millions of users and are powerful enough we go to Google, Facebook and your bank to get them stop gathering this data."*	Red
Marketplace participants: C2B	Product description and quality	Specific user profile rather than raw dataset	*"Tell us what you are looking for, give us those attributes and we will send it to all of our users. We are going to be able to do in a very lasered and targeted way what everyone is trying to do."*	Green

Figure 22: Example of the color-coding process of VETRI for the BM Stress Test.

Figure 23 shows a part of the BM Stress Test for VETRI as an illustration. The full BM stress tests can be found in Appendix B.2, C.2, and D.2. This section will discuss the main findings per obstacle per data marketplace. These results are taken directly from the interview. It should be noted that the statements in the BM Stress Test and the resulting conclusions therefore might be biased, as the data marketplace owners might have answered the questions in a way that makes them look good. Due to limited resources and restricted time, it was not possible to organize facilitated sessions with groups of people and external domain experts that might help to avoid tunnel vision and biased conclusions during the Stress Tests themselves. The BM Stress Test and the statements provide the perspective of an employee that shared his/her thoughts on behalf of the data marketplace. The remaining sections of this chapter should therefore not be perceived as absolute truth, but rather as grounds for further exploration and analysis. In Chapter 7 'Shaping hypotheses' and Chapter 8 'Enfolding literature', we will critically reflect on the results and relationships.

		Trust		Privacy and control		Product description and quality
		Lack of trust in the system and/or its participants	*Harm of own business interests*	*Privacy*	*Control*	
Service domain	**VETRI: Personal Data Marketplace**					
Value proposition	All services in a single platform	Vision: full transfer of personal data to VETRI wallet	Business interests are not harmed, but shifted. Normally you would buy data from big tech companies.		Working with a duplicate system	
Enterprise data marketplace	No					
Data processing and/or analytics tools	No					
Marketplace participants	C2B	Customers need eduction on the technology used		Assumption: some users will not care about privacy, control and the technology used to ensure that at first: they get paid more using VETRI.		A specific user profile with certain attributes and the ability to directly ask for certain data rather than acquiring a raw data set
Industry domain	Health & personal data			Higher sensivity than other data; needs to be extended from current survey practices	Little control because big-techs aggregate personal data in return for their services	

Figure 23: An example of the BM Stress Test for VETRI

6.2.1 VETRI BM Stress Test results

Before discussing the results of the VETRI BM Stress Test it should be noted that VETRI is yet in the beginning period of its existence. The interviewee's perspective on the obstacles is therefore partly also on obstacles expected in the future and not necessarily out of the current experience with VETRI's business model.

For **trust**, several aspects were noted that require attention. Firstly, the vision of VETRI is to have a full transfer of personal data to the VETRI wallet. This implies that trust in the VETRI system is crucial to have data users transfer their data. It was mentioned that this does not necessarily have to be in the operating party, as the use of blockchain technology makes the system independent of the party operating it. Also, because the platform sponsor is a foundation there is no private interest other than serving the user. However, people don't really understand blockchain (or smart contracts) today. Therefore, education on blockchain might be needed to create trust. Concerning the harm of own business interests, it was stated that business interests are not harmed but shifted. Normally data would be bought from big tech companies. This shift is seen as positive from both consumer and user perspectives. The data consumer gives the data user more money (vouchers, gift cards) to buy more of its own product. For **privacy and control**, some aspects that require attention, as well as possible showstoppers (for the full vision) were identified. These possible showstoppers relate to the control obstacle. VETRI can be seen as a competitor of big-techs that currently aggregate personal data in return for using their services. VETRI currently and in the nearby future works and will work with a duplicate system: the data is gathered in the VETRI system as well as in the big tech systems. To achieve full control over personal data, the data should only be gathered in the VETRI system. The monetization aspect is only to give value to the data, not control over the data. Another aspect that relates to privacy and control is the higher sensitivity of personal data. This is an extension of the proven market for monetized survey practices. However, it is assumed that some users will not care about privacy, control, and technology to ensure that at first. Because of how the VETRI system works, they can get paid more than with other systems. A positive note on this obstacle is that VETRI

itself does not have to see the data for the marketplace to work. They ensure the validity of the data and privacy and control for an individual user by using a digital identity. Not every data request needs all details of this digital identity, so a user can control what can be shared and what not. Concerning the **product description and product quality,** only positive elements of the business model were identified. Because of the nature of the marketplace, specific user profiles with certain attributes are sold and can be requested for rather than acquiring raw data sets. A notification is sent to a data user when a data request matches his/her attributes. For **matching providers and buyers,** an aspect that requires is attention is the current use of data aggregators. The vision is to directly involve data consumers (companies), but the strategy is to first focus on getting users involved. It is not possible to attract both sides at once, this is the chicken-and-egg problem. Also, users will have no direct insight on what their data is worth when matched to a certain buyer but can choose whether to participate in the deal for a set price or not. The matching itself is possible thanks to blockchain technology. This partly overlaps with the **pricing mechanism** obstacle. The data consumers will set prices by trial and error. They will initially have to see if the user accepts or not. However, users will be attracted to the fact that they can earn more in general because VETRI removes the intermediaries and the digital identity implemented by the marketplace will ensure that no fake bot data is being sold. For **governance mechanisms,** it was stated that regulations (that are not yet there) should help with development. This is partly an aspect that requires attention because little control can be exercised over legal implications. Rules and regulations are different and more strict for personal data than for other data categories, which possibly means more effort should be put into complying with it. However, new regulations such as the plans for data taxes could stimulate the attractiveness of VETRI for data consumers. This is because the regulations on data taxes would require individual companies to pay for the personal data they aggregate and store, while they might not even use it. It should also be considered that the global scope of the platform implies that they should deal with the different regulations across the globe. **Other** obstacles are related to *network effects, incentivization to participate, competitors,* and *technical issues.* The network effects and competitors could be seen as potential showstoppers. The market strategy is to get millions of users for the data marketplace to succeed. A very big scope should be maintained and achieved to fight against big tech companies. These companies (e.g. Google & Facebook) have the means to destroy all a personal data marketplace is trying to achieve. For the incentivization of participants, VETRI's business model offers interesting positive elements. They state that everybody can participate and there is currently no true alternative to "just not using Facebook". Also, because of their non-profit base, they ensure that a true alternative is offered. The personal data will not shift from one party "misusing it" to another. They also expect that users can be paid up to 50% more than when data is bought from data aggregators due to the use of digital identities (no fake data) and the removal of intermediaries. The use of cryptocurrency tokens in the marketplaces allows for other ways of revenue than taking it from the users. This is because they kept a certain percentage of the tokens themselves. The coins will increase in value once the data marketplace gains popularity. Technical issues could be hacks by users, compiling and organizing the data, and acquiring new funds to continue with development and growth.

6.2.2 the DX network BM Stress Test results

Before discussing the results of the DX Network's BM Stress Test it should be noted that the company is no longer operating. This means the interviewee's perspective is based on past experience with obstacles that the DX Network has encountered.

For **trust**, the DX Network stated that there is a difference in trust in human/business sense or technical sense. Trust in a technical sense might be easier to establish, as one could probably trust a blockchain structure and audited smart contract. However, human or business trust might be harder

as nobody wants to take the first step in trusting a business. An organization should establish a reputation or track record. On the positive note, relating trust to malicious submissions, the DX Network dealt with that by having the data provider pay a listing fee per data point to discourage malicious submissions. As an attention point, it was stated that some data might not be shared because there might be a competitive advantage in it. However, if the price would be high enough it would still be worth sharing. For **privacy and control,** only positive business model elements were identified. This is mainly because the B2B data that was sold on the DX Network was already public, just not structured. It was also noted that in B2B datasets, data goes obsolete. This means that in contrast to personal data, which often stays the same over the years, B2B data is not useful or interesting anymore after a few years. For **product description and quality**, also no aspects that require attention were found. This is mainly because the DX Network used a specified data model. All data was offered in a common format and access was via a single API to all aggregated data in this format. To give the user some control over the quality of the data a whitelist and blacklist mechanism was implemented. This can be compared to an eBay review system. For **matching data providers & buyers,** the DX Network was not involved. It just provided the infrastructure to make exchange possible. The only data that was available was what the providers offered. Although queries could be done in almost natural language (because SPARQL was used as a query language), the queries had to be specific and targeted. For the **pricing mechanism,** the DX Network chose to let the sellers set the price. However, exchange data on the data marketplace DX tokens had to be purchased. As the value of the coin was not stable, the revenue the data providers received from selling data was very variable. This was perceived as an obstacle for the data providers, as they would have wanted to know upfront how much their data was worth. For **governance mechanisms** what had to be considered, and will always require attention, are the licensing terms set by the data provider. These terms allow the buyer of the data to use the data (under certain circumstances). The DX Network stated that they provided a line of defense against web scraping (of open platforms) since it is often difficult to legally protect yourself against that. By using a data marketplace the data that would be scraped would now be monetized. For **other** obstacles, some findings were done for the *incentivization to participate, for the competitors, and the technical issues.* The *technical issues* appeared to be the real showstoppers for the DX Network and seem to form the main reason that the marketplace is no longer operating. The DX Network made use of a technical (command line) interface. The complicated infrastructure led to slow adoption amongst potential users. In retrospect, the DX Network stated that the data marketplace should be focused on the user experience and not on the technology on the backside. The marketplace participants should perceive the interface as simple and easy to work with. This also relates to the payment currency the marketplace used: the DX token, a free-floating currency. Purchasing the tokens that were needed to participate in the marketplace was perceived as difficult. Also, the buyers of the tokens turned out to not be the users of the marketplace. For *incentivization to participate* it was noted that the DX Network wanted to focus on promoting open access to information while receiving a fair share of the wealth that was contributed by sharing. The data providers received a query fee for each time the data point was queried.

6.2.3 Databroker BM Stress Test results

Because Databroker expanded its scope from being an IoT data marketplace to an IoT and B2B data marketplace offering all kinds of data, it is interesting to first discuss the showstopper for their old business model. The showstopper that was identified belongs in the category **other** under *technical issues.* Databroker stated that they expanded their scope because the IoT technology was not mature enough yet. The problem with the current 4G network is that the communication between a sensor (an IoT device) and a network leads to relatively high energy consumption. According to Databroker

5G technology will solve this problem and speed up IoT adoption. However, when the switch was made in 2019 the knowledge of the market and IoT was too limited to have a business model that was ready for the market. Pursuing the initial idea was financially not possible.

The remaining part of the Stress Test was done using Databroker's current business model. For **trust,** it was stated that technologically speaking there seems to be trust. The participants can control with whom and how data is shared and can stop this at any moment. Concerning the *harm to business interests,* an interesting notion was made about the company being an enterprise data marketplace. It was stated that many data exchange use cases are better suited to a private marketplace (the PaaS solution that Databroker offers). For example, a use case on Databroker is the exchange of safety studies on the chemicals used in consumer products. Databroker argues that legacy data exchange technologies make it difficult to share the safety study data. In the infrastructure that Databroker provides no intermediary has to be included, and the risk of exposing the sensitive data becomes less. Referring to the *incentivization to participate* of the **other** obstacle category, motivations to participate in such an exchange include the high costs of conducting safety studies, ethical considerations (e.g. animal testing), and the need to comply with (international) regulations. Regarding **privacy and control**, Databroker sees itself as a mediator between providers and buyers and thus not feels directly responsible for the privacy of the users. Databroker cannot see the content of the data that is shared but can ban users from the marketplace if they are not compliant with the terms and conditions. They also note that concerning control potential marketplace participants still fear that data is leaked to other third parties. The question that should be asked is how to keep control over the data in case of fraud. For **product description and quality,** it was mentioned that the quality of the data cannot be guaranteed, as the user requirements differ and the content that is shared cannot be accessed by Databroker. However, as a result of the use of smart contracts, the buyer can still have a warranty on the data. The payment will be frozen when a complaint is filed about the data provider. For **matching providers & buyers,** the way of working that requires attention is that matching at this moment is done manually. The marketplace does not have enough users to automate this yet by for example designing an algorithm. This is also the case for the **pricing mechanism** used in the marketplace. Prices are either set by external sellers or by the means of negotiation since there are not enough marketplace deals to automate pricing mechanisms. It was mentioned that this might be valuable in ten years. However, this will remain difficult because it is yet unclear what variables to include in a price discovery mechanism. Complex features of data complicate the pricing because data is non-linear, non-rivalrous, a potential liability, and its value increase through refinement, with holder, and over time. Regarding **governance mechanisms,** it was noted that conformance is important for data sharing. Meaning that there should be clarity on who is responsible if something goes wrong in decisions that are based on data. This problem is located on a higher level than the services of Databroker, as they maintain a decentralized platform in which only the two parties who participate in the data exchange have ownership over the data. However, it is still something that needs to be considered to make all kinds of data exchange valuable. For the **other** obstacle category, some interesting remarks were made that would possibly have a positive relationship in dealing with *network effects, incentivization to participate, and technical issues.* For the network effects, it was mentioned that Databroker's strategy to 'solve' the chicken-and-egg problem is to start with attracting data buyers. The data buyers will be assisted in finding the right data provider. When a data provider made one sale on the data marketplace, he might be more interested in staying involved. For the *incentivization,* Databroker also focuses on the ease of use for the data buyers. The data buyers can use either a debit card or credit card to pay. There is no need to buy DTX tokens first. The use of these tokens on the marketplace however makes it easier to automatically (via smart

contracts) share revenue between different partners. Also, by creating a cryptocurrency the project can be funded by an ICO (initial coin offering). Regarding *competitors,* Databroker mentions that many marketplaces have a centralized structure, while only around ten marketplaces worldwide are decentralized. The centralized alternatives are assumed to be outcompeted on security, compliance, transparency, neutrality, automation & scalability.

6.3 Analyzing data: Pattern recognition

This section reflects the fifth step of the theory-building process from case studies: analyzing data. This corresponds to the fifth step of the BM Stress Test, which is pattern recognition. Patterns can be identified either within-case or cross-case. The within-case patterns will be discussed first. Subsequently, the cross-case patterns will be discussed.

6.3.1 Within-case: VETRI

For VETRI the BM Stress Test indicates that almost all business model choices have inconsistencies. This means that a choice leads to a green relationship with one obstacle, but a red or orange relationship with another obstacle. It also seems that VETRI's value proposition of having "all services in a single platform" is the most problematic business model choice. This choice leads to three red relationships. This is mainly because their value proposition implies that their competitors are powerful big-tech companies such as Facebook and Google that have access to a lot of personal data of their consumers. The only way to deliver the value that VETRI intends to is getting millions of users, which leads to bargaining power and should enable them to collaborate with big-techs. In this way, the duplicate system can be discarded and all personal data of a user would only be stored in the VETRI application. The only choices that don't seem to lead to any consistencies are the *platform sponsor* (private) and the *payment currency* (cryptocurrency). For the platform sponsor, the use of a foundation would show the users that there is no private interest and that it is thus a true alternative to giving your data away "for free" to big-tech social media companies. 'For free' is stated in quotation marks, since the user gets to use the service of the company in return for their personal data. As explained the use of cryptocurrency allows VETRI to make revenue in other ways than taking it from its users. However, it could be argued that since it was stated that it was an issue that users don't understand blockchain technology today, users also won't understand the use of cryptocurrencies.

6.3.2 Within-case: The DX Network

The pattern that can be identified from the DX Network's BM Stress Test does not relate to one business model choice, but rather shows one main obstacle that has only red relationships with business model choices: *the technical issues.* The business model choices directly show that innovations like these are socio-technical. This means that innovation is more than what technology can enable on its own. The sociological aspects should make sure that the technology is understood and used by the actual consumers. Without the adoption of the users, there is no value in creating innovation. In this case, the technical command-line interface led to slow adoption, the buyers of the token were not the users of the marketplace and the purchasing process of these tokens was perceived as difficult. The DX Network itself also saw this as the lesson learned from their failure: a data marketplace should be focussed on the user experience. Not the technology on the backside.

6.3.3 Within-case: Databroker

The main pattern that can be identified from the Databroker BM Stress Test is how the coloring changed from red when discussing the IoT data marketplace (Databroker DAO) to only green and orange relationships only when discussing the current B2B and IoT data marketplace business model. This implies that the obstacles were too great to overcome to have a viable IoT data marketplace business model. However, when combining it with B2B data it seems that a viable business model can be achieved. Databroker stated that the knowledge of IoT is still too limited, the technology is not ripe yet, and the market for data that is not related to IoT devices is already existent which makes creating a good business model easier. Business model choices that have only green relationships, according to Databroker, are being an enterprise data marketplace, having a decentralized platform architecture, using a smart contract, and using fiat money as payment currency for the buyer. The fiat money is however converted to cryptocurrency within the Databroker platform.

6.3.4 Cross-case patterns

From a cross-case point of view, we can conclude that all data marketplaces have (or had) a relatively low amount of users and adoption rate. This is while, due to network effects, the marketplaces would offer more value, or would only be able to fully live up to their value proposition, if they are to have a great number of users. It is not clear what the main obstacle is that should be overcome to benefit from these network effects. However, we note that while the data marketplace owners have a clear vision for their platform design and the functions it should carry out, their strategy for platform adoption among their users is less clear. In all cases, the showstoppers (red relationships) relate to the socio-technical implications of technological choices. For VETRI, this is mainly the import data streams from all kinds of online platforms. Even though this might be technically possible, it is not socially beneficial for the promised control over personal data when there is no collaboration with these online platforms. For the DX Network, we already explained the reasons for them being technically sufficient but socio-technically insufficient. Databroker DAO (the IoT data marketplace) also shows that while sharing IoT data is technically possible, it is difficult to turn the technological data trading process and into a viable business model. The technical opportunities are there but it is difficult to create value from them.

CH 7. Shaping hypotheses: The resulting correlational relationships

The seventh chapter reflects the sixth step of the theory-building process from case studies: shaping hypotheses. In the sixth chapter, we performed the BM model Stress Tests and discussed the results per obstacle per data marketplace. As we want to create an explanatory model at the end that includes as many relationships as possible these results should be integrated. Therefore, in this chapter, we collate the information on each obstacle from all three BM Model Stress Tests discussed in Chapter 6. We also explicitly refer to the corresponding business model choice of the data marketplace. In this way, we can answer sub-question 6: *"What are the hypotheses on data marketplace obstacles and business model choice relationships?"*

7.1 The correlational relationships

We structurally ordered the hypotheses on the relationships (or: claimed relationships) by obstacle category in the tables below (3.a t/m 4.f) The relationships are directly taken from the BM Stress Tests and we should thus analyze them further. The relationships in the tables are yet only correlational, they show the BM choice that a data marketplace has made and how this, according to them, is related to an obstacle that should be overcome. The problem with correlational relationships is that the correlations do not yet explain anything; the business model choice and the relationship to the obstacle just co-exist in the cases that we studied. The stress pattern tables and their interpretation below thus form an intermediary result and a stepping stone for the explanatory model that we want to create in the end. However, by mapping the relationships in this way, we can easily refer to them and use them as input for the explanatory model we will later create.

Before we start discussing the explicit relationships between the BM choices and obstacles we should note another problem we ran into. This problem exists because in the taxonomy that is used some BM choices are very explicit and comprehensive and other choices are not. For example, a data marketplace can choose to sell sensor data or not. This is a clear choice that does not include additional matters. If a choice is comprehensive is it easier to identify accurate relationships with obstacles. In this case, the BM choice for sensor data makes creating a viable business model difficult because IoT technology is yet underdeveloped. Other BM choices, such as the value proposition are way less comprehensive. The taxonomy states that theoretically there are four possible value propositions – easy data access and/or tooling, secure data sharing, high quality and unique data, and all services in one platform. For this research, listing the value propositions in this simplified matter does not help us to achieve an accurate relationship with an obstacle. This is because we see the value propositions of data marketplaces in a broader perspective: it indicates every way in which *a firm will deliver value to its customers.* For example, we identified the DX Network to have 'easy data access and or data tooling' as a value proposition. To make this happen, among other things, they implemented a review system with a blacklist and whitelist mechanism to the user some control over product quality. The relationship between easy data access and the review system for product quality is not immediately clear. This is because 'easy data access and/or tooling' cannot fully capture how a firm will deliver value to its customers. The aim of a data marketplace, and thus how they formulate their value proposition, can be to provide easy data access. However, if all data is malicious or of very low quality, there is no use in providing easy access to that. The data will be useless and the data marketplace will not attract any customers making it impossible to deliver any kind of value. Because of this problem we chose to abandon the BM choice options for the value propositions according to the taxonomy for the remainder of our research. Instead of a specific BM choice, we note it down as 'value proposition' in general. This "BM choice" (value proposition)

includes the relationships to obstacles that we could not link to other explicit BM choices. However, they are still very valuable to get insight into data marketplace business model viability.

First, we have depicted the green relationships in table 3.a t/m 3.g. These are the relationships where a BM choice has a positive relationship with an obstacle: the data marketplace owner perceives these obstacles as something their business model can and will deal with. We cannot validate these relationships, it is only based on the experience of data marketplace owners. In the context of this research, it is not possible to criticize what they experience in daily practice. Therefore, we argue that we can use these relationships as direct input for the explanatory model that we will later create. We will now shortly discuss the relationships in all tables. We will refer to the case studies that are the source of the claimed relationships between brackets (described in chapter 6). The R(number) refers to the relationships that can be found in the tables.

Table 3a (Trust): A BM choice for a decentralized architecture and the use of smart contracts will have a positive effect on trust as it creates technological trust (R1-VETRI/DX Network/Databroker). A BM choice for an enterprise data marketplace has a positive effect because creates a private atmosphere that is more trusted (R2-Databroker). When choosing to use finance and alternative data less trust is needed because the data is not private but public (R3- DX Network). We want to note that we do not believe that all finance and alternative data is always public. This statement will only apply to a part of the finance and alternative data. Choosing for health & personal data will not do additional harm to trust as the data would normally be acquired by big-techs for "free" (R5-VETRI). Having a private platform sponsor that does not generate profit on the data sales but only performs asset sales (cryptocurrency token sale) to finance its operations shows that there is no private interest other than serving the user, which generates trust (R4-VETRI). We should note that this is a very specific interpretation of the BM choices. The comparison only holds if the private platform sponsor is a foundation and the asset sales are not for profit, but for maintaining operations. Implementing a pay-per-use system for providers (as well as buyers) can create trust because as fees have to be paid per submission, malicious submissions will only cost them money (R6-DX Network). Again, this is a very specific interpretation of the BM choice. We cannot say that choosing pay-per-use positively affects trust. It can, but it should then be implemented in a specific manner.

Table 3b (Privacy and control): Choosing a decentralized architecture and customer-provided data will positively affect privacy and control because the data marketplace owner has no access to the data. Only issues of privacy control can arise between the data provider and the data buyer (R1 - VETRI/Databroker). Even if a data marketplace chooses to sell customer-provided data, and does not have access to the content, users can still be banned from the marketplace if the data shared is not compliant with the terms and conditions (R2 - Databroker). This would positively affect how users perceive the privacy and control issues. However, we want to note that this can only be done if it can be noticed from the outside that a dataset is not compliant. For example, if the product description states that it is a list of personal email addresses for every person in a certain neighborhood, the data marketplace owner can intervene. If it does not show on the outside, this will be difficult. Providing privacy and control does not even have to be part of the value proposition, the data marketplace can act as only a mediator. In that case, the data providers and data buyers carry all privacy and control responsibilities (R5 - Databroker). Choosing specialized software for data access, such as a digital identity, will increase perceived privacy and control (R3 - VETRI). When choosing to sell finance & alternative data, the data goes obsolete after several years. This means that the consequences for privacy and control are less of a concern (R4 - DX Network).

Table 3c (Product description and quality): Having C2B marketplace participants allows a data marketplace to directly ask the customers for certain data, rather than having to filter the requested data out of existing datasets. The quality of sold data will be perceived as high because it exactly matches the request (R1 -VETRI). This approach also concerns making the BM choice for selling up-to-date data as the C2B customers can then get a notification when the request matches their attributes (VETRI). The value proposition of a data marketplace should include a way to account for the quality of the data being sold. For example, having a review system that is based on a whitelist (high-quality data providers) and blacklist (low-quality data providers) mechanism (R3 - DX Network). Choosing a smart contract will not help in guaranteeing data quality but can give the data buyer a warranty on the data may the acquired data be of insufficient quality (R4 - Databroker).

