Creating a Taxonomy of Business Models for Data Marketplaces

Master Thesis