Table 3d (Matching data providers & buyers): Using a decentralized architecture that includes blockchain technology makes automatic matching possible (R1 – VETRI). If a data marketplace itself is not involved in the matching mechanism the value proposition should include something that makes it easy for the data providers & buyers to make the matches themselves. For example, by creating an easy query language (R2 - DX Network).

Table 3e (Pricing mechanism): The situation where pricing data is less of a problem is when it is not about the absolute value (in money) of the data, but the relative value (in money) of the data. We observed this in the case of a personal (C2B) data marketplace. Because their value proposition includes that they aim to pay customers more for selling their personal data (by having no intermediaries, by guaranteeing that buyers that no bots are included in the and by generating own revenue from token sale only) the absolute value becomes less important (R1 – VETRI). In this case, getting some money for your data, and more than at competitors is better than getting less or none.

Table 3f (Governance mechanisms): In general, few governance mechanisms for anything related to data aggregation or trading exist. So if the value proposition of a data marketplace offers anything that can contribute to this void it can positively affect this obstacle. For example, the DX Network stated that they could provide a line of defense against web scraping, as there currently is no legal protection available (R1 – DX Network). Web scraping would in the end lead to about the same results as acquiring the data on the data marketplace, but would be more (1) costly and time-consuming for the “data buyer” (web scraper) and (2) would not lead to any gains for the data provider (the organization being scraped). We find that value propositions can be boosted by the implementation of certain higher-level regulations. We will look into this in more detail in the next chapter.

Table 3g (Other): If a data marketplace chooses to sell health & personal data everybody can be included since every individual possesses this sort of data. This would be good for increasing the *network effects* (R1-VETRI). Choosing a global scope will also help with creating network effects. We saw this in the case of the DX Network. They aimed to create the function to query many databases from all over the world at once instead of one at a time per region (R2 – DX Network). This function would be more valuable when more data providers join as this increases the chances that a data buyer would find something worth buying. Then it was seen in two cases that, among other reasons, the data marketplace owners chose to use cryptocurrency as this allows for other ways of making revenue than taking it from transactions between the data provider and data buyer (R3 – VETRI/DX Network). Another case showed that creating a cryptocurrency means that the project can be funded by an ICO (Initial Coin Offering). This choice also makes it easier to distribute revenue among partners (R8 – Databroker). A BM choice of fiat currency has the advantage that is it easy to buy on the data marketplace. There is no need to buy crypto tokens first. However, a choice for a

decentralized architecture means that the competition (with other marketplaces) is less, as currently there are more centralized data marketplaces (R9 – Databroker).

Trust (+)	BM choice
R1. Technological trust (blockchain, smart contracts, control over data exchange)	Decentralized/ Smart contracts
R2. Private use cases on private data marketplaces	Enterprise data marketplace
R3. Data is already public (it is just not structured)	Finance & Alternative data (B2B)
R4. No other interest than serving the user	Private platform sponsor/ Asset sales (not-for-profit) / Cryptocurrency
R5. Interests are not harmed but shifted. Data would normally be acquired by big-techs	Health & personal data
R6. Listing fee to discourage malicious submissions	Pay-per-use

Table 3.a: Trust relationships (+)

Privacy and control (+)	BM choice
R1. Data marketplace operator has no access to data content	Decentralized architecture / Customer provided data
R2. Ban users from marketplace if data shared is not compliant to terms and conditions	Customer provided data
R3. Use of a digital identity	Specialized software
R4. Data goes obsolete	Finance & alternative data
R5. Platform operator acts as mediator: No direct privacy responsibilities	*Value proposition*

Table 3.b: Privacy and control relationships (+)

Product description & quality (+)	BM choice
R1. A specific user profile with certain attributes and the ability to directly ask for certain data rather than acquiring a raw data set	C2B
R2. Phone sends notification when request matches your attributes	Up-to-date data
R3. Review system – whitelist and blacklist mechanism	*Value proposition*
R4. Quality cannot be guaranteed, but smart contract gives a warranty on the data	Smart contract
R5: Specified data format; access via single API	API

Table 3.c: Product description & data quality relationships (+)

Matching (+)	
R1. Matching is possible due to blockchain technology	Decentralized architecture
R2. Performing queries is almost natural language (SPARQL)	*Value proposition*

Table 3.d: Matching data providers & buyers relationships (+)

Pricing mechanism (+)	
R1. Ability to pay more: no intermediaries + no bots, revenue comes from tokens sale	C2B / *Value proposition*

Table 3.e: Pricing mechanism relationships (+)

Governance mechanisms (+)	BM choice
R1. Line of defense against web scraping – no legal protection	*Value proposition*
R2. Assumption: Regulations will help instead of counteract	*Value proposition*

Table 3.f: Governance mechanism relationships (+)

Other (+)	BM choice	Other (+)	BM choice
R1. Everybody has personal data; currently there is no true alternative to "just not using Facebook"	Health & personal data	**R4.** Buyers can use debit or credit card to pay. There is no need to buy crypto tokens first	Fiat currency
R2. Ability to query many databases at once, but perceiving it as one big database due to single API structure	Global scope	**R5.** By creating a cryptocurrency the project can be funded by an ICO. It also makes it easier to distribute revenue among partners.	Cryptocurrency
R3. Cryptocurrency allows for other ways of making revenue than taking it from the user	Asset sales (not-for-profit) Cryptocurrency	**R6.** Little amount of decentralized data marketplaces; little competition	Decentralized architecture

Table 3.g: Other relationships (+)

The relationships we just discussed are positive instances of the obstacle categories. If we assume that the experience from the data marketplace owners is legit, these obstacles should not be a problem when making "the right" BM choices. Therefore, we are more interested in the red/orange relationships. These are the consequences of certain obstacles that from a practical perspective also remain difficult By further analyzing these problems we can identify if some obstacles are really "unsolvable" or if there are solutions that the data marketplaces did not yet think of. These relationships are depicted in table 4.a t/m 4.f and described below. In these case studies no orange or red relationships were found with the product description/quality obstacle category. Therefore, this category is not included.

Table 4a (Trust): It is difficult to fulfill the value proposition of a data marketplace without having a track record or reputation that will account for trusting the business (R1 – DX Network). A B2B marketplace will deal with the issue of establishing trust between the businesses that should exchange data (R2 – DX Network). The customers of all data marketplace owners should acquire trust in the operating party to let them live up to their value proposition (R3 – DX Network). Also, technical BM choices that can also be used for creating trust (decentralized architecture, smart contracts, cryptocurrency) will not necessarily create trust as technology is often not understood (R4 – VETRI).

Table 4b (Privacy and control): A data marketplace that operates with B2B marketplace participants should be on top of preventing leaks to 3rd parties. It is not yet clear how to keep control in case of fraud. When choosing health and personal data and cooperating with C2B participants (personal data marketplaces) and offering a value proposition that aims to make the monetization of personal data and control over personal data possible the problem for privacy and control remains that the data marketplaces will in first instance work with duplicate systems as big-techs also gather this data (explained in chapter 6.2.1) (R2/R4 – VETRI). However, this is again a very specific combination of business model choices that go beyond taxonomy definition. For example, if the personal data gathered is a specific instance of health data that cannot be accessed by big-techs this would change the problem. In most cases, choosing for health & personal data will bring with the problem that it has a higher sensitivity than other sorts of data (R5 – DX Network). However, in some cases, it is assumed that privacy and control will have no priority for users. We saw this in a C2B environment (R3 – VETRI).

Table 4c (Matching): If a data marketplace chooses to let the customers provide the data, the availability of the data on the data marketplace is in control of the customers. This means that the matches that can be made are also dependent on this (R1 – DX Network). We observed for both a C2B as well as a B2B business model choice that is difficult to attract both market sides at once (R2 –

Databroker/VETRI). As matching is based on having two market sides involved, this chicken-and-egg problem also forms a direct challenge for making matches. Also, any data marketplace should include a certain kind of approach to match-making in its value proposition. Whether this is manually or automatically, match-making is of high importance (R3 – Databroker). For example, it is good to offer data that is of high quality but if no other party can be found that wants to buy it, it is still useless.

Table 4d (Pricing mechanism): If a data marketplace lets the external sellers set the price the data buyers might be disadvantaged. As it is unknown what variables should be included to calculate the price (there is no uniform pricing mechanism), they will not know upfront if they will get value for their money (Arrow paradox). On the other side, letting buyers set the price will leave the data providers with no insight into what their data is worth (R5-VETRI). This can be adjusted a bit by also including a negotiation mechanism in the data marketplace for discussing the price (R3 – Databroker). Data buyers now often set the price by trial and error. Another disadvantage arises when the data marketplace uses cryptocurrencies. As this stand-alone market is very volatile, the coin value is unstable, which makes the data set that is already difficult to price fluctuate in its value even more (R1 – DX Network).

Table 4e (Governance mechanisms): It was claimed by VETRI that the value proposition of data marketplaces can be given a push in the right direction if more regulations for data trading are developed (R1 – VETRI). We will look further into this in the following chapter. Two clear relationships to governance mechanisms are the BM choice for health & personal data and global scope. A choice to sell health and personal data means that data marketplaces should take into account that the regulations for this category are different than for other categories (for example due to the GDPR) (R2 – VETRI). Also, choosing a global scope means that they should consider that there are different regulations across the globe (R3 – VETRI). We observed that when choosing B2B marketplace participants a great problem is that not clear who has the responsibility for data-driven decisions. This might make businesses reluctant to participate in the marketplace, as they are afraid to carry the full responsibility of the consequences that might result from using the data as grounds for important decisions (R4- Databroker).

Table 4f (Other): The scope and content of the value proposition of a data marketplace are important to assess if a data marketplace business model is viable. For VETRI we saw that the creation of network effects and a big scope are essential for living up to their value proposition (R1 – VETRI). At this moment, this is not the case, which is why it is a potential showstopper. Each addition in the value proposition brings new challenges for the viability of the business model. For example, they now operate by spreading surveys to their users and letting their users sell that (personal) data. By definition, they are a data marketplace because they facilitate data exchange. However, including data that is imported from social media platforms (which they aspire for the future) will lead to new challenges, and might decrease the viability of the business model. Thus, there is a direct relationship between the formulation of the value proposition and the viability of data marketplace business models. Making a BM choice for sensor data includes dealing with IoT technology and possibly the development of the 5G network (R2 – Databroker). There thus is a dependency on the technology readiness level of these technologies. As they are currently underdeveloped, it is a potential showstopper. In the DX Network case, we saw that the choice for a decentralized platform architecture was perceived as difficult and its interface led to slow/too little user adoption (R3 – DX Network). However, again, we cannot say that this is a valid relationship. This is because in this specific case the BM choice for a decentralized platform architecture also meant a technical interface. If a data marketplace finds a way the reap the benefits of using a decentralized architecture and makes the interface for its customers easy to use, it does not have to be a problem.

If a data marketplace chooses for using cryptocurrencies it should keep track of the buyers of the tokens. These buyers should also be users of the data marketplace for the business model to work (R4 – DX Network). Then, probably with any kind of industry domain choice, there will be competition. However, the competition for health and personal data might be the biggest (R5 – VETRI). If these competitors are big-techs they also possess the power to fully destroy a (personal) data marketplace (R6 – VETRI). Of course, this then again is dependent on the value proposition of that personal data marketplace. If they choose to not interfere with the operations of big-techs, they can co-exist.

Trust (-)	BM choice
R1. Need of track record; reputation	Easy data access and/or tooling
R2. Trust between businesses	B2B
R3. Trust in operating party	All services in a single platform
R4. No knowledge about technology (technological trust)	C2B/Decentralized /Smart contracts/ Cryptocurrency

Table 4.a: Trust relationships (-)

Privacy and control (-)	BM choice
R1. Leaks to 3rd parties: how to keep control in case of fraud	B2B
R2. Duplicate systems of personal data storage	All services in a single platform + Health & personal data
R3. Assumption: privacy and control will be no priority for users	C2B
R4. Little control because personal data is aggregated by big-techs	Health & personal data
R5. Personal data has a higher sensitivity than other data	Health & personal data

Table 4.b: Privacy and control relationships (-)

Matching (-)	BM choice
R1. Only infrastructure for exchange: availability of data in control of providers	Customer provided data
R2. Both market sides cannot be attracted at once (chicken-and-egg problem)	C2B/B2B
R3. Matching is done manually since there are not enough users to automate it	Easy data access and/or tooling

Table 4.c: Matching data providers & buyers relationships (-)

Pricing mechanism (-)	BM choice
R1. Unstable coin value	Crytocurrency
R2. Not enough marketplace deals to automate pricing	Price set by marketplace providers
R3. It is unknown what variables should be included to calculate the price	Price set by external sellers/negotiation
R4. Prices are set by trial-and-error	Price set by buyers
R5. Users (providers) have no insight on what their data is worth	Price set by buyers

Table 4.d: Pricing mechanism relationships (-)

Governance mechanism (-)	BM choice
R1. Regulations should help data marketplace progress (but is not yet in place)	*Value proposition*
R2. Different regulations for personal data than other categories	Health & personal data
R3. Different regulations across the globe	Global scope
R4. Who has responsibility for data-driven decisions?	B2B

Table 4.e: Governance mechanism relationships (-)

Other (-)	BM choice
R1. Network effects/big scope are essential to live up to value proposition	*Value proposition*/Global scope
R2. IoT technology not ready yet (5G might help)	Sensor data
R3. Technical interface leads to slow/too little user adoption	Decentralized platform architecture
R4. Token sale: buyers of the token were are not the users of the data marketplace	Cryptocurrency
R5. Big-techs and data aggregators are also competing for data	Health & personal data
R6. Big-techs have the power to destroy a personal data marketplace	*Value proposition*

Table 4.f: Other relationships (-)

It should again be noted that these tables are based on case-study research from a data marketplace owner perspective. They should not be perceived as the ultimate truth, but do give a broad overview of the matter.

7.2 Discussing the results

From what we saw in the BM Stress Test pattern tables we can conclude that the relationships between BM choices and obstacles are vague. This has multiple reasons. We believe part of the reasoning for this is that we overestimated the explanatory power of the taxonomy. The taxonomy uses generic categorization, while the solutions of data marketplaces to obstacles that we saw in the case studies are too specific to connect to a specific business model choice. Many things go beyond taxonomy definition. For example, the taxonomy uses "finance and alternative data" as the industry domain. Technically this is correct, but when connecting it to data marketplaces obstacles even this is too general. For example, following our method of connecting a BM choice to an obstacle, we would say BM choice – finance and alternative data, relationship to privacy and control obstacle: data is already public, so fewer issues for privacy and control. The problem is that not all instances of finance and alternative data that can be sold, only some, are public. So, by formulating the BM choice as "finance and alternative data" it does not always have the same relationship to the privacy and control obstacle. This makes a very hard to make valid claims on specific BM choices and obstacle relationships.

We also see that a lot of relationships are only valid under certain conditions and always need additional explanation. For example, because we structurally followed our BM Stress Test method we resulted in a relationship between asset sales of cryptocurrency tokens and trust. Although practical experience shows the truth in this relationship (see Ch 6.2.1) it cannot be generalized to the BM model choice of asset sales and using cryptocurrencies leads to trust. It is very context-specific and depends on how the business model choices are being realized. Also, the relationships between BM choices and obstacles can be broader than we observed. For example, the case study on VETRI led us to believe that with C2B marketplace participants there will be instances when privacy and control will be no priority. Following this stress test method, the relationship is: C2B – privacy and control are not always a priority. However, it can just as well be the case that this is also applicable for B2B marketplace participants. It only did not come forward in the B2B case study. Another example is that of the BM choice of 'price set by buyers'. During the case study of VETRI, we heard that the price is set by buyers by trial and error. This would let the data providers be in control: they can choose whether they want to accept the price and sell their data for it or not. If it appears that no transactions will take place, the buyers will need to increase the money they initially offered. We see two problems in this approach: first, it is not efficient. The price might have to be adjusted several times before a transaction can be made. Secondly, the same argumentation can be used when external sellers set the price. The method stays the same, only the perspective changes. This again means that we cannot say anything concrete about the relationship of a BM choice with an obstacle. The problem thus is that the choice does not reflect the broader picture, which is what we should prevent when creating an explanatory model. In other words, the case study led us to see the correlations but does not give any clear relation on a certain BM choice and a certain way to overcome an obstacle due to its relationship with the obstacle. We also believe that a lot of the problems we ran into are also due to the matter that there is a direct relationship between the formulation of the value proposition and the viability of a data marketplace business model. For example, we analyzed the case of VETRI as a personal data marketplace. This specific personal data marketplace wants to offer its customers the chance to monetize their personal data and to have control over it. At this moment, the control part is a great issue while the monetization part is slowly

setting a course. So, differences in value proposition will also lead to differences in the impact of obstacles, and thus business model viability.

In conclusion, the results of the Stress Tests are vague and not very comprehensive. However, the correlations that we did identify are not useless. They provide us with some insights that we otherwise would not have had. At this point, we realize that we need to think broader than specific business model choices. Therefore, in the explanatory model that we aim to create, we will use higher-level concepts such as the business model dimensions or business model strategies instead of business model choices. The correlations that we have found will still be for input for the explanatory model, only in a broader sense. We know that the choices you can make within a dimension can be of influence, but we cannot go into detail on the specific choices. For example, although we cannot say that either 'price set by sellers' or 'price set by buyers' will help overcome the pricing mechanism obstacle, we can say that the choices made in the price discovery dimension (that includes price set by sellers/ by buyers) affect the pricing mechanism obstacle.

Before we can create an explanatory model we will first reflect on the negative elements of the obstacles categories. This is because these negative elements were only based on the opinions of the data marketplace owners. Before any of it can be included in the explanatory model we need to critically reflect on it.

CH 8. Enfolding literature: Reflecting on the relationships

The eighth chapter reflects the seventh step of the theory-building process from case studies: enfolding literature. In this step of the theory-building process similar as well as conflicting literature should be considered to form a final explanatory model. The aim is to answer sub-question 7 of this research: *"Can we verify the relationships that resulted from the BM Stress Tests?"*

So far we have identified positive and negative instances of the obstacles that in some way relate to the business model of a data marketplace. As we discussed, this is not necessarily a specific business model choice but can be a broader concept such as a business model dimension or business model strategy. However, until this moment we have assumed that negative relationships are indeed negative and the positive relationships are indeed positive. As we said, we will not argue with the positive relationships. If they form no problem in practice, it does not matter if they are claimed to form a problem in theory. What is necessary, is reflecting on the negative relationships. This is because we only claimed them to be negative based on the opinions of the data marketplace owners. If we critically look into these relationships, we might find another perspective. Only after critically reflecting on these relationships we can include them in the explanatory model.

The reflection on the relationships will be done in two ways: by enfolding (academic) literature and by interviewing experts on the claimed negative relationships. In the first part of this chapter, we will use (academic) literature to look into solutions. The solutions can be connected to the negative relationships in the heat signature (the red and orange cells). If a solution connects to a certain red or orange cell of the matrix, it can be assumed that the relationship between the obstacle and the data marketplace business model is not necessarily negative. Upon implementation of the solution, the obstacle might be controlled and the relationship might change of color, and thus impact.

In the second part of this chapter, we will reflect on the relationships between obstacles and business models by interviewing experts on data marketplaces. In both parts of the chapter, the reflection will be done per obstacle category. The aim of this chapter is not to validate any solutions. It is rather to further analyze the relationships so that this can be included in the explanatory model for and used for future research directions.

8.1 Reflection the relationships: a literature perspective

To identify the solutions in a structured manner, the six categories where orange and red relationships were found will be discussed one by one. Where necessary, literature will be scanned to find out if there are possible solutions that relate to the main obstacle and its relationships with the business model choices. As the aim of this chapter is not to find all solutions for the obstacles and validate them, five relevant articles will be chosen per obstacle to give a first indication of the possible solutions and their implication for the state of the relationship.

8.1.1 Reflection on trust

Looking at the relationships identified from the BM Stress Tests it can be seen that four relationships have an orange impact.

Trust (-)	BM choice
R1. Need of track record; reputation	*Value proposition*
R2. Trust between businesses	B2B
R3. Trust in operating party	*Value proposition*
R4. No knowledge about technology (technological trust)	Decentralized /Smart contracts/ Cryptocurrency

Table 4.a: Trust relationships (-)

Looking at these relationships more closely shows that it consists of two main points. The first point is trust in the operations of the data marketplace itself. This is rather a matter of institutional trust than technological trust. Institutional trust deals with the belief that there are "*impersonal structures that enable an entity (company or individual) to act in anticipation of successful future endeavors"* (Ratnasingam et al., 2005, p. 70). Technological trust is the belief that technology will help users to achieve their goals in uncertain situations (Lee & See, 2016). We can derive from the existence of institutional trust that it is not necessarily the technology that the marketplace participants distrust. It can also be the social aspects of data marketplaces, such as its reputation. This is either between a data provider and a data buyer or between marketplace participants and the data marketplace itself. For the technological trust, is it the knowledge that is lacking about the technology used which leads to a lack of trust. The latter point can be stated to be a relationship that can be improved upon, as education on these topics would lead to more knowledge. The how and what of the educational methods to achieve this could be part of future research. It should then also be researched if increasing knowledge on these topics would indeed increase the trust in data marketplaces.

To explore possible solutions for acquiring institutional trust the article database of Google Scholar was searched on "data marketplace" AND "trust". The results of the query were sorted on relevancy and were scanned in this same order to explore solutions. Articles that noted that a certain technology would take away the obstacle of trust were discarded, as it was earlier concluded that institutional trust and technological trust are different things. The articles we used are depicted in table 5.

Table 5: Solutions mentioned in literature for creating (institutional) trust

	(AUTHOR, YEAR)	TITLE	CITATIONS	SOLUTIONS FOR CREATING TRUST
1.	(Zuiderwijk et al., 2014)	**"Elements for the Development of an Open Data Marketplace"**	16	• Peer-to-peer public dialogue among participants • A critical mass of users • Clearly show where data is coming from and how it was created
2.	(Shakeri et al., 2019)	**"Modeling and matching digital data marketplace policies"**	4	• Regulations and policies on who has access to what
3.	(J. S. Park et al., 2018)	**"Smart contract-based review system for an IoT data marketplace"**	46	• Reputation system or review system: Complete integrity of all reviews to every user of the system
4.	(Täuscher & Laudien, 2018)	**"Understanding platform business models: A mixed methods study of marketplaces"**	272	• Review system • Active community management • Curation of listings • Conflict resolution program • Money-back guarantee

5.	(Chowdhury et al., 2019)	**"Trust Modeling for Blockchain-based Wearable Data Market"**	1	• Direct trust: first-hand experience and evidence • Indirect trust: referrals from one or more intermediate third parties

The authors of these articles mention several solutions for creating trust. For example, a data marketplace could stimulate peer-to-peer public dialogue among participants and clearly show how data was created and where it came from (Zuiderwijk et al., 2014). It should be transparent in the regulation and policies that are employed on who has access to what data (Shakeri et al., 2019). It should make use of reputation or review systems (J. S. Park et al., 2018; Täuscher & Laudien, 2018). There should be complete integrity of these reviews meaning all reviews (positive and negative) should be included and have the same accessibility (J. S. Park et al., 2018). Data marketplaces could also employ active community management, a conflict resolution program, and a money-back guarantee (Täuscher & Laudien, 2018). Direct trust can be established through first-hand experience and evidence (Chowdhury et al., 2019). Indirect trust can be established by getting referrals from one or more intermediate parties. Programs can be set up that incentivize participants to make referrals to their direct contacts.

We can conclude that there are some means to improve the trust relationships between either marketplace participants or between marketplace operator and participants. However, in the context of trust, every e-commerce platform is dealing with this obstacle (Jøsang et al., 2007). This is a consequence of the online nature of such a platform. Because providers and buyers do not directly meet when making an exchange of any goods, the trust information is limited. The same applies to a data marketplace. Therefore, it seems that the relationship between the obstacle and a business model cannot completely be solved. However, the trust relationship can be improved upon by implementing some of the solutions mentioned above. In general, we believed that establishing trust in data marketplaces is of utter importance. This is because trust is necessary when the expected outcomes cannot be guaranteed (Ratnasingam et al., 2005). For data marketplaces guaranteeing these outcomes is very difficult, as we tried to describe in several ways. Therefore, the trusted party, the data marketplace, should be perceived as a party that behaves appropriately, even without enforcement.

8.1.2 Reflection on privacy and control

In the obstacle category of privacy and control there seem to be red and orange relationships between the business model choices and the obstacles.

Privacy and control (-)	BM choice
R1. Leaks to 3rd parties: how to keep control in case of fraud	B2B
R2. Duplicate systems of personal data storage	*Value proposition* / C2B / Health & personal data
R3. Assumption: privacy and control will be no priority for users	C2B
R4. Little control because personal data is aggregated by big-techs	Health & personal data
R5. Personal data has a higher sensitivity than other data	Health & personal data

Table 4.b: Privacy and control relationships (-)

The red relationships are specifically targeted at data marketplaces that aim to sell health & personal data. Their competitors can be stated to be big-techs that have great power over the personal data industry, such as Facebook, Google, and Instagram. These social media platforms also collect and monetize personal data, but often without the greatest part of its users' knowing or understanding (Nyoni & Velempini, 2018). To overcome this obstacle individuals thus need to be empowered to reclaim control on the data that is collected by these platforms. If individuals have control over their data, they can choose what they want to do with it. The question remains if and how this can be achieved. To explore possible solutions for achieving this the article database of Google Scholar was searched on "personal data" AND ("privacy" OR ''control"). The articles that resulted from the query were sorted on relevancy and their titles and abstracts were scanned in this same order to explore solutions. The outcome is summarized in table 6.

Table 6: Solutions mentioned in literature for establishing individual privacy and control

	(AUTHOR,YEAR)	TITLE	CITATIONS	SOLUTIONS FOR INDIVIDUAL PRIVACY AND CONTROL
1.	(Nyoni & Velempini, 2018)	**"Privacy and user awareness on Facebook"**	22	• Raise user's privacy awareness • Privacy settings should be simplified and given more emphasis • Laws that protect user's data and be enforced by regulators
2.	(van Ooijen & Vrabec, 2019)	**"Does the GDPR Enhance Consumers' Control over Personal Data? An Analysis from a Behavioural Perspective"**	56	• Future development of the implementation of the GDPR
3.	(Lazaro & Le Métayer, 2015)	**"The control over personal data: True remedy or fairy tale?"**	33	• Reconceptualization of control: collaborative effort in a socio-technical environment (no individual agency) • Consider that not all individuals have the same capacities
4.	(Malgieri & Custers, 2018)	**"Pricing privacy – the right to know the value of your personal data"**	68	• Increase awareness by introducing a right to know the value of your personal data
5.	(Isaak & Hanna, 2018)	**"User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection"**	277	• Legislation that includes four main principles: public transparency, disclosure for users, control, and notification

From the first exploration of this relationship, it can be concluded that two areas can change the nature of the relationship: legislation and user understanding and awareness. Concerning awareness, privacy settings should be simplified and given more emphasis (Nyoni & Velempini, 2018) and a right can be introduced that allows individuals to know the value of their personal data (Malgieri & Custers, 2018). It should be considered that achieving this is a collaborative effort. No agency can achieve this individually (Lazaro & Le Métayer, 2015). Also, it should be considered that not all individuals have the same capacities (Lazaro & Le Métayer, 2015). This means that different measures will be needed for different groups of people. There should be laws that protect the user's data and they should be enforced by regulators (Nyoni & Velempini, 2018). The legislation should be based on four main principles: public transparency, disclosure for users, control, and notification (Isaak & Hanna, 2018). Also, further development and implementation of the GDPR can help (van Ooijen & Vrabec, 2019). Legislation on personal data aggregation could have the effect that platforms such as Facebook and Instagram have to change their revenue models. It has already been suggested that an alternative could be implemented subscription fees for using such social media platforms (Heisen, 2021). This would lead to a shift in power to the end consumer. The end consumer should however then still be aware of and understand the different options that relate to privacy and control of their personal data.

The orange relationships that state that the possibility of fraud and personal data is part of the privacy and control obstacle is not easy to solve. Fraud, unfortunately, happens in every form of online and offline commerce, and it is unlikely that this will be different for a data marketplace. However, technological solutions and international controls will partly prevent fraud from happening. Personal data will also always be more sensitive than other categories of data since a great part of personal data will not go obsolete. While business data from 10 years ago might not be useful anymore, it is less likely that your name or e-mail address has changed. Therefore, it stays subject to misuse.

The assumption that privacy and control will be no priority for users is supported by the so-called 'privacy paradox'. This is "*the relationship between individual's intentions to disclose personal information and their actual personal disclosure behaviors"* (Norberg, Horne & Horne, 2007, p. 1). This means that even though consumers may say that they are concerned about their privacy, their actual behavior does not reflect this. This relationship will thus probably differ per user (group).

8.1.3 Reflection on matching

In the obstacle category of data providers and data sellers, all relationships relate to the number of users and their adoption speed of a data marketplace.

Matching (-)	BM choice
R1. Only infrastructure for exchange: availability of data in control of providers	Customer provided data
R2. Both market sides cannot be attracted at once (chicken-and-egg problem)	C2B/B2B
R3. Matching is done manually since there are not enough users to automate it	*Value proposition*

Table 4.c: Matching data providers & buyers relationships (-)

The relationships can change, as all will become less problematic when the amount of marketplace participants increases. For example, the first relationship notes that when a data marketplace is independent of matching mechanisms, but only provides the infrastructure for exchange, the availability of data is only in control of the providers. When there are no providers, no buyers can be matched to them. This is contrary to the marketplace where a matching mechanism is employed and providers are actively involved by the data marketplace operator. When the data providers are involved, the data buyers will be attracted to the marketplace and vice versa. This is of course also related to the concept of network effects. When the number of marketplace participants gets larger, matching will get easier as the odds that a dataset is offered or needed are higher. However, this will also have consequences for pursuing a manual matching approach in a business model. Manual matching is not necessarily an obstacle with few amount of users. However, it can be assumed that this will not be viable anymore when the amount of users gets too large. In conclusion, matching at this moment is mainly perceived as an obstacle because of the limited amount of data marketplace participants. Therefore, strategies should be employed to attract providers and buyers. Only after this has been done, does it make sense to look into the most efficient matching mechanisms.

8.1.4 Reflection on pricing mechanism

In the obstacle category of the pricing mechanism, the identified relationships can be redefined to two main points.

Pricing mechanism (-)	BM choice
R1. Unstable coin value	Crytocurrency
R2. Not enough marketplace deals to automate pricing	Price set by marketplace providers
R3. It is unknown what variables should be included to calculate the price	Price set by external sellers/negotiation
R4. Prices are set by trial-and-error	Price set by buyers
R5. Users (providers) have no insight on what their data is worth	Price set by buyers

Table 4.d: Pricing mechanism relationships (-)

The first point is that, as also already established from literature, it is very difficult to give data a price tag because of its unique characteristics (in comparison to physical products). Whether the business model of a data marketplace lets the marketplace providers, the buyers, or external sellers set the price, they all deal with the difficulty of estimating what variables are important when calculating the price of a dataset. There are not enough marketplace deals yet to automate pricing by machine learning algorithms based on prior sales data. The variables that are important for price calculation can thus not be derived from this. Prices being set by trial-and-error is an approach that could give insight on this, but this will not lead to efficient transactions in the long run. When buyers set the price, this aspect should even be given some ethical consideration. This is because the data providers, users in the case of a personal data marketplace, have no insight into how much their data is worth. As they initially are the data owners it could be perceived as misleading, and therefore ethically challenging, to have them sell their data when there is no uniform pricing standard. To explore whether solutions are being invented to solve the obstacle of data pricing the article database of Google Scholar was searched by using the key “pricing data”. The articles that resulted from the query were sorted on relevancy and their titles and abstracts were scanned in this same order to explore solutions. Abstracts or titles that implied that similar solutions were proposed were skipped. The outcome is summarized in table 7.

Table 7: Solutions mentioned in literature for pricing data

	(AUTHOR, YEAR)	TITLE	CITATIONS	SOLUTIONS FOR PRICING DATA
1.	(Azcoitia & Laoutaris, 2020)	**“Try Before You Buy: A practical data purchasing algorithm for real-world data marketplaces”**	-	• ‘Try Before You Buy’ algorithm – pricing is based on the accuracy of the dataset for the buyer
2.	(Liang et al., 2018)	**“A Survey on Big Data Market: Pricing, Trading and Protection”**	137	• Data pricing models: certain market structures fit certain pricing models with particular algorithms/schemes, goals, and approaches
3.	(Stahl & Vossen, 2016)	**“Data Quality Scores for Pricing on Data Marketplaces”**	14	• Data Quality Scores: Relate data pricing to quantifiable data quality criteria: content-related,

				technical, intellectual, instantiation-related. (Novelty to be considered in future research)
4.	(Shen et al., 2019)	**"Pricing Personal Data Based on Information Entropy"**	3	• Pricing functions: proposal for a personal data pricing function based on information entropy
5.	(Chen et al., 2019)	**"Towards model-based pricing for machine learning in a data marketplace"**	19	• Model-based pricing: instead of selling raw data, sell ML model instances. The price is dependent on the accuracy of the model instead of the dataset.

The conclusion that can be derived from this literature search is that a lot of research is being done in the field of data pricing (mechanisms). This ranges from (algorithmic) data pricing models (Azcoitia & Laoutaris, 2020; Liang et al., 2018) or functions (Shen et al., 2019) to scoring frameworks based on data quality (Stahl & Vossen, 2016). Research even suggests it should not be tried to find a solution to price the raw data itself, but rather the machine learning instances that follow (Chen et al., 2019). The price can then be dependent on the accuracy of the model instead of on the dataset. There is no best practice yet in terms of data pricing, presumably also because of the early stage in the development of data marketplaces. It will probably also be very difficult to select a best practice, but the broad range of solutions implies that the relationship between this obstacle and the related business model can be improved upon.

The second point might be an issue on a smaller scale but is however still important to consider for future development. The BM choice that is related to this is using cryptocurrency as payment currency. It was found that own cryptocurrencies in data marketplaces can be used to fund the project by an ICO (initial coin offering). This practically means that the revenue from the token sale is re-invested in the data marketplace. However, creating an own token also means that the value of this coin is unstable. If prices are set by the seller, the seller will not know upfront how much its data will be worth in fiat, as the variation in crypto value might be great. It can be assumed that this is not an attractive perspective for the data sellers. If the prices are set by the buyer, it can be assumed that they will make use of these fluctuations. This is also the case if a data marketplace does not create its own token, but uses other established cryptocurrencies. For example, coins such as Ether or Bitcoin are also still heavily volatile (McCoy & Rahimi, 2020).

Although the exact influences of using cryptocurrency on the expected use or adoption of data marketplace cannot be determined, it should be at least be considered that their development seems to be partly intertwined. The relationship however is not necessarily negative as the development of so-called "stablecoins" is also progressing (*Wat Zijn Stablecoins? | Binance Academy*, n.d.). These coins follow the value of fiat currency and therefore discards the risk of great fluctuations in price. In a further stadium of data marketplace development, this might be a possible solution.

8.1.5 Reflection on governance mechanisms

The obstacle category of governance mechanisms was also already touched upon in the privacy and control section, where it appeared that legislation is a necessary pre-condition to achieve individual privacy and control. However, it was suggested during the interviews that regulations should also help stimulate the progress of a data marketplace more broadly.

Governance mechanism (-)	BM choice
R1. Regulations should help data marketplace progress (but is not yet in place)	*Value proposition*
R2. Different regulations for personal data than other categories	Health & personal data
R3. Different regulations across the globe	Global scope
R4. Who has responsibility for data-driven decisions?	B2B

Table 4.e: Governance mechanism relationships (-)

One of the data marketplace operators even said *"I think regulation is going to be our best friend"* (personal communication, May 11, 2021). Although this is a subjective statement, it is not unplausible. However, it should not be forgotten that it could also work the other way around. Upon implementation of the *General Data Protection Regulation (GDPR)* in 2018, it was also questioned whether it would be a burden or a foundation for digitization (Mikkelsen et al., 2017). A promising business might be derailed by regulatory issues. Researching the influence of all regulations on the development process of data marketplaces would be a new research topic. However, it is interesting to give a first indication of what regulations could be looked at based on the insights from this research. During the interviews, three (future) regulations were mentioned: (1) Payment Services Directive (PSD2) (2) Data taxes / digital service taxes (3) Web scraping. The Payment Services Directive (PSD2) obliges banks to give third parties access to the checking accounts of their customers (De Nederlandsche Bank, n.d.). The advantage for a data marketplace is that they can now also access the financial/banking data of their customers and thus use it as a product of exchange on the data marketplace. It is thus an opportunity for innovation. The regulation is currently implemented in Europe. Data taxes are taxes on data collection or selected gross revenue streams of large digital companies (Asen, 2021; Madsbjerg, 2017; Zaretsky, 2021). The assumed advantage for data marketplaces is that this regulation will give big-techs a fiscal reason to not acquire all data, which allows data marketplaces to reclaim some power over these big-techs. The progress of implementing this regulation differs per country. Then, lastly, regulations for web scraping overlap with data marketplace operations. This is because on data marketplaces the data could be exchanged that can also be scraped from the web. This is not illegal, but scraped data cannot be used for unlimited commercial purposes (Paul, 2020; Waterman, 2020). We can therefore assume that the data licenses that data marketplaces would work with will provide an advantage over compliance with web scraping regulations. We summarized the implications of these regulations in table 8.

Table 8: Implications for data marketplaces concerning data regulations (on) PSD2, data taxes, and web scraping

	REGULATION	SOURCE	SCOPE	DATA REGULATION IMPLICATION
1.	Payment Services Directive (PSD2)	(De Nederlandsche Bank, n.d.)	EU – implemented	• Banks are obliged to give third parties access to the checking accounts of their customers (with their permission) • Opportunities for innovations: smart household books that make predictions based on previous payment transactions, devices that keep track of stick, orders, and immediately settles the bill. • Main advantage: access to financial/banking data
2.	Data taxes / digital services taxes	(Zaretsky, 2021) (Asen, 2021) (Madsbjerg, 2017)	Scattered (scope and status differs per country)	• Taxes on data collection or selected gross revenue streams of large digital companies • Assumed advantage: a fiscal reason for big-techs to not acquire all data
3.	Web scraping	(Waterman, 2020) (Paul, 2020)	Global - implemented	• Web scraping is not illegal when following compliance rules • The data however cannot be used for unlimited commercial purposes • Assumption: data licenses of data marketplaces provide an advantage over web scraping compliance

The implications only reflect assumptions on the opportunities or advantages of the regulations. Even these assumptions can be questioned. For example, implementing taxes on data collection might theoretically seem valuable for (personal) data marketplace business models, as it gives them space to handle the data that big-techs will not anymore because it costs them money. However, if the monetary value of data is already difficult to assess, how will you know how to tax it correctly? And as Zaretsky (2021) puts it: "*If Google makes a dollar by selling my data and pays 10 percent tax on that transaction, it is still better off than had it not collected and sold my data at all"* (Zaretsky, 2021). In other words: would it matter? Also, there is no consensus in the EU on a tax regime, and multiple countries are setting up their own tax regime. Considering globalization, especially digitally, having multiple standards may lead to confusion and double taxation.

The final governance relationship that was indicated is the choice for a B2B business model and the responsibility for data-driven decisions. In other words: who is responsible if something goes wrong? The impact it can have will probably differ per application area, depending on what the sold data will be used for and who will use it. It will presumably be different if the municipality uses the data for a major decision or if the local bakery uses it to improve that bread baking time schedule. The matter perhaps can be solved legally. However, in most circumstances, it remains a moral dilemma. There is not yet much know about the moral perspective on using new technologies (McManus & Rutchick, 2019).

8.1.6 Reflection on other

The 'other' obstacle category shows very broad relationships that are indirectly also related to other obstacle categories.

Other (-)	BM choice
R1. Network effects/big scope are essential to live up to value proposition	*Value proposition*/Global scope
R2. IoT technology not ready yet (5G might help)	Sensor data
R3. Technical interface leads to slow/too little user adoption	Decentralized platform architecture
R4. Token sale: buyers of the token were are not the users of the data marketplace	Cryptocurrency
R5. Big-techs and data aggregators are also competing for data	Health & personal data
R6. Big-techs have the power to destroy a personal data marketplace	*Value proposition*

Table 4.f: 'Other' relationships (-)

One of the most important ones is the notion of network effects. An increased number of marketplace participants would improve the value of the service the marketplace offers. Concerning the 'matching' more valuable matches can be made, and concerning 'trust' a track record and reputation can be built. Network effects result from the adoption of the data marketplace by its participants and for this reason, it seems valuable to further study their perspective. This would result in requirements that a data marketplace should possess to be adopted by the crowd. The relationship is important to safeguard as after a critical inflection point is reached the growth of a marketplace follows an exponential trajectory (Hagiu & Rothman, 2016). However, a critical side note on this matter is that flaws in a business model are amplified when growing too early or quickly. Hagiu & Rothman (2016) state that before scaling, and thus profiting from the network effects the value proposition between buyers and sellers must first be compelling, trust must be built, incentives to keep participants involved must be created, and regulators need to be involved after the value proposition is clear. It thus seems important to find the right balance between not being dependent on network effects only to overcome certain obstacles, and profiting from the network effects to live up to the full value proposition a data marketplace proposes.

Although the statement that IoT technology is not ready to use on a data marketplace might be biased, it has more often been stated that the development of a 5G network will provide several opportunities for IoT applications (Agiwal et al., 2019; Goudos et al., 2017; Wang et al., 2018). Therefore, it is likely that with the development of IoT technology the market for sensor data will become more attractive. The relationship is thus dependent on the development of a separate technology (which can be perceived as an additional obstacle).

The relationship between a technological interface and the slow user adoption can be explained by referring to the technology acceptance model (Venkatesh & Bala, 2008). Apart from the usefulness, high perceived ease of use will lead to faster customer adoption of the technology. Therefore it can be suggested that a simple, customer-friendly interface will lead to faster adoption of a data marketplace.

In the pricing mechanism section, the obstacles of using a cryptocurrency in a data marketplace were already discussed. However, the token sale that results from the choice of developing a cryptocurrency is not directed related to the pricing mechanism. It however is related to the funding of a project. The goal of a token sale is to form a community with the participants on the platform: the ones that buy the tokens should also be users of the data marketplace. In that way, the cycle is closed. If the buyers of the tokens do not participate in the data marketplace, the cycle is not closed and the revenue model is flawed.

Lastly, and partly overlapping with what was discussed earlier, is the competition of the big-techs and possibly data aggregators. It is difficult to define a strategy on how to compete with these big techs. However, the EU aims to revamp the operational practices of these parties by implementing The Digital Services Act and the Digital Markets Act (Keane, 2021). Apart from the regulations that were already mentioned these rules may positively affect the competitive relationship. However, the implementation of these acts is still being discussed by the 27 EU member states (Amaro, 2021). The role of data aggregators is still quite unclear. On the one hand, they have a working business model that also includes selling data. Therefore, they could be seen as competition. On the other hand, they could be mediators between the data providers and the data marketplace. An example of this are small local medicine shops that might be too small to collect to the marketplace themselves, but the collated information would be of appropriate size (Banerjee & Ruj, 2018). Data aggregators can also be used in the first phase of a data marketplace start-up to get the first marketplace buyers involved, as is VETRI's strategy (personal communication, May 5, 2021).

8.2 Reflection on the relationships: an expert opinion perspective

To reflect on the relationships between the obstacles and business models from another perspective than the data marketplace operator we asked experts on data marketplaces for their opinions. We interviewed three individuals. We considered these individuals to be experts as they are involved in the TRUSTS project and thus are knowledgeable on and experienced with the topic of data marketplaces. The questions were not related to the specific business model choices of the case studies, but rather to their ideas behind the identified relationships in each obstacle category. The relationships above are based on the experience of data marketplace operators. As a theory-building process should also include conflicting literature, it is valuable to also question these relationships in this way. This would validate or dispute whether those relationships are indeed the (only) main points to consider in each obstacle category. In this way, we can critically assess the identified relationships before including them in the explanatory model. We structured the interview protocol by discussing the relationships in each obstacle category (trust, privacy and control, matching, pricing mechanism, governance mechanism, and other) The questions that we asked the data marketplace experts can be found in Appendix E. The answers to the questions were analyzed by transcribing the interview and marking the remarks that we think can form important input for the explanatory model. These remarks are described below.

For **trust,** the case study research led to an insight into the difference between technological and institutional trust. The data marketplace experts argue that institutional trust is not of concern. The issue of trust mainly arises between data providers and data buyers and does not involve the marketplace itself. It was argued that trust-building relationships take months or even years, while data marketplaces are meant for relatively quick data exchanges. That is why their proposed solution is an environment that allows for trust-less interaction. No trust would be needed because of distributed computing solutions. The downside of this approach is that technological solutions are often hard to implement for SMEs and start-ups due to the technical complexity. For personal data marketplaces, it was also noted that people would trust the use of these technologies without understanding because of marketing and communication. It was assumed that the trust issue might even be less, as the user primarily sees it from the benefit perspective (just as with Facebook). However, not all technological solutions data marketplaces implement necessarily relate to creating trust. For example, a smart contract might be useful for automatic payment triggers but does not create transparency. It was even argued that they would do the opposite, as coded contracts would be even harder to understand than written contracts.

For **privacy and control,** all relationships were mostly agreed upon by the data marketplace experts. It was stressed that the issue of control does not only involve leaks to third parties but rather keeping control over the context of data usage that was agreed upon. This includes the scope of usage and how it is going to be shared and used onwards. Because there is no enforceable IP on data it cannot be controlled how it is going to be used. This notion has an overlap with how the obstacle of trust was described. A suggested solution was the combination of a trust-less environment with selling pre-determined insights from the data on the data marketplace instead of raw data. By doing this, the selling party can force a certain degree of control on the data usage of the buying party. This would only be possible if the algorithms that can process this data are available. For control regarding personal data marketplaces specifically, and thus their competition with the big-techs who are also after personal data, the red relationships from the BM Stress Test can be validated. This is mainly because the value of behavioral data and usage data is in vast datasets and the developed algorithms to dig through them. The Big Tech (Facebook, Google, Amazon, Apple) are way ahead on any personal data marketplace that starts from scratch without any user base of an existing platform. Therefore it was suggested to reconsider and redefine what a

personal data marketplace is and what kind of data they sell.

For **matching** the main insight from the case study research was that there is a lack of data marketplace participants to create an efficient matching mechanism. According to the data marketplace experts, one of the core problems is that companies are fearing to share their data. Three main reasons for this were mentioned. The first reason is that companies tend to be overcautious. By selling data they could lose a competitive advantage. Even if they don't know what their data could be valuable for, and do not see what kind of advantage they would be giving away they would rather not share because the perceived risk is greater than the benefit. This is a principal-agent problem. The decisions are being made internally in the company and when someone decides on selling the data this comes with great responsibility. It was suggested that there is a need for distinguishment in the type of data being sold. The perceived risk might differ if the data is strategic or operational. A starting point could be a checklist for data providers to identify the sweet spot in generic data that could be exchanged. The second reason was described as 'the fear of missing out'. This means that it is not necessarily about the competitive advantage that is lost, but rather about the jealously of somebody creating value while the data seller is then no longer allowed or able to participate in that value creation. They would not have been able to find that value themselves, but also do not feel like the other party deserves it. The final mentioned reason was more practical: the IT architecture differs per company. SME's or start-ups that have been around for 10 years or less do not have sophisticated IT systems, while large companies that have been around for longer often have legacy IT systems. This means that, while data is described as being "the new oil", it is quite an effort to prepare it for sale. Solutions could be internal data stewards or data catered teams who would create data pipelines. However, the question remains if the necessary management attention and qualified resources in combination with the overcaution and fear of losing out lead to a good business case. From the described perspective it is not attractive for companies to participate in a data marketplace. Nonetheless, it was also suggested that there are other opportunities. For example, a data marketplace operator could act as a soft consultant that guides a companies' internal data governance management and helps them to create their own team. In that way, it would be easier and more attractive for companies to conform to the requirements of participating in a data marketplace as a data provider.

For the **pricing mechanism** obstacle, the data marketplace experts acknowledged the identified relationships. Data is very context-specific and you do not know exactly what you are pricing, what you are buying, and how the value trajectory will develop. There is no direct solution for this. However, it was stated that data pricing currently is a big topic and that literature on this phenomenon has dominated the study. This was also seen when identifying the solutions for data pricing. It was also noted that the difference in pricing mechanism for physical products and data also differs in the number of people that are interested in buying it. The amount of people that buy data is not that big, meaning that there also is no general experience on what will be accepted. For a small number of buyers, the pricing mechanism is closer to a contractual negotiation. This is what was also seen in one of the case studies. Manual matches, including price negotiation, seem suitable on a small scale. However, this is not an option if data marketplaces aim to become as successful as other electronic commerce platforms. For totally new fields it will be a matter of trial and error. For established fields, there is some more knowledge on how it would be. Regarding the payment currency, either fiat or crypto, the main relationship that using a cryptocurrency endangers your (revenue) business model was acknowledged. This is mainly because cryptocurrencies are part of a different ecosystem that cannot be controlled. However, at this stage in the development of data marketplaces, it is not the core problem that needs to be solved. Therefore, it is not urgent to identify possible solutions.

For **governance mechanisms,** it was asked if data marketplaces needed additional regulation to have a viable business model, as was suggested during the case studies. The data marketplaces experts were no experts in rules and regulations but did mention that the main legal discussion is about the liability. Especially as decentralized marketplaces are the trend, it might be difficult to estimate who is legally liable when something goes wrong with the data. Also, a comparison with the implementation of the GDPR was made. The implementation of this regulation resulted in huge confusion in the beginning. It was not clear what it exactly was, when it was applicable, and what it meant for an individual, organization, or business. It was, therefore, suggested that it could be a strategy for a data marketplace operator if they can give the guarantee to a business that they will comply with regulation once they joined the data marketplace. This would be attractive since it might be hard for a business to comply with its own stack of technology. If the business would then join the data marketplace, the risk could be shared.

In the interest of creating a broad view of data marketplaces obstacles that should be overcome the question for the **'other'** category was aimed at identifying more obstacles and solutions from a TRUSTS perspective. The main point that was made was that "plain vanilla" marketplaces do not seem to have a viable business case. Plain vanilla in this case means that a data marketplace would just sell raw data. The raw data does not scale because it is very unique. A single piece of data cannot be used in many different contexts but is only valuable for a certain application in a certain context. It is also not possible to build an algorithm for each separate piece of data. Therefore, it was mentioned that to achieve scalability a stable context should be created. This would be possible by being niche-specific. If an industry has a defined set of industry standards, players, and requirements, the probability that data being offered is useful for another party is higher than on a generic data marketplace. In conclusion, selling consultation and solutions rather than raw data sets was seen as an opportunity that has a more viable business model.

We have now performed the three research elements that can lead us to the final result: an explanatory model on the relationships between the business model of data marketplaces and the obstacles that they should overcome. Initially, this research aimed to find specific relationships between the business model choices (from the taxonomy) and the obstacles. However, after having conducted the Business Model Stress Test with three case studies we realized that this method will not lead us to the intended result. The explanatory models that we will describe in the following chapter will use the business model dimensions as a basis, as we explained in Chapter 7.2.

CH 9. Reaching closure: explanatory model(s)

The ninth chapter reflects the eighth step of the theory-building process from case studies: reaching closure. In this case, we aim to reach closure by creating an explanatory model that illustrates the relationships between the data marketplace obstacles and their business models. Based on the insights from all chapters, including the stress test results and the reflection on the relationships (Chapter 6, 7, and 8), we can make our final judgment and create an explanatory model. Including all findings in one empirical model would not result in a clear model. Therefore, the observations are discussed per marketplace requirement and corresponding obstacles. It should be noted that the relationships in these diagrams are not statistically validated, but do give an initial insight into what appears to be important in practice.

In chapter 3, we described the framework that resulted from explorative theorizing on data marketplace obstacles (see figure 9). In this framework, the obstacle categories were connected to the marketplace requirements they would hinder. Trust and governance were connected to safeness, privacy & control to no repugnance, matching & other to thickness, product description & product quality to provenance information, and pricing to no congestion. To provide structure, the empirical models will now also be discussed per marketplace requirement and corresponding obstacles. To make it easier to understand the empirical models, we have created a legend in figure 24. The business models are modeled as a large orange rectangle with four smaller rectangles in it. These smaller rectangles represent the four domains of a business model according to van de Ven (2020). The business model dimensions, that leave options for different business model choices within this element, that relate to the obstacle(s) are depicted within such a domain. The obstacle categories are visualized as yellow boxes and the external factors as grey boxes. The obstacles can affect the overall viability of a data marketplace business model (see figure 9), and the domain and business model dimension choices can affect the impact of the different obstacle categories and external factors. Arrows between obstacles or external factors represent that we identified an influence of one construct on the other.

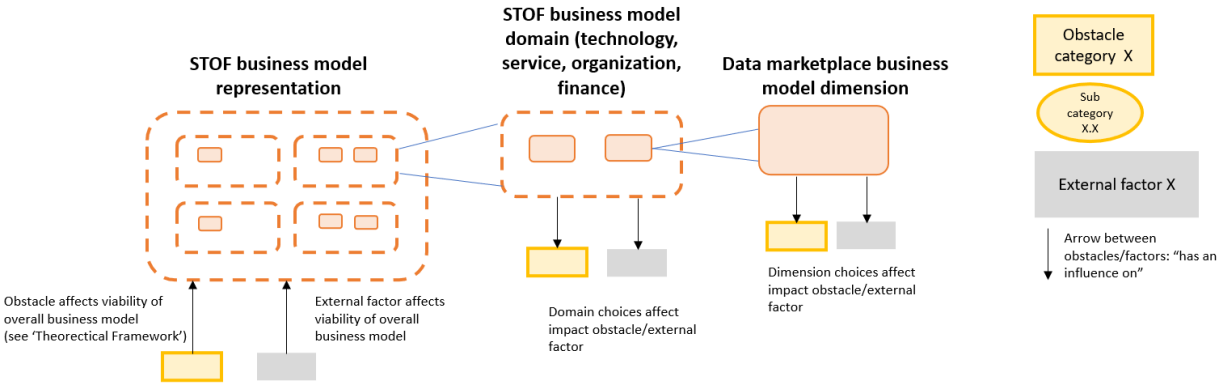

Figure 24: Legend for the empirical models

9.1 Safeness – trust & governance

In the first model, depicted in figure 25, we visualize the relationships between the business model and the trust and governance obstacle categories.

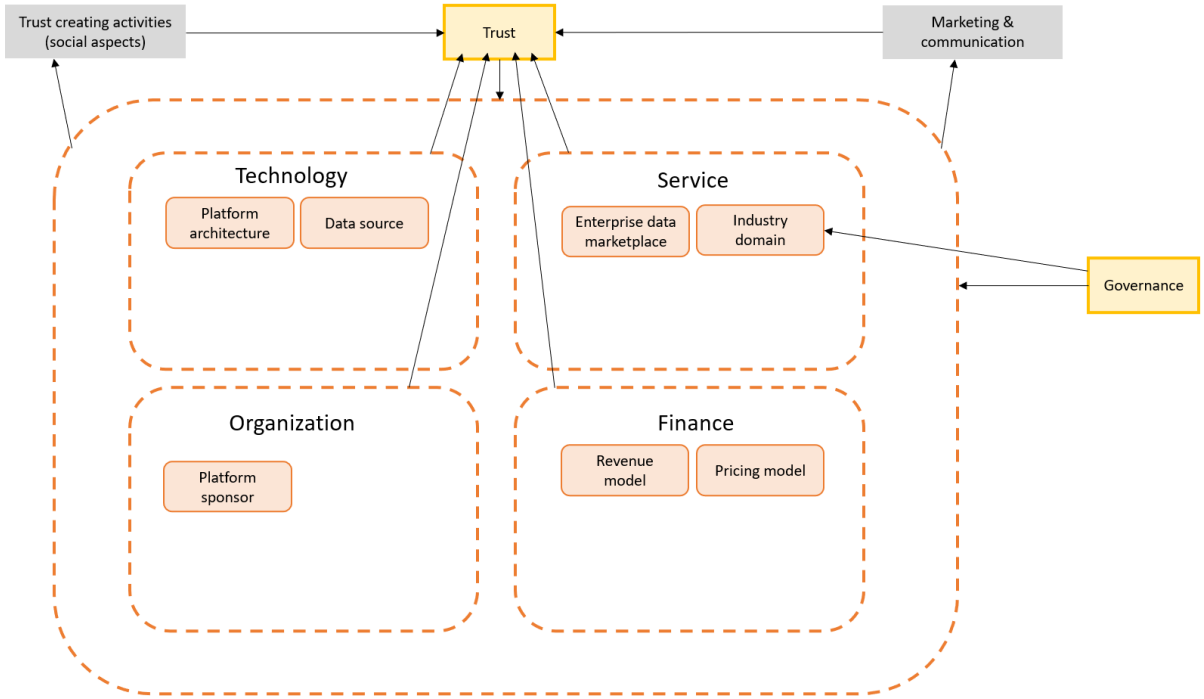

Figure 25: Relationships between the obstacles of trust and governance and the business model of a data marketplace based on the BM Stress Tests and reflection on those results.

To start with **governance**, we saw that this obstacle can affect the overall development of data marketplaces. It is unclear whether this would be positive or negative. On the one hand, governance could lead to more opportunities for the business models (R2, table 3f). On the other hand, it could lead to confusion and difficulties in the implementation and realization process (Ch. 8.1.5) Governance has a specific relationship with the chosen industry domain of a data marketplace. For example, as personal data is more sensitive, regulations are different and often more strict than for business data (R2, Table 4e). However, this also leads to opportunities when a personal data marketplace provides a system that outperforms other options on securing sensitive data (R3, table 4b). For all forms of data, whether it is business, IoT, or personal, the discussion on liability is problematic (R4, table 4e, Ch. 8.2: governance). This does not have a direct consequence for actually selling the data, but does influence what can be done with the data, e.g. innovate. If the options for innovations are limited because of liability issues the data exchange on a data marketplace will probably also be limited.

Trust has a higher amount of relationships with a data marketplace business model. Firstly, the amount of trust perceived influences the viability of a data marketplace business model via the safeness marketplace requirement. In the technology domain platform, the platform architecture (R1, table 3a) and data source (R3, R5, table 3a) seem to matter to achieve trust. That is because there are two kinds of trust relationships in a data marketplace. That between a data provider and a data buyer (R1, table 4a) and that between a data provider/data buyer and a data marketplace owner (R2, table 4a). If the data is provided directly by a customer (the data source) we could assume that its relationship with the buyer is more important (Ch. 8.2: trust). In that case, more effort should be done to improve the trust relationship. Examples are mentioned in Chapter 8.1.1.

The platform architecture of a data marketplace is included as a decentralized architecture would allow for a trust-less data marketplace (R1, table 4a + Ch. 8.2: trust). In the service domain being an enterprise data marketplace could influence trust (R2, table 5a). According to Chowdhury et al. (2019) direct trust can be built through evidence and first-hand experience (Ch. 8.2). Therefore, we see participating in a private marketplace with private use cases as a step in the direction of participating in an 'open' data marketplace because it builds trust through experience.

Industry domain matters because the requirements per domain for trust might differ. For example, one party might view it primarily from a benefits perspective while another might only see risks (R3, table 5b, Ch. 8.2: trust). For the organization domain, it was observed that the platform sponsor in combination with the revenue model could matter (R4, table 3a). This is because the business interest of a data marketplace reflects its true intentions. For example, a private sponsor such as a foundation and a not-for-profit revenue model on data transactions could radiate higher trustworthiness. The revenue should then come from some other sort of asset sales. The pricing model can be related to trust if a data marketplace decides to also adopt a pricing model on the data provider side. If a data provider has to pay a listing fee per data point/set/stream it will presumably be discouraged to offer malicious submissions on the marketplace (R6, table 3a).

Then, two constructs were identified that do not directly relate to the business model choices that can be made according to the taxonomy. The first construct is marketing & communication (Ch. 8.2: trust). The second construct is trust-creating activities, as described in the section about reflection on the trust relationships (Ch. 8.1.1). These two constructs possibly contradict a bit. If more trust needs to be built for users to adopt the data marketplace, trust-creating activities could contribute to this. If technological trust appears to be enough users could adopt the data marketplace even without understanding the technology used. This will be achieved on basis of marketing and communication about the data marketplace. The ease of use of the technology constructs should then be high, as it appeared that a technological interface is not well adopted (R3, table 4f).

9.2 No repugnance – privacy & control

In the second model, depicted in figure 26, we visualize the relationships between the business model and the privacy and control obstacles.

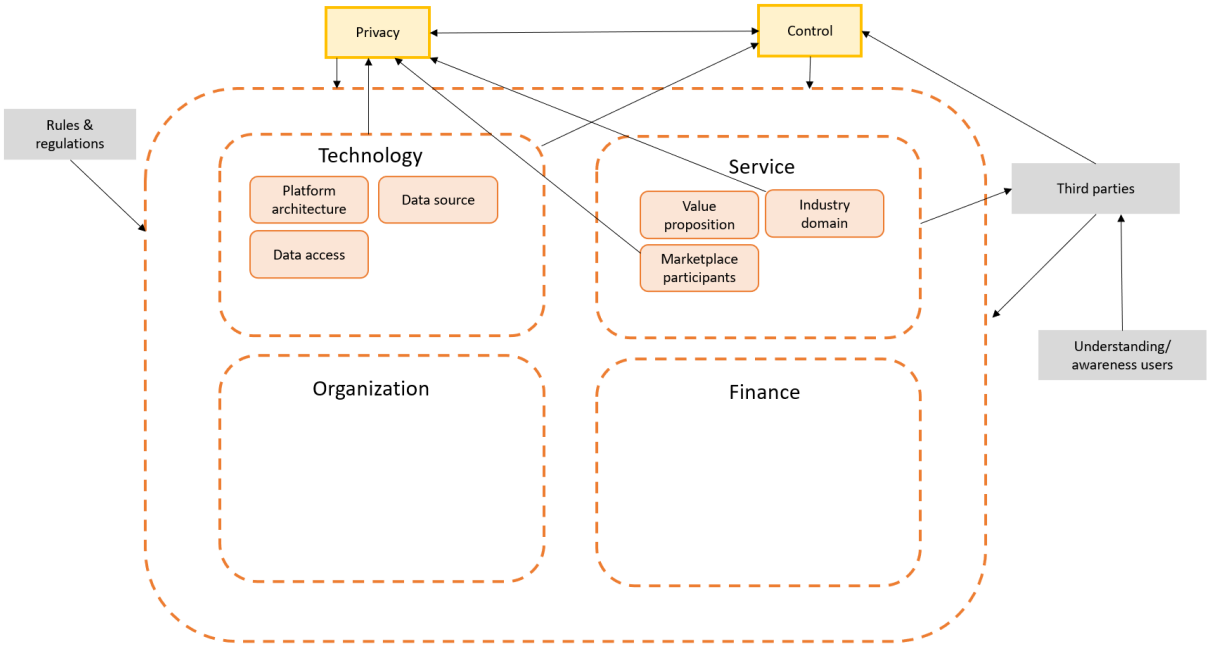

Figure 26: Relationships between the obstacles of privacy and control and the business model choices of a data marketplace based on the BM Stress Tests and reflection on the results.

Privacy and **control** are interrelated concepts. A higher degree of privacy leads to a higher degree of control and vice versa, which is why we placed them in the same obstacle category (Ch.3.5.2). The amount of perceived privacy and control influences the viability of a data marketplace business model through the 'no repugnance' marketplace requirement. The business model of a data marketplace is subjected to rules and regulations on privacy and control matters such as the GDPR. Technology enables a data marketplace to comply with these rules and regulations and can therefore provide privacy and control to a certain degree. For example, a decentralized architecture could enable selling data with limited access to the actual data content (R1, table 3b). Also, specialized software such as the inclusion of a digital identity could prevent fraud and give control (R3, table 3b). However, 'to a certain degree' is explicitly mentioned as sharing data, which is the aim of a data marketplace, is the opposite of keeping data private.

However, privacy might matter less in cases where the data is less sensitive (R4, table 3b). Also, for certain marketplace participants, privacy and control might have no priority (R3, table 4b). They would rather take a benefits perspective. This for example means that making money is prioritized over worrying about privacy. This can be either consciously or unconsciously as a result of the privacy paradox (Ch. 8.1.2) Therefore, industry domain and marketplace participants are of influence on the privacy construct. The data marketplace could also be seen as a mediator. Because they only facilitate the data transactions, they would have no direct privacy responsibilities (R5, table 3b). These responsibilities would be for the data provider and data buyer. However, this does of course not solve the problem. The main troubling relationship does not directly connect the obstacles and the business model choices but involves some external factors. Whether it is for a personal, IoT, or B2B data marketplace, this external factor is a third party that influences the control and the viability of a data marketplace business model. For personal data marketplaces, this third party can be specified to be a big tech (R5/R6, table 4e). Living up to a value proposition that includes C2B

participants and a health & personal data industry domain often includes competition with big techs as competitors who aggregate the same data. Because these big techs have a major head start it will be very difficult to guarantee a user full control over its personal data. It could be improved to some degree by either again rules and regulations or privacy awareness and understanding of the users if that leads to users getting more reluctant to share their data with big techs (Ch. 8.1.2). In the case of a B2B or IoT data marketplace, this third party is every other party that could get hold of the data after it has been sold on the data marketplace (Ch. 8.2: privacy and control). Because there are no intellectual property rights on data, its usage cannot be controlled. The data could be shared forwards and used for other purposes than intended. Therefore, such a data marketplace can also give no guarantee of control. In this case, the understanding and awareness would relate to data licenses that intend to limit the sharing and re-use of data. For us at this moment, the question remains if it would be traceable if these licenses are violated. The third party thus is the main factor that influences the viability of the business model when it comes to the obstacles of privacy and control.

9.3 Provenance information – product quality & product description

In the third model, depicted in figure 27, we visualize the relationships between the business model and the product quality and product description obstacles.

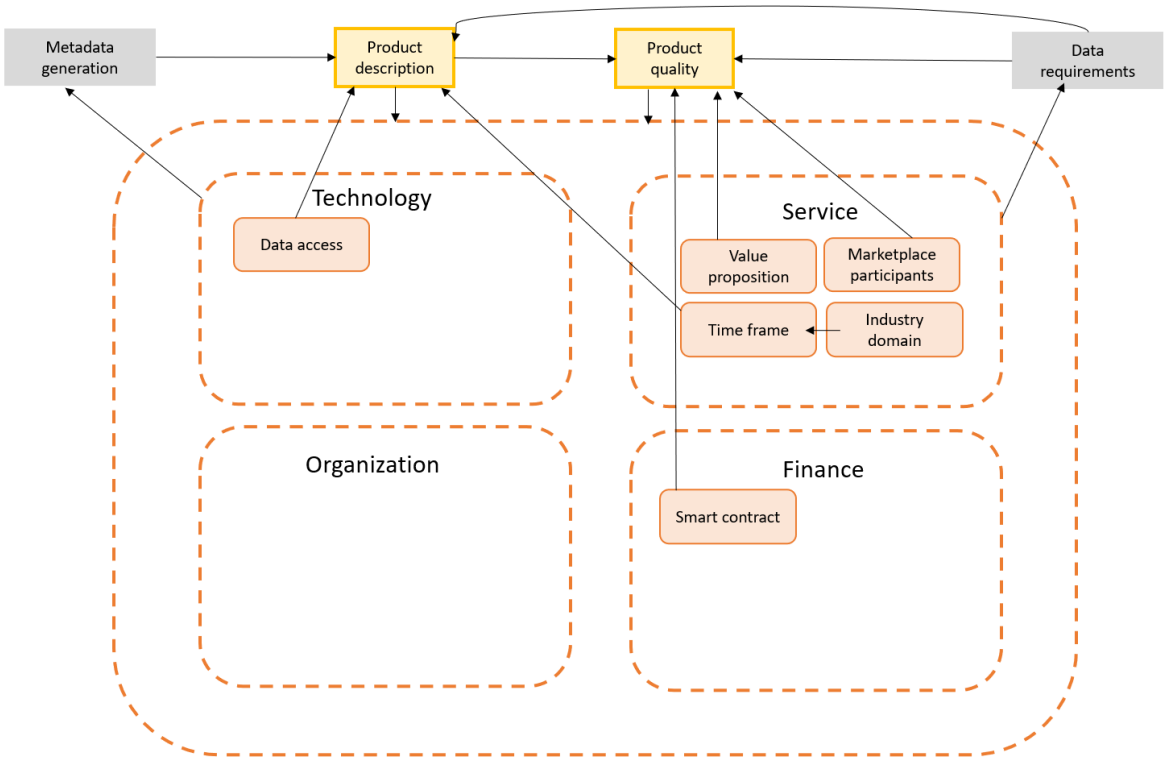

Figure 27: Relationships between the obstacles of product description and product quality and the business model choices of a data marketplace based on the BM Stress Tests and reflection on the results.

The constructs of **product description** and **product quality** are also related. A better product description can lead to better product quality, based on the data requirements of the buyer (Ch. 3.5). Data requirements can also be directly related to product quality as a higher degree of overlap between data requirements and the offered data leads to higher perceived product quality. The combination of the two constructs influences the viability of a data marketplace business model via the 'provenance information' marketplace requirement.

For the service domain of a business model two business model choices directly relate to the **product quality**. Firstly, the value proposition a data marketplace offers can affect product quality. If a data marketplace aims to guide the data buyers in the purchasing process an indication of the quality can be given by implementing a review system (R3, table 3c). Some value propositions of data marketplaces are even primarily focused on offering high-quality data. In the finance domain, the use of a smart contract cannot guarantee the quality of the data but can give a warranty on it (R4, table 3c). This relationship reflects some kind of control over the product quality. Marketplace participants are said to be of influence on the product quality because the kind of data that can be sold differs. For C2B a data profile can be sold, rather than a raw dataset. The C2B structure also allows for the ability to directly ask customers for certain data attributes if they are not yet part of their profile. We assume that this would also be possible for B2B, but this would technically be more challenging as the IT architecture per organization differs (Chapter 8.2: matching). The business model choice for data access can have an influence on the **product description** when a specified data format is used that is connected to a certain form of access. An example would be public B2B data that is all structured using the same entity types and data properties that can all be accessed via a certain API (R5, table 3c). The time frame choice in combination with the industry domain can also matter for the product description as different combinations require a different product description approach (R2, table 3c). For example, we imagine that (near) real-time data that often results from IoT sensor streams can be described by the variables that the IoT device measures. Describing a static strategy dataset of a business of choice requires a different approach.

During the case studies, we did not find any issues with these obstacles. However, we can assume that this also relates to the fact that the data marketplaces operated on a small scale. Controlling product descriptions and product quality is likely to be more difficult on a larger scale when manual controls would be too labor-intensive. Therefore, we also included the external factor of metadata generation identified from literature in the empirical model (Ch. 3.5). The generation of metadata influences the product description as metadata is the *information* about data that is needed to generate *knowledge*. An efficient business model requires an accurate method for generating metadata so that data buyers know what to expect when acquiring the data itself.

9.4 Thickness – matching & other

In the fourth model, depicted in figure 28, we visualize the relationships between the business model and the **matching** and **other** obstacles. The 'other' category consists of *network effects, competitors, technical issues,* and *incentivization*, as defined before performing the BM Stress Test (Ch. 5.2).

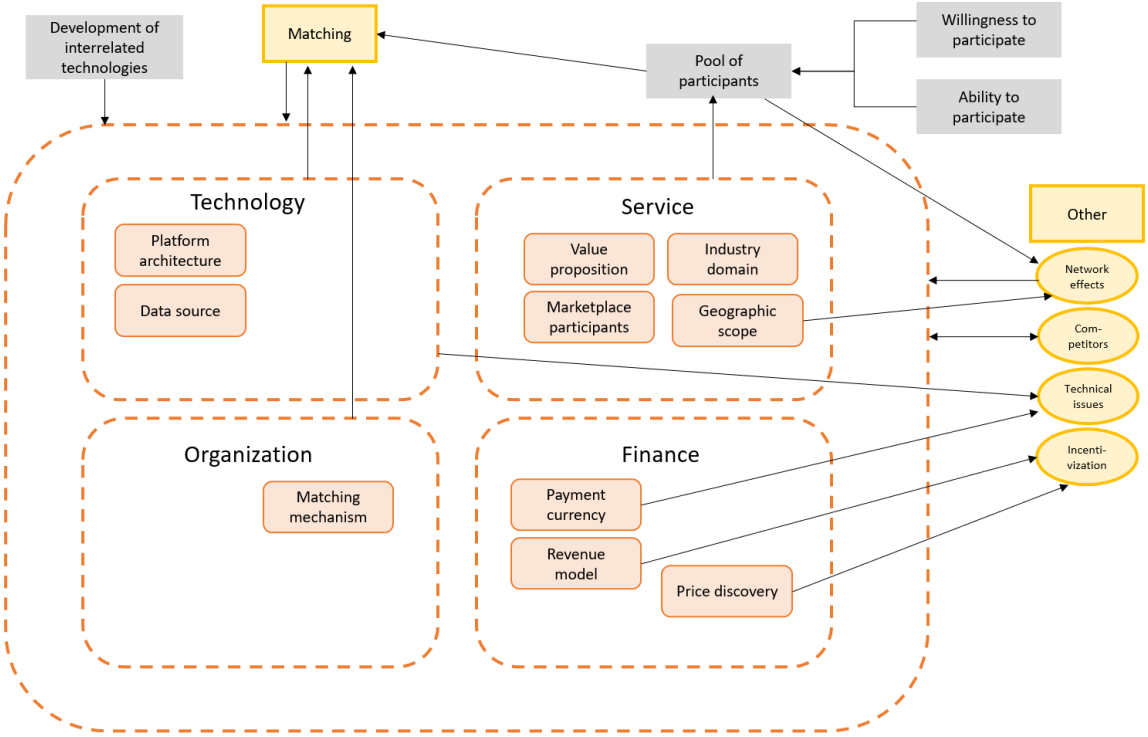

Figure 28: Relationships between the obstacles of matching and 'other' and the business model choices of a data marketplace based on the BM Stress Tests and reflection on the results.

This empirical model includes a lot of relationships. However, two main external factors influence the viability of a data marketplace business model. The first major factor is the pool of participants. A data marketplace is a multisided platform for data trading, which means that the basic needs for it to succeed in doing business are having participants on both sides and "unlocking" the data that the providers possess. In general, we can say that the pool of participants is influenced by the willingness to participate and the ability to participate. We derived this from the findings in Ch. 8.2: matching.

For data providers, this willingness might be low due to overcaution on losing a competitive advantage or the fear of missing out on the value of their data that another party unlocks after buying it (Ch 8.2: matching). For data buyers, the data bought should be perceived as useful and correspond with their requirements for them to be willing to join. The ability is more of a technical issue (Ch 8.2: matching). Both parties should be technically able to manage the datasets that would be exchanged on the marketplace. The IT systems that are in place at the data providers should allow for the creation of data pipelines to the data marketplace. That could be seen as a lot of effort. Data buyers should possess the skills to integrate the acquired data into their analytical tools to create value from the data. This poses a problem if the acquired data cannot be fitted with the analytical tools, which cannot be known upfront. The service domain of business model choices also influences the pool of participants. Different value propositions attract different participants and some industry domains might be more suitable than others. For example, anybody can sell personal data but not everybody has an IoT device whose data streams can be sold (R1, table 3g). Of course, this is not the only consideration as personal data is also considered to be more sensitive. Also, the larger the geographic scope, the more participants can be attracted (R2, table 3g). We argue that the eventual

pool of participants affects how well matching can be performed. On the one hand, a larger pool would increase the odds of finding a match. On the other hand, a small but niche-specific pool would also increase the odds that the data being offered is useful for the other party, thus increasing the quality of the match (Ch 8.2: matching). The size of the pool of requirements of the participants in the pool determines in what way a data marketplace can profit from network effects.

Then, the second external factor we identified is the development of interrelated technologies. Data marketplaces are no independent constructs. They depend on other technologies to develop a viable business model. For example, an IoT data marketplace might depend on the development of 5G technology to offer a good business case (R2, table 4f + Ch. 8.1.6). Other examples are the development of blockchain, cryptocurrencies, and corresponding stablecoins which are technologies that are greatly used in the marketplaces (R3/R4, table 4f).

As for the **matching** obstacle itself that influences the viability of a data marketplace business model, platform architecture, data source, and matching mechanism seem to have a direct influence. A decentralized architecture that includes blockchain technology could enable automatic matching (R1, table 3d). Since matches can only be performed if data is offered on the marketplace, the data source matters for the amount of control a data marketplace owner has over making these matches. For example, if the data is customer provided the data marketplace owner is dependent on them for the number of matches that can be made (R1, table 4c) Also, logically, matching is related to the matching mechanism that is in place.

Then for the **other** obstacles, there were four main categories defined. The pool of participants and geographic scope both influence the network effects that can be created (R1, table 4f, Ch. 8.1.6). These effects will partly determine the value of the business model. Competitors also influence the viability of the business model. These could be either data aggregators or big-techs, that as discussed from a problem for personal data marketplaces (R5/R6, table 4f). We can assume data marketplaces in return could also affect their business models, although this is dependent on the scale on which they operate and the market power they then possess. The technology domain should deal with the technical issues that are encountered (R3, table 4f). Some technical issues however also relate to the finance domain when it comes to the business model choice regarding payment currencies and the corresponding transaction process (R4, table 4f). Incentivization is mainly related to the revenue model and price discovery choices of a data marketplace. To ensure the participation of data providers they should see the benefits rather than the risks (Ch 8.2: matching). As benefits are often perceived in monetary value, the finance domain of the business model should reflect this (e.g. revenue model and price discovery).

9.5 No congestion – pricing

In the fifth model, depicted in figure 29, we visualize the relationships between business model choices and the **pricing** obstacle.

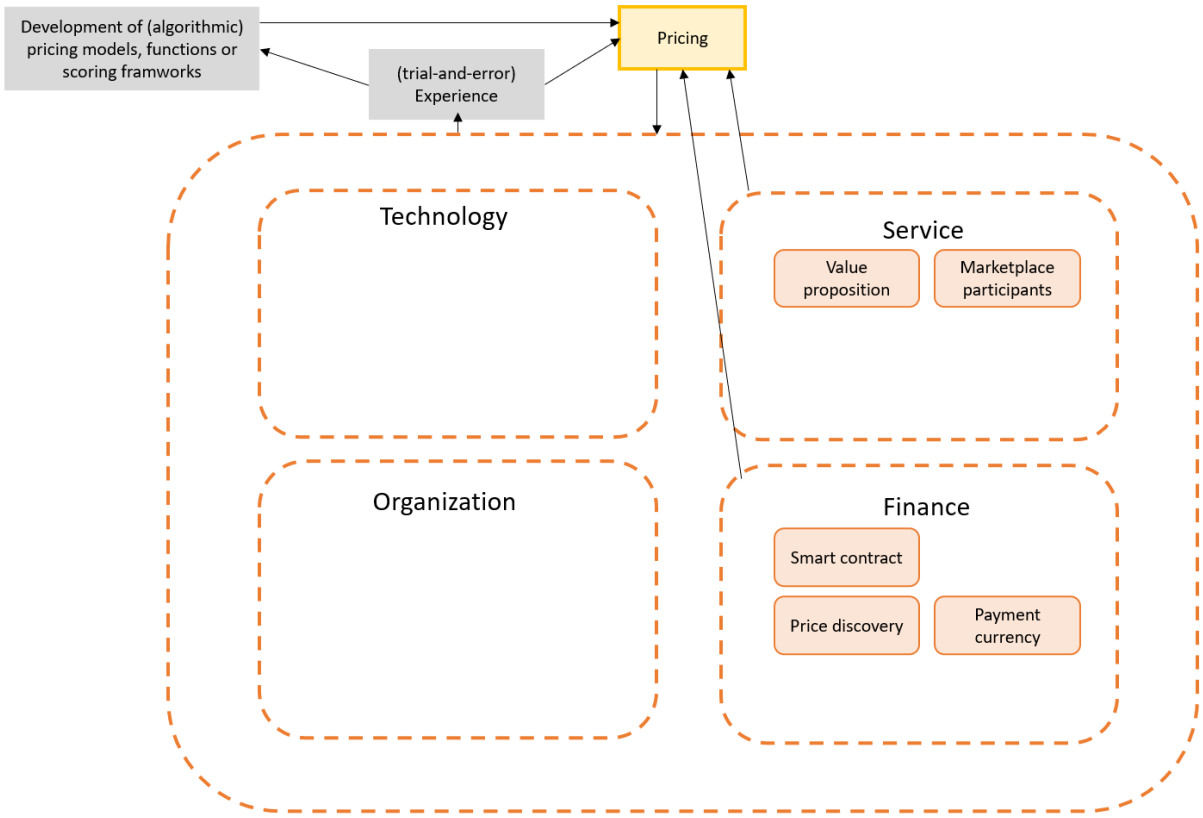

Figure 29: Relationships between the obstacle of pricing and the business model choices of a data marketplace based on the BM Stress Tests and the reflection on the results.

Firstly, it should be noted again that dealing with the pricing obstacle is broader than making "the right" business model choices. The context-dependency, unique characteristics, and unpredictable value trajectory make it hard to price (Ch. 3.5). Currently, a lot of academic research is being done on data pricing, which eventually might help data marketplaces to find a solution to overcome this obstacle (Ch. 8.1.4). Of course, the experience of data marketplaces can also contribute to this. Pricing can also be based on experience, and adjustment of the prices according to demand and supply (Ch 8.2: pricing). Although there is no perfect solution for the problem of data there are some business model choices that can be of influence. For the service domain, the value proposition and the marketplace participants could matter. Some marketplace participants would rather care about receiving any kind of compensation for their data, than receiving no compensation at all (R1, table 3d). This is then not necessarily related to the correct value of the data based on its characteristics. This is mainly applicable to a personal data marketplace, where there is no alternative offered to give away your data in return for the use of online platform services (R1, table 3g). The effort should however weigh up to the compensation received. For example, Barber (2018) describes how he signed up to several data marketplaces for two weeks and made approximately 0.3 cents. It can also be argued that the value proposition of a data marketplace matters for the establishment of a pricing mechanism. Providing only the infrastructure for raw data exchange would require different pricing than offering a form of consultancy on the data being exchanged (Ch 8.2: other). For the finance domain, the mode of price discovery, the payment currency, and the implementation of a smart contract are related to the pricing obstacle. Since there is no uniform standard for data pricing the

choice for price discovery seems like an advantage. If the data buyer sets the price and the data provider is unaware of the value of its data the monetary compensation might be too low (R5, table 4d). If the data provider sets the price the data buyer might pay too much for what the data is worth in the context of the application (R3, table 4d). The latter results from the issue that a dataset cannot be tried or shown before buying it, unlike with most physical products (the so-called Arrow paradox). The reflection on this issue did show a possible solution, the so-called 'Try before you buy' algorithm ((Azcoitia & Laoutaris, 2020) (Ch. 8.1.4). The choice of payment currency relates to pricing because choosing cryptocurrency seems to add additional instability as the crypto market is very volatile (Ch. 8.1.4) Fiat currency does not have this problem, but making this choice also means that there is no opportunity for project funding from a token sale (R5, table 3g). Lastly, implementing smart contracts makes automatic micropayments possible (Ch. 6.2.3: Databroker). This does however not solve how much these payments should be.

CH10. Discussion

This thesis aimed to create an explanatory model on the relationships between business model choices and data marketplace obstacles to assess the viability of data marketplaces. Firstly, explorative theorizing led us to a model on how we can see the relationships between a data marketplace business model, data marketplace obstacles, and the marketplace requirements for efficient market operation. The outcome of living up to these requirements would be a viable data marketplace business model. This model formed the basis for the empirical part of this research, in which we performed three BM Stress Tests on three different marketplaces. By taking this exploratory approach we hoped to identify several links between business model choices of data marketplaces and obstacles for their viability. By reflecting upon these identified links we aimed to create an explanatory model. This chapter aims to discuss and reflect on the research process and the results of the study. We will use the sub-questions of this research for structuring the reflection. This will help us to assess what we can and cannot conclude about the relationships between business model choices and obstacles based on this research. We will first give a structured overview of the answers to the sub-questions we formulated. Based on what is lacking about these answers, we will discuss the challenges and limitations of the research approach we chose to conduct to achieve the research objective. Both the answers on the sub-questions and the discussion on the challenges and limitations will lead us to the practical implications and recommendations, or in other words: how the results of this research can be valuable for increasing the viability of data marketplaces from an obstacle perspective. Subsequently, we discuss the scientific contributions of this research. The chapter ends with some recommendations for further research directions.

10.1 Challenges and limitations by sub-question

As we said, we started off the research by explorative theorizing to form a model that could act as our basis for the empirical part of the research (**Chapter 3**). This is where we encountered the first challenge. When creating the model we aimed to represent its constructs and the relationships between the constructs as accurately as possible. For the sake of simplicity, we chose to model the obstacles to be directly related to one marketplace requirement. However, in reality, certain requirements and obstacles might overlap. For example, control was modeled to be of influence on the no repugnance marketplace requirement because it has a direct relationship with privacy. However, it can also be argued that control influence the safeness requirement as not having control over the data allows others to misuse it. This is also true for the construct of 'harm of own business interests' that was initially scaled under trust. Throughout the research, it became more clear that this is not necessarily an issue of trust. One market side might trust either the data marketplace owner or the other side, but might still not be willing to participate because they do not want to risk it. The choices to model the obstacles under certain requirements were based on logical argumentation that seemed most applicable in that stage of the research. It can however be influenced by the subjectivity of the researcher. We do not believe that this challenge influences the practical implications and recommendations that result from our research. However, to transform the explorative theorizing we performed in a theoretical framework more research should be done that does include such notions.

In **Chapter 4,** we selected the case studies. We chose to do this by theoretical sampling of the main types of data marketplaces that currently exist according to Kelly (2021): IoT, B2B, and personal. The advantage of choosing these categories for theoretical sampling is that it creates a rich and diverse view of the obstacles that are faced and how these are being dealt with in different types of data marketplaces. This helped us to identify relationships between obstacles and business model choices from three different perspectives. However, the different types of data marketplaces (B2B, IoT, and

personal) appeared to be quite different in their operations. Since the different categories have other difficulties that result from the existing obstacles is therefore difficult to generally speak of "the viability of a data marketplace". A data marketplace always revolves around the exchange of data, but the additional aspects differ. For example, if it would be stated that a personal data marketplace at this moment does not have a viable business model because of the competition for data with big-techs this is not necessarily correct. If the value proposition of the data marketplace is to give its users control over their personal data and monetize it this would be rather difficult in the current situation. However, if the value proposition only includes monetizing personal data this would just be an extension of receiving rewards for filling in surveys. This would take away a part of the obstacles that are faced. The same reasoning applies when comparing IoT to B2B to personal data marketplaces. Therefore, we believe that now we have explored the research area of data marketplace business models and obstacles it is valuable to go more into depth into a certain category of data marketplaces. The theoretical sampling would then not be based on the types of data marketplaces but on certain characteristics of the business model of one type of data marketplace. For example, data marketplaces that would all be categorized under 'personal data marketplace' by definition but have a different value proposition. Also, we chose to select these cases because (1) out of all considered cases their websites and whitepapers were most elaborate and gave us the most detailed information, and (2) these data marketplaces were willing to participate in the research. The case selection was not based on other characteristics of the data marketplace other than it being IoT, B2B, or personal. As a consequence, two out of three cases turned out to be not-for-profit organizations. This does not disprove the relationships we found but does mean that the relationships regarding the financial part of having a viable business model might be underexposed in this research.

In the first part of **Chapter 5**, we prepared the business model stress tests by classifying the three data marketplaces into the taxonomy created by van de Ven (2020). This corresponds to sub-question 1 of the research:

> **SQ 1:** *"How can we describe the business models of three different data marketplaces by using an existing taxonomy?"*

We answered this question by analyzing the websites and the whitepapers of VETRI, Databroker (DAO), and the DX Network. The business models of these data marketplaces could be well described by using the taxonomy. However, what a taxonomy cannot describe is the business model strategy of a data marketplace. This turned out to be a limitation of the research approach we chose to perform to find the relationships between business models and data marketplace obstacles. We will later describe this in more detail. In the second part of Chapter 5, we prepared the other half of the BM Stress Test by making a final selection of the obstacles that we wanted to take into consideration. This corresponded to sub-question 2 of the research:

> **SQ 2:** *"What obstacles can we select for performing a BM Stress Test on the chosen data marketplaces?"*

By the means of explorative theorizing in Chapter 3, we had already defined several obstacle categories for data marketplaces: trust, privacy and control, product description and product quality, matching providers & buyers, pricing mechanism, and other. For the final selection, we combined the findings from literature with our observations of a TRUSTS focus group on data marketplace business models. We found that the BM Stress Test for trust should differentiate between *lack of trust in the system and/or its participants*, and *harm of own business interest*. Also, privacy and control were separated and 'other' was specified to consist of *network effects, incentivizing to participate,*

competitors, and technical issues. After the preparation of both elements we were enabled to start performing the BM Stress Tests and answer sub-question 3:

SQ 3: *"How can we relate the identified obstacles to the business model choices?"*

We chose to do these tests by conducting interviews with employees of the data marketplaces that we selected as case studies. During the interviews, we asked them if obstacle X was perceived as an obstacle for their data marketplace and if yes, how they accounted for it in their business model. In the transcription of the interviews, we marked the answers that could be possibly be used to fill the BM Stress Test Matrix. These answers were first checked against statements made on the websites and whitepapers. If no contradictions were found, they were filled in the BM Stress Test. In this way, the obstacles were related to the business model choices of the selected data marketplaces. From the context of the conversation, we were also able to derive how big the impact of this obstacle was. This enabled us to answer sub-question 4:

SQ 4: "*What is the impact of the obstacles on the business model choices?*"

To visualize the impact we used a color indication. A red color indicated that we found a certain obstacle to be a possible showstopper. An orange color indicated that we found that this aspect requires attention. A green color reflected that there we found no direct negative effects or in some cases even positive effects of certain obstacles. A grey color reflects that we could identify no direct relationship from either the information on the website, the whitepaper, or the interview data. What we should note about the BM Stress Test relationships and the impact assessment is the possibility that it includes biased statements. Because data marketplaces are currently in their infancy only limited resources were available to conduct the case studies. The cases could only be studied from one perspective: that of the data marketplace owner. Of course, these data marketplace owners fully believe in the viability of a data marketplace on work daily basis to overcome the obstacles and achieving their goals. Because of this, their opinions are on the one hand valuable because they give insight into practical experience. On the one hand, their opinions might be biased because they want to make a decent impression. It cannot be known for sure if the business model choices that were said to relate to certain obstacles have the results they intend to have and the true impact of the obstacles is reflected in the interview context. The results have not yet been achieved since the marketplaces are not yet fully operational; they only have or had several successful use cases. Business Model Stress Tests are often used to identify certain patterns from, either within-case or cross-case. Therefore, we expected it to be useful to include sub-question 5 in the research:

SQ 5: "*What are the patterns between the business model choices and obstacles?"*

The most useful insight that resulted from the pattern recognition analysis is that business model choices have inconsistencies. This means that a certain choice can lead to a green relationship with one obstacle but a red or orange relationship with another. Another thing we noticed is that it is difficult to generalize from within-case patterns to cross-case patterns. We believe that this is due to from due to the context-specificity of the BM choices that are made in a specific (type of) data marketplace. This is also what we noted when trying to answer sub-question 6:

SQ6: *"What are the hypotheses on data marketplace obstacles and business model choice relationships?"*

We already went into detail on the discussion of these results in Chapter 7.2 but will elaborate on it a bit further. Initially, we imagined that because a data marketplace business model taxonomy represents all business model choices a data marketplace can make, it would be appropriate to use it as a means to map the business model choice and obstacle relationships. However, we now realize

that a business model choice is not an isolated concept. Our research approach underexposed the importance of the choice in a broader context: the firm's strategy (as described in Ch 3.2). A business model is only an abstract representation of some aspect of a firm's strategy (Seddon et. al, 2003) and this is why we believe that it is difficult to formulate concrete relationships between business model choices and data marketplaces obstacles. The business model choice is not just a "random" choice between the options, it is part of a larger construct and will only have a certain impact if it is strategically used correctly. To make this more clear we can use an example: If a data marketplace chooses to use cryptocurrency rather than fiat currency its implications for the impact of a data marketplace obstacle would not directly be clear. You could say that it would add additional instability on top of the fact that data is hard to price due to its characteristics, as the crypto market is very volatile. This practically means that the data price is not only influenced by its content, but also by the exchange rate of the chosen cryptocurrency token. But, this would be different if a data marketplace would use stablecoins. So, we cannot say that choosing cryptocurrencies has a certain relationship with the impact of an obstacle. We only know that there is some kind of connection between the choices that are made in the payment currency dimension and the pricing obstacle. This aligns with what DaSilva & Trkman (2014) say about the connection between business models and strategies. To recap, they stated that a "*strategy (a long-term perspective) sets up dynamic capabilities (a medium-term perspective) which then constrain business models (present or short-term perspective) to face either upcoming or existing contingencies"* (DaSilva & Trkman, 2014, p.383). According to this perspective, a firm should not need to change its business model to respond to contingencies, but can strategically prepare itself by developing dynamic capabilities. Looking at our results, we can support this claim. Not everything should necessarily be about the business model choice you make for your data marketplace, it is rather about how you choose to handle the consequences of making this choice. Any choice should work as long as you consider the implications of making that choice. We believe that this can, for example, be based on financial reasons (e.g. the one choice is more expensive than the other choice), personal preferences, employee experience, but also based on the impact of obstacles. Our explorative research approach and its results enable us to comment on the considerations that can (and should) be taken into account when making a business model choice from an obstacle perspective. We will address this in section 10.2: Practical implications of the results. As we described, our BM Stress Test consisted of positive (green) and negative (red/orange) relationships. Because this was only based on the context of the interviews, and thus might give a biased perspective, we deemed it necessary to reflect on the relationships before creating the explanatory models. This corresponded to sub-question 7:

SQ 7: *"Can we verify the relationships that resulted from the BM Stress Tests?"*

Based on what the discussed above, we could not verify the specific business model choice and obstacle relationship. However, we could adjust the relationships to a broader concept such as business model dimension or business model strategy. Therefore, we answered this sub-question by reflecting on the negative instances of the obstacles in all obstacle categories. We did this by exploring the literature for possible solutions and asking data marketplace experts for their opinions. We analyzed the findings and included the results in the explanatory models. These explanatory models were structured by marketplace requirement and include the expected influences on data marketplace business models of obstacles as well as external factors.

Reflecting on the overall results of our research we believe that the viability of data marketplace business models does not only have to do with the fact that they are defined to be a 'data marketplace'. Their viability has to do with how well their business model choices enable them to comply with the marketplace requirements under influence of the obstacles and external factors.

Therefore, we believe that to say something concrete about viability it would be more useful to individually assess the business model choices and strategy of a specific data marketplace. However, our research does lead to practical implications of business model choices that can be considered and broader recommendations that go beyond business model choices. The practical implications and recommendations will respectively be discussed in the following sub-sections.

10.2 Practical implications of the results

In the challenges and limitations, we have discussed that based on the results of this study we cannot say that it is a specific business model choice a data marketplace makes that matters for the viability of a data marketplace. We thus cannot make conclusions about business model choice X, direct impact on obstacle Y, and indirect impact on the viability of a data marketplace.

However, based on our results we can hypothesize on the importance of the context in which business model choice X is made for overcoming obstacles and assessing and/or increasing viability. Or, in the words of DaSilva & Trkman (2014), how a data marketplace can strategically prepare itself by developing dynamic capabilities based on the business model choices they make. In the previous section, we illustrated this with the example of choosing cryptocurrency as a payment currency. We will now do this for all business model dimensions in the four STOF domains that we can say something about based on the results of our research. This will give a first overview of the hypotheses of considerations that should be taken into account when making a BM choice (from an obstacle perspective). Table 9 shows an overview of the considerations per business model dimension. We will discuss the table per STOF domain in the following sub-sections.

Table 9: Overview of the considerations that can be taken into account when making a business model choice from an obstacle perspective (per STOF domain and BM dimension)

	Dimension	**Considerations for making a BM choice (obstacle perspective)**
Service domain	Value proposition	-Details of the value proposition
	Enterprise data marketplace	-Opportunity to build trust -Allows for more sensitive cases of data exchange
	Data processing and/or analytics tools	-Opportunity to increase perceived product quality (but presumably fewer “out-of-the-box” data exploration opportunities for customers)
	Marketplace participants	-Different participants have different norms and values, resulting in a different impact of obstacles -Know your customer
	Industry domain	-Different implications for different industry domains: sensitivity, obsoleteness, technicalities of the data
	Geographic scope	-Pool of participants (big-scale network effects vs. niche market specificity) -Regulations
	Time frame	-Different approach for product description (combination with industry domain)

Technology domain	Platform architecture	-Difference in number of competitors -Possibility for trust-less data marketplace environment -Difficulty of technicalities
	Data access	-Privacy (/security)
	Data source	-Dependency on the capabilities of data providers
Organization domain	Matching mechanism	-Pool of participants (big-scale network effects vs. niche market specificity)
	Platform sponsor	-Foundation (not-for-profit) as platform sponsor
Finance domain	Revenue model	-Cryptocurrency token as an asset sale
	Pricing model	-Pricing model that allows charging both market sides
	Price discovery	-Should incentivize participation
	Smart contract	-Warranty for data quality -Automatic (micro)payments -Difficult to understand
	Payment currency	-Importance of simplicity for the user -Opportunity to fund data marketplace (ICO) when creating own token (crypto) -Dependency on development of the (volatile) third market (crypto)

10.2.1 Service domain

For the choice for a **value proposition,** we have argued that its relation to obstacles can go beyond the formulation in the business model taxonomy. We have seen and described that for example for a personal data marketplace that aims to offer control over personal data as well as monetization, the obstacle of big-tech competition has a different impact than for a personal data marketplace that only aims to offer monetization. To assess the viability of a data marketplace, it is thus very important to look at the exact definition of the value proposition and its consequences for the obstacle environment. A simpler value proposition will likely be easier to realize but will also be less valuable for the data economy.

The choice for an **enterprise data marketplace** can provide an opportunity to build trust within a group of participants through first-hand experience. Also, this BM choice provides the means to handle more sensitive cases of data exchange (because it is private), which can initially attract different groups of participants than an "open" data marketplace (where every individual or organization can participate).

In the cases studies we conducted there were no instances of data marketplaces that offered **data processing and/or analytics tools.** However, we saw that having no control over data – in the sense that you do not know where it will end up and what will be done with it after it is offered on the marketplace – can keep individuals or organizations from participating. This will be less when a data

marketplace chooses to only sell analytics and it is known what the data providers are “giving away” upfront. Also, offering analytics rather than raw data lowers the risk that data consumers will deem the data of low quality and unusable because it offers ready-made insights. However, we assume that it will also lead to disadvantages as it will lead to fewer “out-of-the-box” data exploration opportunities for data buyers.

The **marketplace participants,** the choice-obstacle relationship mainly concerns that different marketplace participant groups have different norms and values. One group might perceive the risks of participating in a data marketplace as higher than the benefits, while another group might see just the contrary. Therefore, we believe that for this choice concerning the link to obstacles it is important to identify what is important to them (know your customer). A data marketplace owner can then start to develop the right measures for its customer groups. For example, if in a B2B market loss of competitive advantage is a problem, try to find a sweet spot in the data that is valuable but less sensitive and can be shared to enlarge the willingness to participate.

When choosing an **industry domain** for a data marketplace it is mainly important to consider the different implications of that domain. One domain will contain more sensitive data (e.g. health & personal data), another domain includes data that will go obsolete (e.g. financial & alternative data), and another will have larger dependencies on the development of other technologies (e.g. sensor & mobility data depends partly on IoT technology development). Even within a specific domain choices can (and should) be made. For example, personal data can include usage and behavior data but it does not have to. Specificity in defining the data that will be sold on a data marketplace will lead to the most accurate relationships with obstacles and therefore is important to get a clear idea of the viability.

The **geographic scope** of a marketplace matters for the pool of participants that can be included. On the one hand, a large geographic scope will enable a data marketplace to profit from network effects that can be created. In that case, a large pool would mean that the odds are increased that matches between data providers and data buyers can be made. On the other hand, a small but niche-specific pool could increase the quality of the match, meaning that the data is deemed useful because the data requirements are well aligned (due to the known characteristics of the niche market). Choosing a larger (global) scope will also mean dealing with different global regulations on data trading.

When choosing a **time frame** the approach for product description should be considered. This is often in combination with the industry domain. For example, a static strategy dataset will require a different approach than a real-time sensor dataset.

10.2.2 Technology domain

When choosing a **platform architecture** we believe that it is important to consider that a decentralized architecture, using blockchain technology, allows for creating a trust-less environment. This can be valuable for data marketplace adoption. Also, the number of existing decentralized marketplaces is less than the centralized ones meaning there is less competition. However, a decentralized architecture easily becomes very technical. These technicalities should not be reflected in the customer interface. This should remain as simple as possible to make it easy for the user to adopt and use the data marketplace.

When considering the choice for **data access** a data marketplace owner can consider the impact on privacy (and security). We saw this in the case of a personal data marketplace that aims to integrate digital identity in its system to access the data. This would increase the perceived security and privacy of the data marketplace.

In the cases that we studied all data marketplaces worked with customer-provided data as a **data source.** Based on this study we find that the main consideration for this choice should be the dependency on the capabilities of data providers. Even if the providers would be willing to

participate and sell their data they might not be capable of doing so because of legacy IT systems or lack of resources in their organization. Data marketplace owners should address these shortcomings.

10.2.3 Organization domain

For the **matching mechanism,** we find that the same argumentation applies as for geographic scope. The advantages of profiting from network effects of many-to-many matching should be weighed up against the value of more specific (for example one-to-one) matches.

For the **platform sponsor,** a data marketplace could consider a foundation as a platform sponsor to increase trust. However, this would mean that its business model would be not-for-profit. Its intention would then be to make a contribution to the data economy for the good of society rather than for its own.

10.2.4 Finance domain

For the choice of the **revenue model** of a data marketplace, we found that we can further specify the option for choosing asset sales as a revenue model. When a data marketplace chooses to include cryptocurrencies as payment currency (see **payment currency**) and creates its own token, it can use the token sale as a way to generate revenue. This revenue model could mean that less revenue has to be taken directly from data provider–data buyer transactions, which makes it more attractive for these stakeholders to participate in the data marketplace.

For **pricing model,** we found that it might be interesting to implement a pricing model for both market side – data providers and, logically, data buyers. The pricing model for data buyers could be implemented to decrease the number of malicious submissions, making the data marketplace more trustworthy.

The mode of **price discovery** of a data marketplace should incentivize participation. The choice enfolds a complex consideration: the stakeholders who set the price might have an advantage. If the data buyer sets the price and the data provider is unaware of the value of its data the monetary compensation might be too low. If the data provider sets the price the data buyer might pay too much for what the data is worth in the context of the application. This could be avoided by using negotiation as price discovery mode (although these transactions might be less quick and efficient), let the marketplace provider set the price based on some uniform pricing standard (when available in the future) or technical solutions that allow buyers somehow to try out the datasets before buying it, overcoming the Arrow paradox.

Choosing to implement a **smart contract** will enable data buyers to have an automatically integrated warranty on the data quality as payments can be frozen when a complaint is filed about the data provider. Smart contracts also make automatic (micro)payments possible, making it easier to share revenue amongst many different parties that might be involved in the transaction. However, we found that it is important to realize that smart contracts will not contribute to the transparency of data marketplaces and data transactions, may this be important for data marketplace participants. Coded contracts will likely be even harder to understand than written contracts.

Choosing a cryptocurrency as **payment currency** for a data marketplace can be an opportunity to fund the marketplace by an Initial Coin Offering (ICO) (if an own token is created). This does however mean that the marketplace will depend on the development of a very volatile third-party market. Also, the marketplace should in some way control the buyers of their tokens to also be participants of the data marketplace for the system to work. If a data marketplace wishes to avoid the volatility of cryptocurrencies it can choose to use stablecoins. These coins follow the value of fiat currency and therefore discards the risk of great fluctuations in price. When choosing to use a cryptocurrency the data marketplace should take into consideration that the technical details (e.g. having to buy tokens first) might be perceived as more difficult and time-consuming for data providers. Therefore, a data

marketplace can also choose to combine the use of fiat currencies and cryptocurrencies. A data buyer would then pay for the data in fiat, which is then converted to a cryptocurrency during the transaction.

We have now discussed what data marketplaces can and should take into consideration for their business model strategy when making business model choices. However, as we discussed, the viability of a data marketplace is not only affected by the business model choices that are made. As was visualized in the explanatory models, we also identified some external factors that the business model choices, as defined in the taxonomy by van Ven (2020), have little to no influence on. It is unknown to what extent the business model choices and corresponding strategy affect the obstacles in comparison to the external factors. Therefore, it is very difficult to assess the (current) viability of data marketplaces based on obstacles and business model choices. What we can say is that data marketplaces have to deal with many obstacles that slow down their development. Some obstacles are bigger than others and therefore have a bigger impact on the viability of the business model. What we can also derive from our findings is that it is under all circumstances important to try and comply with the marketplace requirements of safeness, no repugnance, provenance information, thickness, and no congestion. Complying with these principles would lead to efficient market operation and a viable data marketplace. Therefore, we want to convert our findings to a more general proposition:

> *"The better the marketplace requirements are safeguarded by considering the business model strategy for a business model choice (from an obstacle perspective) and the influence of the external factors, the higher will be the viability of the data marketplace."*

We say "from an obstacle perspective" because a choice can also be made for other reasons than decreasing the impact of obstacles, e.g. financial reasons. We have already discussed the business model choice and corresponding business model strategy part. However, based on the results of this study, we can also make some general recommendations that go beyond business model choices from the taxonomy to safeguard these requirements. We will discuss this in the next sub-section.

10.3 General recommendations beyond explicit business model choices

Based on the findings some general recommendations can be made that would possibly decrease the impact of obstacles and with that increase the viability. These suggestions (partly) go beyond explicit business model choices and are connected to the marketplace requirements, as these would have a viable data marketplace as an outcome.

1. **Safeguard safeness:** For safeguarding the safeness requirement for a data marketplace trust and governance are the constructs to consider. A trust relationship is hard to build in an environment that is focused on quick interaction and exchange. Therefore, it would be valuable to decrease the need for trust by creating a trust-less environment as far as possible. Where not possible, prioritize trust-creating activities as part of the business model. This is currently modeled as an external factor, but could even be part of the value proposition. This can then also be included in the promotion of a data marketplace. This would be useful as the way trust is perceived by marketplace participants is also influenced the marketing and communication of the data marketplace owner and among data marketplace participants. It is thus beneficial to stress the trustworthiness of a data marketplace in marketing and communication. For example, a referral program that stimulates public dialogue among participants will create (indirect) trust. It should be considered that different users might have different needs to achieve trust. Some might only need to be notified of the benefits, some might require some education or explanation on the technologies used. Governance was modeled as a factor that cannot be influenced by the data marketplace itself but does affect the development of the business model. Therefore, it can only

be suggested to work hand in hand with (upcoming) regulations and design a data marketplace business model accordingly. The data marketplace should anticipate the regulations and their consequences and profit from its implementation when possible. Value for its participants can be added when joining a data marketplace gives a guarantee of compliance. It would be attractive for participants if they would not have to be bothered with governmental regulatory measures.

2. **Safeguard no repugnance:** For safeguarding the no repugnance requirement for a data marketplace privacy and control are the constructs to consider. Mainly for privacy, the business model is under the influence of rules and regulations that should be complied with. A data marketplace is by definition the opposite of keeping data private as it enables the exchange of data. However, the privacy implications did not seem to limit the development of data marketplaces. It is rather the direct relationship that privacy has with control. As explained it is very difficult to keep or allow for control over data because it has no IP rights. Because it is hard to guarantee control over raw data, data marketplaces might also take the approach of selling pre-determined insights or providing consultancy on decisions that can be made based on the data. In this case, it is exactly known what is being sold and it, therefore, allows for more control over the data exchange. This way of working will still allow businesses to realize data-driven innovations based on the insights acquired from a data marketplace. Giving the users of personal data marketplaces control over their personal data is a different aspect of control. In this case, it could be recommended to not include behavioral and usage data in the first instance as this includes a lot of competition from big-techs. However, it of course depends if the value proposition of the personal data marketplace is only to allow its customers to monetize their data or to also give them control over it.

3. **Safeguard provenance information:** For safeguarding the information provenance requirement for a data marketplace product description and product quality are the constructs to consider.
   Safeguarding is difficult due to two external factors: metadata generation and data requirements. These factors form great difficulties as metadata is hard to generate and the data requirements are context-dependent. The suggestion that can be made partly overlaps with the suggestion made for safeguarding the no repugnance requirement. That is because selling consultation on data or insights on data does not include describing a dataset and makes it easier to guarantee quality. For raw datasets, it only seems possible to describe datasets in a standard manner if they concern a standard subject. For example, a weather dataset on measured temperatures throughout Europe would always include the country and the measured temperature. Strategic datasets of different businesses would probably contain different information. In all circumstances, the data requirements of data buyers should be mapped. A larger overlap between the data requirements and the dataset will lead to higher perceived product quality. Because these requirements can be very broad, it seems easier to deal with the consequences if the context in which data is sold is more stable. For example, a data marketplace that is developed for a niche industry will get more easily progress in getting familiar with the industry standards, the players, and thus the data requirements. This would increase the odds that the data can be well described, deemed useful and therefore leads to higher quality data exchange. To overcome this obstacle otherwise, an efficient method of metadata generation needs to be developed that also considers the data requirements of a buyer.

4. **Safeguard thickness:** For safeguarding the thickness requirement for a data marketplace matching and 'other' obstacles such as network effects, competitors, technical issues, incentivization are the constructs to consider. This requirement includes the external factor that

can be said to have the greatest influence on the business model viability of data marketplaces. If a data marketplace has no pool of participants, no data can be exchanged and thus no business model can be exploited. This pool will be big if there is a high willingness to participate and a high ability to do so. Both aspects in general are the main concern to safeguard the thickness of the marketplace. Therefore, to make data marketplace business models viable, it should be researched precisely under what circumstances users would be willing to participate. It can even be suggested to create "risk profiles" based on the sensitivity of the data. Risk-averse businesses that would want to participate in a data marketplace would be enabled to exchange public data, and risk-seeking businesses would be enabled to exchange more sensitive data such as strategy data. Public data would then lead to a lower reward in comparison to more sensitive because it also carries a lower risk for sharing. Just as in investing, this would be a high-risk high-reward approach. Because willingness on its own is not enough, the ability also requires attention. Data marketplaces could provide a perfect infrastructure for data exchange, but if the data cannot be unlocked at the data provider due to their restricted resources there will also be no business case. To increase the viability of a data marketplace business model it is thus not only important to consider own resources, but also the resources of the participants that would be involved on the platform. Another remark can be made on the relationship between the pool of participants and the matching on a data marketplace. A more narrow pool of participants, such as in a niche industry, would increase the odds of useful matches being made. Therefore, it seems valuable for a data marketplace to start their business in this way. Once experience has been gained, the business can expand to other markets. Also, the dependence on the development of interrelated technologies should not be underestimated. A data marketplace does not operate in solitude and is dependent on how its business model choices are intertwined with the technologies that are needed for that choice. For example, an IoT data marketplace depends on technology that enables communication of the IoT device with a network at low energy consumption. Something that does not yet seem possible until the realization of a 5G network. Data marketplaces should ensure that they stimulate the creation of network effects. An example could be offering a reward for bringing in new participants. Furthermore, a close eye should be kept on competitors and the technical back-end should not result in a technical front-end. Data providers, as well as buyers, should be incentivized to participate, either through financial or other means.

5. **Safeguard no congestion:** For safeguarding the no congestion requirement for a data marketplace pricing is the construct to consider. It is hard to make suggestions on this construct as a real standard would either be based on the development of pricing models or on the experience that data marketplaces have gained from previous matches. To gain experience, participants are needed. Attracting those participants and making the matches is therefore vital. Apart from practical experience, a lot of academic research is being done on the subject of data pricing. It is therefore useful for data marketplaces to stay up-to-date on the newest developments in this research field.

10.4 Scientific contributions

In this research, we provided an exploration of a new research context: the relationships between business models of data marketplaces and data marketplace obstacles. Our literature research showed that there are three broad research topics included in data marketplace research (1) data marketplace business model classification *(*Schomm et al. (2013); Stahl et al. (2014); Koutroumpis et al. (2017); Spiekermann (2019); Fruhwirth et al. (2020); van de Ven (2020)) (2) technological solutions (Sharma et al. (2020); Azcoitia & Laoutaris (2020); Perera et al. (2017)**;** Mišura & Žagar (2016); Parra-Arnau (2018)) and (3) isolated remarks on data marketplace obstacles (Spiekermann (2019); Lawrenz et al. (2019), G. Smith et al. (2016); Nikandor & Elo (2019)). However, data marketplaces form a complex socio-technical system, meaning that these topics are all interrelated. To our best knowledge, this is the first work that takes an obstacle perspective on data marketplace business models and tries to identify these interrelationships. By taking this approach, our study makes several contributions to existing literature.

The first contribution is that we managed to provide a comprehensive overview of data marketplace obstacles and their implications. In other words, by the means of explorative theorizing, we managed to identify, interrelate and examine the constructs that are important to consider when assessing the viability of data marketplace business models from an obstacle perspective. We showed how the requirements for efficient market operation (Roth 2002, 2007) can be applied in the context of data marketplaces by linking them to the data marketplace obstacles, something that has remained unclear so far (Koutroumpis, 2017). The overview on data marketplace obstacles is a contribution in itself as previously no overviews existed in literature. Furthermore, interrelating the obstacles and the marketplace requirements gives (business model) context to and stresses the importance of overcoming the obstacles. In this way, it gives more depth to isolated remarks on challenges, difficulties, barriers, or obstacles by different authors such as Spiekermann (2019), Lawrenz et al. (2019), G. Smith et al. (2016), and Nikandor & Elo (2019).

The research also adds to the literature by presenting a first explanatory model on the relationships between data marketplace business models and data marketplace obstacles. This model is useful for practitioners because it leads us to insights on the practical implications of making a certain business model choice (from an obstacle perspective) (See Ch. 10.2). It also gives an overview of other factors that should be acknowledged when designing a data marketplace business model (strategy). Previous studies have not extensively dealt with non-technical topics like this (Abbas et al., 2021). We provide details and context to the claim by Aaltonen et al. (2021) that data commodification is a complex socio-technical practice, instead of just technical. We complement the current technology-focused data marketplace research with empirical research from a non-technical business model perspective. Therefore, our research is a contribution to data marketplace commercialization from an academic perspective.

We contribute to BM literature by providing practical tooling of the Business Model Stress Test (Bouwman et al., 2018). The BM Stress Test was applied in a different mode than it normally would since uncertainties instead of obstacles are used. In this study, we thus show a new context for application. Even though it appeared that to create an explanatory model this approach underexposed the context-dependency and importance of a firm's business model strategy, we believe our BM Stress Test set up will be useful as a basis to assess the viability of individual data marketplaces. The strategy a data marketplace chooses to pursue for a certain business model choice can then be included in the impact assessment (e.g. if the BM choice is cryptocurrency, choosing a stablecoin will correspond to a different impact than a volatile coin).

10.5 Recommendations for future research

First of all, this research aimed to create a broad overview of the relationships between business model choices and data marketplace obstacles. Therefore, it was chosen to select the cases by the means of theoretical sampling of an IoT, B2B, and personal data marketplace. As we already stated in the challenges and limitations section, their business operations turned out to be quite different, which made it more difficult to generalize the implications of their business model choices. Performing a case study research within one data marketplace category will possibly lead to more specific implications and therefore recommendations specifically on IoT, B2B, or personal data marketplace business models.

Secondly, this case study research only included the perspective from the data marketplace owner. Since a data marketplace is a multi-sided platform, it involves multiple stakeholders whose perspectives on the viability of data marketplaces are not included in this research. For example, data providers and data buyers, the pool of data marketplace participants, might have another view on the obstacles. Understanding their perspective might deepen the knowledge of some of the identified relationships. For example, on this willingness and ability to participate in a data marketplace. It could be interesting to talk to tech companies (for IoT & B2B) and see under what circumstances they would be willing to participate. Their ability to participate is also a very interesting research topic: what resources do data providers, that could make a big contribution to society, need to share their data on a data marketplace? Even if not all obstacles can be solved, it might be possible to make the circumstances more attractive. The same goes for individuals that would be the target group of personal data marketplaces: what are their needs and wishes and how can they be attracted to participate in a data marketplace? The third-party service providers were also left out of this research. It would be interesting to see what their role could be in overcoming the data marketplace obstacles. For example, one of the suggestions was to sell consultation rather than raw data. The third-party service providers could have a big role in this as they are concerned with uploading algorithms and applications to add value to the data for the data buyer. We believe that including more stakeholders will help with identifying the innovation sweet spot for data marketplaces. This means that four criteria are satisfied: desirability, viability, feasibility, and integrity (Shahbazi, n.d.).

Thirdly, our explanatory models could be improved and extended. Only a limited amount of case studies has been done to identify the obstacles and the relationships. Other case studies might lead to new relationships or even new obstacles. This will in its turn lead to more practical implications and recommendations. By performing more research, our framework that resulted from explorative theorizing can be enhanced to a theoretical framework. Academia can use the overarching obstacle structure that we provided as a starting point for their research.

CH 11. Conclusion

Now that we have discussed the answers on and challenges and limitations of the sub-questions, the corresponding practical implications, and the more general recommendations, we can give a final answer to the main research question: *"How can we use the relationships between business model choices and obstacles to assess the viability of data marketplace business models?"*. We will discuss our answer to this question first in the overall conclusion. Afterward, we will discuss the societal and managerial relevance of the research and the relevance to the CoSEM program.

11.1 Overall conclusion

This research started based on the observation that many data marketplaces shut down – implying that their business models are not viable – and the lack of literature on the why behind this. We considered this to consist of the relationships between the business choices of data marketplaces and the obstacles that they are facing. These relationships would map the friction and help identify the difficulty in making a data marketplace business model viable.

By the means of explorative theorizing we came up with a framework that showed the relationships between the desired outcome – a viable data marketplace business model –, the marketplace requirements, the obstacles, and the business model choices. During the BM Stress Tests, the different business model choices indeed showed to have a different impact on different obstacles. However, this turned out to be very context-specific. The business model choice is not just a "random" choice between the options and is not right or wrong by definition. It is part of a larger business model strategy and its impact on an obstacle will also depend on the dynamic capabilities of the data marketplace. Apart from BM choices, we saw that external factors such as trust creating activities, marketing and communication, rules and regulations, third parties, understanding and awareness of users, methods for metadata generation, insight in data requirements, development of interrelated technologies, the willingness of customers to participate, the ability of customers to participate, development of (algorithmic) pricing models and practical experience also influence the impact of obstacles. Some of these factors, such as rules and regulations on data exchange, cannot be controlled by a data marketplace owner. For these reasons, we argue that the obstacles cannot fully be overcome by only making certain business model choices.

Therefore, to use the relationships between business model choices and obstacles to assess the viability of data marketplace business models we need to adopt a broader perspective. Based on the results of this research, we could not make conclusions about business model choice X, direct impact on obstacle Y, and indirect impact on the viability of a data marketplace. However, we could hypothesize on the importance of the context in which business model choice X is made for overcoming obstacles and assessing and/or increasing viability. Or, in the words of DaSilva & Trkman (2014), how a data marketplace can strategically prepare itself by developing dynamic capabilities based on the business model choices they make. A selection of our most interesting findings concerning the considerations that one can take into account when making business model choices include that: a choice for an enterprise data marketplace can be an opportunity to build trust and allows for more sensitive cases of data exchange, data processing and/or analytics tools in a data marketplace form an opportunity to increase perceived product quality, a decentralized blockchain-based platform architecture allows for a trust-less data marketplace, a data marketplace is dependent on the capabilities of its providers may the act as their data source, cryptocurrencies can be used to fund a data marketplace but its user interface should remain simple. This is also true for the use of a smart contract, which allows for automatic (micro)payments and can give a warranty on data quality, but is not transparent and can be difficult to understand.

However, this broader perspective on data marketplace business model viability also includes external factors that go beyond the direct implications of business model choices. We believe it is also concerned with how well a data marketplace can live up to the marketplace requirements regarding choices that cannot be directly found in a taxonomy. This is why we based our recommendations on the following proposition: *"The better the marketplace requirements are safeguarded by considering the business model strategy for a business model choice (from an obstacle perspective) and the influence of the external factors, the higher will be the viability of the data marketplace."*

The marketplace requirements we speak of are safeness, no repugnance, provenance information, thickness, and no congestion. In short, to safeguard safeness we recommend creating a trust-less environment for as much as possible. Also, we think it would be beneficial to stress the trustworthiness of a data marketplace in marketing and communication. Furthermore, a data marketplace should anticipate the upcoming regulations and profit from its implementation when possible. To safeguard no repugnance data marketplaces could consider selling pre-determined insights or provide consulting on the decisions that can be made based on the data. This allows for more control over the data exchange. This same argumentation can for safeguarding provenance information. Product description and product quality will be less of a problem when insights are sold instead of raw data. Another option is the development of an efficient method of metadata generation that also considers the data requirements of a buyer. To safeguard thickness the main point to consider is the pool of participants. Whether working with a narrow pool or big pool of participants for making matches, a data marketplace should ensure that their business model leads to a high willingness to participate and a high ability to do so among its participants. To safeguard no congestion the only recommendation that we can make is staying up-to-date on the newest developments in the research field of data pricing mechanisms.

In summary, our research provided an exploration of a new research context: the relationships between business models of data marketplaces and data marketplace obstacles. It has become clear what obstacles should be dealt with and how these obstacles hinder the marketplace requirements that have a viable data marketplace business model as an outcome. Based on the results, we cannot assess if data marketplace business models are viable or not based on business model choices and the relationships with the obstacles. It is not unequivocal what business model choices will lead to the best results and it is unclear how the choices compare to external factors that can also influence the data marketplace viability. We believe that is data marketplace viability is rather based on the how than the what, meaning that any choice can work as long as it is implemented in a way that considers the effect on the obstacles.

So, is data really the new oil? The metaphor would hold because both oil and data are not valuable by themselves. It becomes valuable when gathering, connecting, and using it correctly. To achieve this, among other things, a mode of storage and transportation is needed. It could be argued that a data marketplace is an infrastructure for the transportation of data. A marketplace "transports" the data from one market side to the other. However, this mode of transport is yet underdeveloped and has to overcome many obstacles and deal with uncertain external factors. From this perspective, data cannot be claimed to be the new oil. Not yet.

11.2 Societal and managerial relevance

From a societal point of view, this research is relevant because data marketplaces have lots of potentials to stimulate the data economy and leverage the abundant amount of data for economic benefit. The data can be a source of innovation that society can profit from. However, we are not at the point yet where data marketplaces have the same impact as other electronic commerce platforms such as Amazon or eBay. Therefore, any research that aims to provide a more thorough understanding of the issues around data marketplace commercialization is a step towards fulfilling data marketplace potentials. Especially since little research has been conducted on non-technical data marketplace topics like ours, our explorative research can lead to points of departure for other researchers that are interested in this research context.

In terms of managerial relevance, we give practitioners an overview of the considerations that can be taken into account when designing a data marketplace business model or business model strategy. We also give recommendations on how to safeguard the marketplace requirements with the intent to overcome the data marketplace obstacles and create a viable data marketplace business model. This is not all-encompassing but can be used as a starting point for further investigation.

11.3 Relevance to the CoSEM program

The Complex Systems Engineering and Management (CoSEM) program focuses on realizing innovations in a complex socio-technical environment. A data marketplace is an example of such an innovation. It is socially complex since it involves many stakeholders such as data marketplace owners, data providers, data buyers, and third-party service providers. It is technically complex since it involves the implementation of many different technologies such as blockchain, smart contracts, and cryptocurrencies. Not only should the wishes and needs of the stakeholders socially be aligned but the technology should also enable it to do so. This alignment acts under the pressure of certain obstacles or external factors, as identified in this study. If a CoSEM engineer should design an innovation in the complex social-technical environment of data marketplaces, these obstacles and factors should be considered in the decision-making process.

Bibliography

Aaltonen, A., Alaimo, C., & Kallinikos, J. (2021). The Making of Data Commodities: Data Analytics as an Embedded Process. *Journal of Management Information Systems*. https://doi.org/10.1080/07421222.2021.1912928

Abbas, A. E., Agahari, W., van de Ven, M., Zuiderwijk, A., & de Reuver, M. (2021). Business Data Sharing through Data Marketplaces: A Systematic Literature Review. *34th Bled EConference 2021*.

Ackoff, R. L. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, *16*(1), 3–9.

Agiwal, M., Saxena, N., & Roy, A. (2019). Towards Connected Living: 5G Enabled Internet of Things (IoT). *IETE Technical Review (Institution of Electronics and Telecommunication Engineers, India)*, *36*(2), 190–202. https://doi.org/10.1080/02564602.2018.1444516

Amaro, S. (2021, May 27). *Big Tech mergers: EU prepares new rules but three countries want more control*. https://www.cnbc.com/2021/05/27/big-tech-france-germany-the-netherlands-want-more-control-over-mergers-.html

Asen, E. (2021, March 25). *Digital Tax Update: Digital Services Taxes in Europe*. Tax Foundation. https://taxfoundation.org/digital-tax-europe-2020/

Azcoitia, S. A., & Laoutaris, N. (2020). *Try Before You Buy: A practical data purchasing algorithm for real-world data marketplaces*. http://arxiv.org/abs/2012.08874

Baden-Fuller, C., & Morgan, M. S. (2010). Business models as models. *Long Range Planning*, *43*(2–3), 156–171. https://doi.org/10.1016/j.lrp.2010.02.005

Banerjee, P., & Ruj, S. (2018). *Blockchain Enabled Data Marketplace -- Design and Challenges*. http://arxiv.org/abs/1811.11462

Bouwman, H, de Vos, H., & Haaker, T. (2008). *Mobile service innovation and business models*. Springer Science and Business Media.

Bouwman, Harry, Heikkilä, J., Heikkilä, M., & Leopold, C. (2018). Achieving agility using business model stress testing. *Electron Markets*, *28*, 149–162. https://doi.org/10.1007/s12525-016-0243-0

Carlos Roca, J., José García, J., & José de la Vega, J. (2009). The importance of perceived trust, security and privacy in online trading systems. *Information Management & Computer Security*, *17*(2), 96–113. https://doi.org/10.1108/09685220910963983

Chen, L., Koutris, P., & Kumar, A. (2019). Towards Model-based Pricing for Machine Learning in a Data Marketplace. *Proceedings of the 2019 International Conference on Management of Data*, 18. https://doi.org/10.1145/3299869.3300078

Chesbrough, H., & Rosenbloom, R. (2003). The Dual-Edged Role of the Business Model in Leveraging Corporate Technology Investments. In *Taking Technical Risks: How Innovators, Managers, and Investors Manage Risk in High-Tech Innovations*. MIT Press. https://books.google.nl/books?hl=nl&lr=&id=BW9gBEGak_IC&oi=fnd&pg=PA57&dq=Chesbrough+%26+Roosenbloom+2003&ots=2D2_QE4U7y&sig=gVSvyx0P7jigKQhKxlHhRabFa4Y&redir_esc=y#v=onepage&q=Chesbrough %26 Roosenbloom 2003&f=false

Chowdhury, M. J. M., Ferdous, M. S., Biswas, K., Chowdhury, N., Kayes, A. S. M., Watters, P., & Ng, A. (2019). Trust modeling for blockchain-based wearable data market. *Proceedings of the International Conference on Cloud Computing Technology and Science, CloudCom*, *2019-December*, 411–417. https://doi.org/10.1109/CloudCom.2019.00070

Cosgrove, A., & Kuo, J. (2020, May 6). *Why Data Marketplaces Tend to Fail*. https://www.harbrdata.com/insight/why-data-marketplaces-tend-to-fail/

DaSilva, C. M., & Trkman, P. (2014). Business Model: What It Is and What It Is Not. *Long Range Planning*, *47*(6), 379–389. https://doi.org/10.1016/J.LRP.2013.08.004

Databroker.global. (2020). *Databroker: The Next Generation Data Marketplace*.

De Nederlandsche Bank. (n.d.). *Wat u moet weten over PSD2*. Retrieved June 10, 2021, from https://www.dnb.nl/innovatie-in-betalen-en-bankieren/fintech-ai-en-innovatie/wat-u-moet-weten-over-

psd2/

Economides, N., & Lianos, I. (2021, February). *Giving Away Our Data for Free is a Market Failure*. https://promarket.org/2021/02/01/free-data-market-failure-digital-platforms/

Eisenhardt, K. M. (1989a). Building Theories from Case Study Research. In *Source: The Academy of Management Review* (Vol. 14, Issue 4). https://www.jstor.org/stable/258557

Eisenhardt, K. M. (1989b). Building Theories from Case Study Research. *The Academy of Management Review* , *14*(4), 532–550. https://www.jstor.org/stable/pdf/258557.pdf?casa_token=77I6WodVWj0AAAAA:KeHWo-GfikCWdezK_bjfAfwwyvPpsnu7QLiVqIhJpLRYtFsGlTnO49Spp_TS6Tb7MwtNDG_6Z9g8rmPbHBykIBz3FKt6pzQR-kHqHCwrvQ97CenptiY

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory Building from Cases: Opportunities and Challenges. In *Source: The Academy of Management Journal* (Vol. 50, Issue 1). https://www.jstor.org/stable/20159839

Fruhwirth, M., Rachinger, M., & Prlja, E. (2020). Discovering Business Models of Data Marketplaces. *53rd Hawaii International Conference on System Sciences* . https://hdl.handle.net/10125/64446

Goudos, S. K., Dallas, P. I., Chatziefthymiou, S., & Kyriazakos, S. (2017). A Survey of IoT Key Enabling and Future Technologies: 5G, Mobile IoT, Sematic Web and Applications. *Wireless Personal Communications*, *97*(2), 1645–1675. https://doi.org/10.1007/s11277-017-4647-8

Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact on JSTOR. *MIS Quarterly* , *37*(2), 337–355. https://www-jstor-org.tudelft.idm.oclc.org/stable/43825912?casa_token=_o7z35cTxaAAAAAA%3ABBAQdcCJyz-6ISuQ6Md2eiBWneJR2lBvA2mVcnr-ldpVDoQo9KjKKw0l29RGG4RdtRKqmKfra88qpOEN7tCrSiPMFoAtyFG_GSFkhwpgaTCZ4No3A6gk&seq=1#metadata_info_tab_contents

Hagiu, A., & Rothman, S. (2016). Network Effects Aren't Enough. *Harvard Business Review*.

Hayashi, T., & Ohsawa, Y. (2020). TEEDA: An interactive platform for matching data providers and users in the data marketplace. *Information (Switzerland)*, *11*(4), 218. https://doi.org/10.3390/INFO11040218

Heisen, E. (2021, May 7). *Straks betalen voor Facebook en Instagram? Dit zijn de mogelijke gevolgen | NU - Het laatste nieuws het eerst op NU.nl*. https://www.nu.nl/tech/6132035/straks-betalen-voor-facebook-en-instagram-dit-zijn-de-mogelijke-gevolgen.html?_ga=2.157171826.874855626.1622797422-580711327.1602490601

Holst, A. (2021, January 20). *IoT connected devices worldwide 2019-2030 | Statista*. https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/

Isaak, J., & Hanna, M. J. (2018). User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. *Computer*, *51*(8), 56–59. https://doi.org/10.1109/MC.2018.3191268

Jøsang, A., Ismail, R., & Boyd, C. (2007). A survey of trust and reputation systems for online service provision. *Decision Support Systems*, *43*(2), 618–644. https://doi.org/10.1016/j.dss.2005.05.019

Keane, J. (2021, April 21). *EU's new sweeping rules for Big Tech could soon look different* . https://www.cnbc.com/2021/04/21/eus-new-sweeping-rules-for-big-tech-could-soon-look-different-.html

Kelly, L. (2021). *Data Marketplaces: Ultimate Guide 2021 | Datarade*. https://about.datarade.ai/data-marketplaces

Koutroumpis, P., Leiponen, A., & Thomas, L. D. W. (2017). *ETLA Working Papers The (Unfulfilled) Potential of Data Marketplaces*. http://pub.etla.fi/ETLA-Working-Papers-53.pdf

Koutroumpis, P., Leiponen, A., & Thomas, L. D. W. (2020). Markets for data. *Industrial and Corporate Change* , *29*(3), 645–660. https://doi.org/10.1093/icc/dtaa002

Lawrenz, S., Sharma, P., & Rausch, A. (2019). *The Significant Role of Metadata for Data Marketplaces*. https://data.iota.org/

Lazaro, C., & Le Métayer, D. (2015). *The control over personal data: True remedy or fairy tale ?*

Lee, J. D., & See, K. A. (2016). Trust in Automation: Designing for Appropriate Reliance:

Http://Dx.Doi.Org.Tudelft.Idm.Oclc.Org/10.1518/Hfes.46.1.50_30392, *46*(1), 50–80. https://doi.org/10.1518/HFES.46.1.50_30392

Liang, F., Yu, W., An, D., Yang, Q., Fu, X., & Zhao, W. (2018). A Survey on Big Data Market: Pricing, Trading and Protection. In *IEEE Access* (Vol. 6, pp. 15132–15154). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2018.2806881

Madsbjerg, S. (2017, November 14). *It's Time to Tax Companies for Using Our Personal Data - The New York Times*. https://www.nytimes.com/2017/11/14/business/dealbook/taxing-companies-for-using-our-personal-data.html

Malgieri, G., & Custers, B. (2018). Pricing privacy – the right to know the value of your personal data. *Computer Law and Security Review*, *34*(2), 289–303. https://doi.org/10.1016/j.clsr.2017.08.006

Manyika, J., Chui, M., Bisson, P., Woetzel, J., Dobsbs, L., Bughin, J., & Aharon, D. (2015). *The Internet of Things: Mapping the value beyond the hype* . www.mckinsey.com/mgi.

Mattioli, M. (2014). Disclosing Big Data. *Minnesota Law Review*, *99*. https://heinonline.org/HOL/Page?handle=hein.journals/mnlr99&id=557&div=&collection=

McCoy, M., & Rahimi, S. (2020). Prediction of Highly Volatile Cryptocurrency Prices Using Social Media. *International Journal of Computational Intelligence and Applications*, *19*(4). https://doi.org/10.1142/S146902682050025X

McManus, R. M., & Rutchick, A. M. (2019). Autonomous Vehicles and the Attribution of Moral Responsibility. *Social Psychological and Personality Science*, *10*(3), 345–352. https://doi.org/10.1177/1948550618755875

Mikkelsen, D., Soller, H., & Strandell-Jansson, M. (2017, January 23). *The EU data-protection regulation-compliance burden or foundation for digitization? | McKinsey*. https://www.mckinsey.com/business-functions/risk/our-insights/the-eu-data-protection-regulation-compliance-burden-or-foundation-for-digitization

Mišura, K., & Žagar, M. (2016). Data marketplace for Internet of Things. *Proceedings of 2016 International Conference on Smart Systems and Technologies, SST 2016*, 255–260. https://doi.org/10.1109/SST.2016.7765669

Nikander, P., & Elo, T. (2019). Will the data markets necessarily fail? A position paper. *30th European Conference of the International Telecommunications Society (ITS): "Towards a Connected and Automated Society."* http://hdl.handle.net/10419/205201www.econstor.eu

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. *The Jounal of Consumer Affairs*, *41*(1), 100–126. https://web-b-ebscohost-com.vu-nl.idm.oclc.org/ehost/pdfviewer/pdfviewer?vid=0&sid=3443102d-ab8d-44c4-98c8-6fcec995c437%40pdc-v-sessmgr01

Nyoni, P., & Velempini, M. (2018). Privacy and user awareness on Facebook. *South African Journal of Science*, *114*(5–6), 1–5. https://doi.org/10.17159/sajs.2018/20170103

Osterwalder, A., & Pigneur, Y. (2010). *Business model generation: a handbook for visionaries, game changers, and challengers* (1st ed.). John Wiley & Sons.

Park, J.-S., Youn, T.-Y., Kim, H.-B., Rhee, K.-H., & Shin, S.-U. (n.d.). *Smart Contract-Based Review System for an IoT Data Marketplace*. https://doi.org/10.3390/s18103577

Park, J. S., Youn, T. Y., Kim, H. Bin, Rhee, K. H., & Shin, S. U. (2018). Smart contract-based review system for an IoT data marketplace. *Sensors (Switzerland)*, *18*(10), 3577. https://doi.org/10.3390/s18103577

Parra-Arnau, J. (2018). Optimized, direct sale of privacy in personal data marketplaces. *Information Sciences*, *424*, 354–384. https://doi.org/10.1016/j.ins.2017.10.009

Pateli, A. G., & Giaglis, G. M. (2004). European Journal of Information Systems A research framework for analysing eBusiness models A research framework for analysing eBusiness models. *European Journal of Information Systems*, *13*, 302–314. https://doi.org/10.1057/palgrave.ejis.3000513

Paul, T. (2020, December 29). *Is web scraping legal ? A guide for 2021*. https://www.linkedin.com/pulse/web-scraping-legal-guide-2021-tony-paul/?trk=read_related_article-card_title

Perera, C., Wakenshaw, S. Y. L., Baarslag, T., Haddadi, H., Bandara, A. K., Mortier, R., Crabtree, A., Ng, I. C. L., McAuley, D., & Crowcroft, J. (2017). Valorising the IoT *Databox* : creating value for everyone. *Transactions on Emerging Telecommunications Technologies*, *28*(1), e3125. https://doi.org/10.1002/ett.3125

Qu, S. Q., & Dumay, J. (2011). The qualitative research interview. In *Qualitative Research in Accounting and Management* (Vol. 8, Issue 3, pp. 238–264). Emerald Group Publishing Ltd. https://doi.org/10.1108/11766091111162070

Ratnasingam, P., Gefen, D., & Pavlou, P. A. (2005). The Role of Facilitating Conditions and Institutional Trust in Electronic Marketplaces. *Journal of Electronic Commerce in Organizations*, *3*(3), 69–82. https://pdfs.semanticscholar.org/4e37/c194604a7e50bf930c45e0235df973127bfb.pdf

Roth, A. E. (2002). The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica*, *70*(4), 1341–1378. https://doi.org/10.1111/1468-0262.00335

Roth, A. E. (2007). What Have We Learned from Market Design? *Economic Journal*, *118*. https://doi.org/10.1162/itgg.2008.3.1.119

Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science*, *33*(2), 163–180. https://doi.org/10.1177/0165551506070706

Sawy, O. A. El, & Pereira, F. (2013). *VISOR: A Unified Framework for Business Modeling in the Evolving Digital Space*. 21–35. https://doi.org/10.1007/978-3-642-31765-1_3

Schomm, F., Stahl, F., & Vossen, G. (2013). Marketplaces for data: An initial survey. *SIGMOD Record*, *42*(1), 15–26. https://doi.org/10.1145/2481528.2481532

Seddon, P., Lewis, G., Seddon, P. B., & Lewis, G. P. (2003). *Association for Information Systems Strategy and Business Models: What' s the Difference? Recommended Citation Strategy and Business Models: What's the Difference 1 ?* 10–13. http://aisel.aisnet.org/pacis2003/17

Shafer, S. M., Smith, H. J., & Linder, J. C. (2005). The power of business models. *Business Horizons*, *48*(3), 199–207. https://doi.org/10.1016/j.bushor.2004.10.014

Shahbazi, K. (n.d.). *How to hit the innovation sweet spot and why it's not all that straightforward - Board of Innovation*. Retrieved August 11, 2021, from https://www.boardofinnovation.com/blog/how-to-hit-the-innovation-sweet-spot/

Shakeri, S., Maccatrozzo, V., Veen, L., Bakhshi, R., Gommans, L., De Laat, C., & Grosso, P. (2019). Modeling and matching digital data marketplace policies. *Proceedings - IEEE 15th International Conference on EScience, EScience 2019*, 570–577. https://doi.org/10.1109/eScience.2019.00078

Sharma, P., Lawrenz, S., & Rausch, A. (2020). Towards Trustworthy and Indepedent Data Marketplaces . *ICBCT20*. https://doi.org/10.1145/3390566.3391687

Shen, Y., Guo, B., Shen, Y., Duan, X., Dong, X., & Zhang, H. (2019). *Pricing Personal Data Based on Information Entropy*. https://doi.org/10.1145/3305160.3305204

Siggelkow, N. (2007). Persuasion with Case Studies. In *Source: The Academy of Management Journal* (Vol. 50, Issue 1). https://www.jstor.org/stable/20159838

Smith, G., Ofe, H. A., & Sandberg, J. (2016). Digital service innovation from open data: exploring the value proposition of an open data marketplace. *Proceedings of the Annual Hawaii International Conference on System Sciences*, *2016-March*, 1277–1286. https://doi.org/10.1109/HICSS.2016.162

Smith, J. (2018a, July 24). *Data Marketplaces: The Holy Grail of our Information Age | Hacker Noon*. https://hackernoon.com/data-marketplaces-the-holy-grail-of-our-information-age-1211a6fec390

Smith, J. (2018b). *The DX Network: Towards Open Data Markets* . https://view.attach.io/Bk~R48Dyf

Spiekermann, M. (2019). Data Marketplaces: Trends and Monetisation of Data Goods . *Intereconomics* , *54*(4), 208–216. https://doi.org/10.1007/s10272-019-0826-z

Stahl, F., Schomm, F., & Vossen, G. (2014a). *The data marketplace survey revisited* . Münster: Westfälische Wilhelms-Universität Münster, European Research Center for Information Systems (ERCIS). www.ercis.orghttp://www.ercis.org/www.ercis.org

Stahl, F., Schomm, F., & Vossen, G. (2014b). Data Marketplaces: An Emerging Species . *DB&IS* , 145–158.

https://books.google.nl/books?hl=nl&lr=&id=aD8oBgAAQBAJ&oi=fnd&pg=PA145&dq=Stahl,+F.,+Schomm,+F.,+%26+Vossen,+G.+(2014a).+Data+Marketplaces:+An+Emerging+Species.++In+DB%26IS+(pp.+145-158).&ots=q2PZTbG2hU&sig=XVqQWDnwyxmuRPBSkl0CXpW_nPo&redir_esc=y#v=onepage&q&f=false

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

Stahl, F., & Vossen, G. (2016). Data quality scores for pricing on data marketplaces. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9621*, 215–224. https://doi.org/10.1007/978-3-662-49381-6_21

Tang, D., Dujii-Hwang, J., Colwill, E., Arora, S., Shah, A., Mendizabal, A., & Callejas, R. (2018). *Value of data - The dawn of the data marketplace* .

Täuscher, K., & Laudien, S. M. (2018). Understanding platform business models: A mixed methods study of marketplaces. *European Management Journal*, *36*(3), 319–329. https://doi.org/10.1016/j.emj.2017.06.005

van de Ven, M. (2020). *Creating A Taxonomy of Business Models for Data Marketplaces*. Technische Universiteit Delft.

van Niekerk, M., & van der Veer, R. (2018). *DataBroker DAO*. https://www.allcryptowhitepapers.com/wp-content/uploads/2018/11/Databroker-DAO.pdf

van Ooijen, I., & Vrabec, H. U. (2019). Does the GDPR Enhance Consumers' Control over Personal Data? An Analysis from a Behavioural Perspective. *Journal of Consumer Policy*, *42*(1), 91–107. https://doi.org/10.1007/s10603-018-9399-7

Venkatesh, V., & Bala, H. (2008). Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences*, *39*(2), 273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

VETRI. (2017). *VETRI: Value you data* . https://vetri.global/wp-content/themes/vetri/documents/whitepaper.pdf

Wang, D., Chen, D., Song, B., Guizani, N., Yu, X., & Du, X. (2018). From IoT to 5G I-IoT: The Next Generation IoT-Based Intelligent Algorithms and 5G Technologies. *IEEE Communications Magazine*, *56*(10), 114–120. https://doi.org/10.1109/mcom.2018.1701310

Wat zijn stablecoins? | Binance Academy. (n.d.). Retrieved June 9, 2021, from https://academy.binance.com/nl/articles/what-are-stablecoins

Waterman, T. (2020, January 29). *Web scraping is now legal. Here's what that means for Data Scientists*. https://medium.com/@tjwaterman99/web-scraping-is-now-legal-6bf0e5730a78

Yoffie, D. B., Gawer, A., & Cusumano, M. A. (2019). *A Study of More Than 250 Platforms Reveals Why Most Fail*. Harvard Business Review. https://hbr-org.tudelft.idm.oclc.org/2019/05/a-study-of-more-than-250-platforms-reveals-why-most-fail

Zaretsky, R. (2021, March 3). *Big Tech Has Our Data. Should We Tax Them on It? | Tax Policy Center*. https://www.taxpolicycenter.org/taxvox/big-tech-has-our-data-should-we-tax-them-it

Zuiderwijk, A., Loukis, E., Alexopoulos, C., Janssen, M., & Jefferey, K. (2014). Elements for the Development of an Open Data Marketplace. In *Conference for E-Democracy and Open Government* (pp. 309–322). www.donau-uni.ac.at

Appendix A: Used academic literature

A.1 Overview of the articles

AUTHOR (YEAR)	TITLE	AIM OF STUDY	CITATIONS
(FRUHWIRTH ET AL., 2020)	**"Discovering Business Model of Data Marketplaces"**	Development of taxonomy for data marketplaces	7
(TÄUSCHER & LAUDIEN, 2018)	**"Understanding platform business models: A mixed methods study of marketplaces"**	Development of taxonomy for digital platforms	248
(LAWRENZ ET AL., 2019)	**"The Significant Role of Metadata for Data Marketplaces"**	Outlines the significance of metadata for data trading and data quality	6
(SHARMA ET AL., 2020)	**"Towards Trustworthy and Independent Data Marketplaces"**	Proposal for a data sharing platform	1
(AZCOITIA & LAOUTARIS, 2020)	**"Try Before You Buy: A practical data purchasing algorithm for real-world data marketplaces"**	Example of pricing mechanism	-
(KOUTROUMPIS ET AL., 2017)	**"The (Unfulfilled) Potential of Data Marketplaces"**	Requirements for data marketplaces	30
(SPIEKERMANN, 2019)	**"Data Marketplaces: Trends and Monetisation of Data Goods"**	Identification of data marketplace trends	22
(HAYASHI & OHSAWA, 2020)	**"TEEDA: An Interactive Platform for Matching Data Providers and Users in the Data Marketplace"**	Description of a matching mechanism	3
(STAHL ET AL., 2016)	**"A classification framework for data marketplaces**	Development of classification framework	48
(PERERA ET AL., 2017)	**"Valorising the IoT *Databox* : creating value for everyone."**	Focus on data control	38
(G. SMITH ET AL., 2016)	**"Digital Service Innovation from Open Data: Exploring the Value Proposition of an Open Data Marketplace"**	Focus on open data	35
(SCHOMM ET AL., 2013)	**"Marketplaces for Data: An Initial Survey"**	Classification of data providers	98
(STAHL ET AL., 2014)	**"The data marketplace survey revisited"**	Continuation of classification	21
(PARRA-ARNAU, 2018)	**"Optimized, direct sale of privacy in personal data marketplaces"**	Focus on personal data marketplaces	15
(MIŠURA & ŽAGAR, 2016)	**"Data Marketplace for Internet of Things"**	Focus on IoT data marketplaces	37
(NIKANDER & ELO, 2019)	**"Will the data markets necessarily fail?: A position paper"**	Focus on fundamental economic constructs	-

A.2 Overview of the obstacle categorization

Obstacle category	Colour indication
Trust	Light blue
Data pricing	Yellow
Product description & quality	Purple
Privacy and control	Dark blue
Governance mechanisms	Green
Matching mechanism	Brown
Other	Dark green

AUTHOR (YEAR)	TITLE	POTENTIAL OBSTACLES MENTIONED
(FRUHWIRTH ET AL., 2020)	"Discovering Business Model of Data Marketplaces"	-Selling data might harm own business interests -No established rules for pricing datasets (Arrow paradox)
(TÄUSCHER & LAUDIEN, 2018)	"Understanding platform business models: A mixed methods study of marketplaces"	Marketplace functions: - Creating trust - Allowing for acceptable price discovery
(LAWRENZ ET AL., 2019)	"The Significant Role of Metadata for Data Marketplaces"	-Generating metadata for datasets is difficult: product description is a challenge
(SHARMA ET AL., 2020)	"Towards Trustworthy and Independent Data Marketplaces"	-Main problem towards the data economy is the lack of trust -Bias towards data providers or data buyers -Difference between physical product trading and data trading -Data requirements cannot be known upfront
(AZCOITIA & LAOUTARIS, 2020)	"Try Before You Buy: A practical data purchasing algorithm for real-world data marketplaces"	-Recognizing the value of certain datasets
(KOUTROUMPIS ET AL., 2017)	"The (Unfulfilled) Potential of Data Marketplaces"	-Data providers are not aware of the quality and legal status of the dataset they are selling -Data in databases cannot be protected by copyright
(SPIEKERMANN, 2019)	"Data Marketplaces: Trends and Monetisation of Data Goods"	-Selling data might harm own business interests -Low willingness to buy data (Arrow paradox) -Absence of legal frameworks: no legal certainty for trading data
(HAYASHI & OHSAWA, 2020)	"TEEDA: An Interactive Platform for Matching Data Providers and Users in the Data Marketplace"	- Insufficient means for data providers to learn what kind of data is desired by the data buyers
(STAHL ET AL., 2016)	"A classification framework for data marketplaces	-Data should not end up for trade on the black market -Difference in buyer requirements: constant standard vs. high individuality
(PERERA ET AL., 2017)	"Valorising the IoT *Databox* : creating value for everyone."	-Users do not have full control over their data
(G. SMITH ET AL., 2016)	"Digital Service Innovation from Open Data: Exploring the Value Proposition of an Open Data Marketplace"	-Legislation is geographically dependent -Types of obstacles: institutional, task complexity, use and participation, legislation, information quality, and technical details
(SCHOMM ET AL., 2013)	"Marketplaces for Data: An Initial Survey"	-Sold data should be of high quality to make correct decisions
(PARRA-ARNAU, 2018)	"Optimized, direct sale of privacy in personal data marketplaces"	-Privacy is related to trust: soft privacy vs. hard privacy -Difficult to establish what a good deal is in terms of monetary value and privacy
(MIŠURA & ŽAGAR, 2016)	"Data Marketplace for Internet of Things"	-Malicious users can misuse the sale of data -Users should understand what data they are selling and to whom they are selling it -Trade-off between monetary value and privacy
(NIKANDER & ELO, 2019)	"Will the data markets necessarily fail?: A position paper"	-Taking place in a data marketplace is perceived as risk because competitive advantage might be lost -Users do not know where their data might end up -Monetary value depends on the context in which the data is used

Appendix B: VETRI

B.1 Interviews preparations and resulting BM Stress Tests

INTRO:

**First of all, I read the information on your website and in your white paper and I think it is very cool what you are trying to achieve. Just for my personal understanding: you aim to give users full control over their personal data. However, the Facebook or Google data that is also data import for the VETRI wallet app, also still goes to FB & Google at the moment right? So to achieve full ownership and control such parties would have to become partners?*

**My research focuses on three kinds of marketplaces (IoT, B2B, and personal) and also the differences in obstacles they experience. So far, I talked to the IoT and B2B marketplaces and they believe that the personal data marketplace is the hardest to realize because of the sensitivity of the data being traded. How do you see this?*
**Do you think there are other obstacles for personal data marketplaces than for an IoT or B2B data marketplace?*
**VETRI is designed as a not-for-profit platform, unlike other data marketplaces. Do you think this aspect influences if a business model for a personal data trading environment is viable or not?*

GENERAL QUESTIONS:

1. Trust

**Trust is often mentioned as the main obstacle for data marketplaces. The website describes that you use blockchain verify rather than trust. Could you describe if you still experience any issues regarding trust?*
→ Could you describe how you address this in your business model?

2. Privacy and control

**Obstacles that are often named for data trading platforms, in general, are privacy and control. But, you state that you can also see it as an opportunity: if you reach out in a privacy-respecting way to customers, you are likely to build deeper trust. Are there still any implications of privacy that might limit your development?*

3. Product description & product quality

**For data marketplaces in general, metadata is needed to describe a product, but can be hard to derive and quality is hard to define because it the on the buyer's requirements. Is it correct to assume that as you sell a profile rather than a raw dataset, this is less of a problem?*
→ Can you now guarantee the product quality?

4. Matching data sellers and data buyers

**Are you in any way involved in matching the data sellers and data buyers?*

5. Pricing

**It is often suggested that data pricing is a key challenge (as it is difficult to estimate upfront what a dataset is worth), and that is why it is difficult to make a business model viable. You state in your whitepaper that the price of the data is based on the desirability of the shared data as perceived by the data consumers. Does this mean that the data user does not have any indication of the value it gets for his/her data?*

→ Is that perceived as an issue for the users?
**How does the data consumer pay for the data? Is it pay-per-use or package-based pricing?*

6. Governance mechanisms

** It has been said that there is an absence of legal/regulatory frameworks for trading data. How is this an obstacle for developing a personal data marketplace like VETRI?*

7. Other

**Do you expect any other obstacles for VETRI and how do you address this in your business model?*

B.2 VETRI BM Stress Test

See next pages.

		Trust		**Privacy and control**		**Product description and quality**	**Matching providers & buyers**	**Pricing mechanism**	**Governance mechanisms**	**Other**			
		Lack of trust in the system and/or its participants	*Harm of own business interests*	*Privacy*	*Control*					*Network effects*	*Incentivizing to participate*	*Competitors*	*Technical issues*
Service domain	**VETRI: Personal Data Marketplace**												
Value proposition	All services in a single platform	Vision: full transfer of personal data to VETRI wallet	Business interests are not harmed, but shifted. Normally you would buy data from big tech companies.		Working with a duplicate system				Regulations should help development	Market strategy is to get millions of users for it to succeed		Big-tech companies (e.g. Google & Facebook) have the means to destroy a personal data marketplace	
Enterprise data marketplace	No												
Data processing and/or analytics tools	No												
Marketplace participants	C2B	Customers need eduction on the technology used		Assumption: some users will not care about privacy, control and the technology used to ensure that at first: they get paid more using VETRI.		A specific user profile with certain attributes and the ability to directly ask for certain data rather than acquiring a raw data set	Currently using data aggregators; you cannot attract both sides at one, chicken-and-egg problem	Users earn more because intermediaries are removed; data consumers market own product					
Industry domain	Health & personal data			Higher sensivity than other data; needs to be extended from current survey practices	Little control because big-techs aggregate personal data in return for their services				Different regulations for personal data than other data categories (e.g. GDPR), data taxes)		Everybody has personal data; currently there is no true alternative to "just not using Facebook"	Data aggregators & big-tech companies are also competing for your data	
Geographic scope	Regional								Need for global scope also results in different regulations across the globe	Scope should be big to fight against big tech companies			
Time frame	Up-to-date					Phone sends notification when request matches your attributes					Participating the monetizing part is not required; you choose whether you want to		
Technology domain													
Platform architecture	Decentralized	Less trust should be required in the operating party, but people don't understand blockchain today		VETRI does not have to see the data to allow the marketplace to work.			Matching is possible thanks to blockchain technology						
Data access	Specialized software			Use of a digital identity									Hacks by users
Data source	Multiple sources												Compiling and organizing the data

Organization domain													
Matching mechanism	One-to-many												
Platform sponsor	Private	Foundation ; no private interest											
Finance domain													
Revenue model	Asset sales (not-for-profit)	No other interest than serving the users						Ability to pay more: no intermediaries + no bots, revenu comes from token (VLD) sale	Regulations will help instead of couteract		The non-profit base ensures that a TRUE alternative is offered		Little money avaiable; breaks from growth because new funds have to be acquired
Pricing model	Package based pricing (future)												
Price discovery	Set by buyers		Marketing trick: Data consumer gives user more money to buy more of its own product		Monetization is only to give value to data, not control over data.		Users have no direct insight on what their data is worth, but can choose whether to participate or not for a set price.	Data consumers (buyers) will set prices by trial-and-error. They will see whether a user accepts or not.			It is likely that users can be paid up to 50% more due to the use of digital identities (no fake data) and the removal of intermediaries		
Smart contract	Yes	Less trust should be required in the operating party, but people don't understand the technology today											
Payment currency	Cryptocurrency										Allows for other ways of revenue than taking it from the users		
	No direct relevance												
	Possible showstopper												
	Requires attention												
	No direct negative effects/positive effects												

Appendix C: the DX Network

C.1: Interview preparations

INTRO:

I am trying to understand the obstacles that data marketplace have to face and how these can be overcome. As the DX network is no longer operating I wondered if I could first ask some questions about this.

**Could you describe why the DX network is no longer operating? *If I am correct, you are now still part of startuptracker.io, which combines the information from providers with data from start-ups. The DX network was also focused on start-up data. Could you describe the difference in the ambition of the two platforms?*
**In the white paper you mention the possible incremental and radical innovations that the DX network could enable. What innovation were you able to achieve before you shut down?*
**On the website it is also described that there are three main types of data marketplaces: B2B, IoT, and personal. I also use this classification in my research. Do you think there is a difference in the obstacles that they face?*

GENERAL QUESTIONS:

1. **Trust:**

**the DX Network delegated the aspects of trust to the data consumers, while you facilitated the whitelist and blacklist mechanism. Did this fully remove the obstacle of trust for the DX network?*

2. **Privacy and control:**

**It is stated in the white paper that there is a great difference between trading personal data and trading company data in relation to privacy. Privacy can also be related to the unwillingness of organizations to share their data because they are afraid that competitors will harm their business interests (maybe especially if you are a start-up). Could you describe how you experienced this?*

3. **Product description & product quality:**

**It is often suggested that describing a dataset is hard, and therefore it is also hard to guarantee the quality of a data product. Could you describe if you see this as an obstacle?*
How did you address this in your business model?

4. **Matching data sellers and data buyers:**

**Data consumers could purchase data across all publishers with a single API request. Did you also have a role in matching the data sellers and data buyers?*

5. **Pricing:**

**If I understand correctly the pricing mechanism of the data on the platform was based on the number of times a certain data point was queried. One of the main difficulties of pricing data is setting a price for it, as unliked physical assets it can be duplicated for free and its utility is hard to measure in practice. How did you experience this with your pricing mechanism?*

6. **Governance mechanisms**

**"Could you describe if you experience any obstacles concerning legal data trading frameworks?" "If yes: How do you address this type of problem?"*

7. **Other**

**"Could you describe if there are any other obstacles for data marketplaces and how you tried to address this in your business model?"*

C.2: The DX Network BM Stress Test

See next pages.

		Trust		Privacy and control		Product description and quality	Matching providers & buyers	Pricing mechanism	Governance mechanisms	Other			
		Lack of trust in the system and/or its participants	*Harm of own business interests*	*Privacy*	*Control*					*Network effects*	*Incentivizing to participate*	*Competitors*	*Technical issues*
Service domain	**the DX Network: B2B data marketplace**												
Value proposition	Easy data access and/or tooling	Establishing a reputation, a track record, as organization				Whitelist and blacklist mechanism. (like a Ebay review system)	Performing queries is almost natural language by using SPARQL.	DX tokens had to be purchased to participate on the platform. The value of the coin was not stable.	The DX Network provides a line of defense against web scraping (of open platforms). It is often hard to legally protect yourself against that.		Promote open access to information, while receiving a fair share of the wealth that is contributed to by sharing		Data marketplace should be focussed on the user experience. Not the technology on the backside.
Enterprise data marketplace	No												
Data processing and/or analytics tools	No												
Marketplace participants	B2B	Trust in human or business sense: nobody wants to take the first step in trusting a business.											
Industry domain	Finance & Alternative data	For B2B a lot of data is already public, just not structured.	Some data might not be shared because there is a competitive advantage in having it. But if the price is high enough, it is worth to share.		In B2B datasets, the data goes obsolete.				Licensing terms (set by data provider)				
Geographic scope	Global									Ability to query many databases at once, but perceiving it as one big database due to single API structure			
Time frame	(Near) real-time												
Technology domain													
Platform architecture	Decentralized	Trust in technical sense: you can probably trust a blockchain structure and audited smart contract											Technical (command line) interface: complicated infrastructrue led to slow adoption
Data access	API					Specified data model: common format, access via a single API to all aggregated data in this format							
Data source	Customer provided data						The DX network just provided the infrastructure to make exchange possible. What the providers put there was what was available. Queries had to be specific and targeted						

Organization domain													
Matching mechanism	Many-to-many									Share structured data with whoever is interested			
Platform sponsor	Independent												
Finance domain													
Revenue model	Asset sale												
Pricing model	Pay-per-use	Data collector (provider) pays a listing fee per datapoint to discourage malicious submissions									Data collectors (providers) receives a query fee for each time the datapoint is queried.		
Price discovery	Set by external sellers							The sellers could set the price, but because they got paid in DX tokens the value was very variable. This was perceived as an obstacle.					
Smart contract	Yes	Trust in technical sense: you can probably trust a blockchain structure and audited smart contract											
Payment currency	Cryptocurrency												Free floating currency: DX token. The buyers of the token were not the users of the marketplace. Purchasing tokens is "difficult".
	No direct relevance												
	Possible showstopper												
	Requires attention												
	No direct negative effects/positive effects												

Appendix D: Databroker (DAO)

D.1: Interview preparations

INTRO:
I am trying to understand the obstacles that data marketplace have to face and how these can be overcome. As Databroker DAO was for IoT data only and extended its scope to Databroker, I hoped I could first ask some questions on the reasoning behind this.
**Could you describe why you decided to extend your scope from only IoT data to all kinds of data?*
**Do you think there are other obstacles for IoT marketplaces only than there are for all kinds of data marketplaces?*

GENERAL QUESTIONS:

1. **Trust**:
**Trust is often mentioned as the main obstacle for data marketplaces. The website describes that you use blockchain verify rather than trust. Could you describe if you still experience any issues regarding trust?*

**Trust can also be related to the unwillingness of organizations to share their data because they are afraid that competitors will harm their business interests. Could you describe how you see this?*

2. **Privacy and control:**
**"It is often suggested that privacy implications of trading data might limit development. How do you experience this? Could you describe how you address this in your business model?"*

3. **Product description** & **product quality:**
**"I read that because of the unique characteristics of data, in comparison to physical products, it is hard to give an accurate description that reflects the use of the dataset for the data buyer. Databroker lets data providers enter the description manually. How do you experience using this method for product description purposes?"*

**"Is it possible to guarantee the quality of a dataset for a data buyer?"*

4. **Matching data sellers and data buyers:**
**"It was described in the literature that there are insufficient means to discover what data is required by data buyers. You provide a DataMatch service. How do you experience the use of this service?"*

5. **Pricing:**
**"Could you describe if you experience pricing of datasets as an obstacle?"*
**" Why set the price in EUR and then convert it to a digital currency?"*

6. **Governance mechanisms**
**"Could you describe if you experience any obstacles concerning legal data trading frameworks?" "If yes: How do you address this type of problem?"*

7. **Other**
**"Could you describe if there are any other obstacles for data marketplaces and how you address this in your business model?"*

D.2: Databroker (DAO) BM Stress Test

See next pages.

		Trust		Privacy and control		Product description and quality	Matching providers & buyers	Pricing mechanism	Governance mechanisms	Other			
		Lack of trust in the system and/or its participants	*Harm of own business interests*	*Privacy*	*Control*					*Network effects (chicken-and-egg)*	*Incentivizing to participate*	*Competitors*	*Technical issues*
Service domain	**Databroker (DAO): IoT/B2B**												
Value proposition	Easy data access and/or tooling			Databroker is just a mediator between providors and buyers and not directly responsible for the privacy of the users.			Matching is done manually. The platform does not have enough users to automate it yet.			Start with the data buyers. Assist them in finding the right data provider. When a data provider made a sale, he might stay involved in the platform.	80% of the data that is generated within a company is not used at all. This can generate revenue.		
Enterprise data marketplace	Yes		Many data exchange use cases are better suited to private marketplaces. Example on Databroker: exchange of safety studies on the chemicals used in consumer products.										
Data processing and/or analytics tools	No												
Marketplace participants	B2B				Still fear that data is leaked to other 3rd parties. How to keep control over the data in case of fraude?				Who is responsible if something goes wrong in decisions that are made based on the data? : A company will be legally blocked. Conformance is really important for data sharing.				
Industry domain	Any data (first: sensor data)												Scaled up from IoT to any data because IoT technology was not mature enough yet (energy consumption was too high) . The 5G network might change that in the future. For now, B2B is easier to convert into a business model.
Geographic scope	Global												
Time frame	Multiple												
Technology domain													
Platform architecture	Decentralized	Technologically speaking there is trust: participants can control with whom and how data is shared and can stop at any moment.										Many marketplaces have a centralized structure. Only around 10 worldwide are decentralized. Centralized alternatives are assumed to be outcompeted on security, compliance, transparency, neutrality, automation & scalability.	
Data access	Multiple options												
Data source	Customer provided data			Databroker cannot see the content of the data shared, but can ban users if they are not complaint to the terms and conditions.				Currently there not enough marketplace deals to automate pricing mechanisms. Maybe in 10 years this will be valuable.					

Organization domain													
Matching mechanism	Many-to-many												
Platform sponsor	Independent												
Finance domain													
Revenue model	Commissions												
Pricing model	Pay-per-use												
Price discovery	Negotiation/set by external sellers							Complex features of data complicate pricing because data is: non-linear, non-rivalrous, a potential liability and its value increases through refinement, with holder, and over time. This should be considered in price discovery. It is difficult to have a discrete, specified price. E.g. what variables should be included to calculate this?					
Smart contract	Yes					The quality of the data cannot be guaranteed, but thanks to smart contracts the buyer can still have a warranty on the data. The payment is freezed when a complaint is filed about the data provider.							
Payment currency	Fiat money (but crypto for storage)										Buyers can use debit or credit card to pay. There is no need to buy DTX tokens first.		By creating a cryptocurrency the project can be funded by an ICO (initial coin offering). It also makes it easier to

	No direct relevance
	Possible showstopper
	Requires attention
	No direct negative effects/positive effects

Appendix E: Experts interview preparation (TRUSTS group)

INTRODUCTION:

Explain the methods and purpose of my research
*3 case studies (B2B, IoT → B2B, Personal)
*Business Model Stress Test against obstacles identified from literature – see the relationships

"I was hoping you could give your insights on how the negative relationships might be solved. Independently by the data marketplaces, or maybe by the TRUSTS initiative that is overarching on these marketplaces."

GENERAL QUESTIONS:

(1) Trust

Literature often states that a certain technology, e.g. blockchain, smart contracts, will take away the obstacle of trust. However, it was also noted that there is a difference between technological trust and institutional trust. This translates to two questions:

-"*How will users trust and thus adopt a data marketplace, not considering any technological solutions?"*
-*"Will users trust technological solutions without understanding how the technology works? Probably the greatest part of the population does not know anything about blockchain or cryptocurrencies.*"

(2) Privacy and control

The biggest issue for privacy and control were for personal data marketplaces, as they cannot live up yet to their value proposition: monetizing personal data and giving control over it. This is because they still have to work with duplicate systems, as big-techs such as Facebook and Google also aggregate the personal data. I think TRUSTS also wants to include the management of personal data.

-*"How do you see this competition with big-techs in relation to what personal data marketplaces are trying to achieve?"*

Another issue for B2B business models was the fear of fraud and leaks to third parties.
-*"How do you think you can deal with that?*"

(3) Matching

Matching data providers and data buyers was not yet really a problem, but mainly because the marketplaces had a limited amount of participants.
-"*What is TRUSTS its strategy for attracting these participants?"*

(4) Pricing mechanism

Another major issue because of the characteristics of data is data pricing. Either because the variables to calculate a price are unknown, there are not enough deals yet to make any kind of prediction, or because when using cryptocurrency the value of a coin is unstable.
-"*What is your idea on data pricing? Would there need to be a uniform standard?*"

A lot of data marketplaces develop their own token because then they can fund their project selling the token. One of the data marketplaces I studied that shut down said that if they could start over they would use stable coins, so not their own currency.

-"*How do you think that the payment currency a marketplace chooses is related to them having a viable business model or not?*"

(5) Governance
In general, it seemed that the conclusion from the case studies was that implementing regulations should help the development of data marketplaces. For example, 'data taxes' were named as an opportunity, since that would make companies pay for the personal data they possess and would therefore rather have a data marketplace handle it.

-"*Are there any other regulations upcoming that could stimulate the development? Or should there be?"*
-*"If there is a global scope, you might have to deal with different regulations. Is that also the case for TRUSTS?*"

It was also mentioned that an issue could be that it is unclear who has the responsibility for data-driven decisions. Just like with for example self-driving cars.
-"*Do you see this as a problem for the development of data marketplaces?"*

(6) Other
"In your TRUSTS work do you have encountered any other obstacles and have you thought of solutions for those obstacles?"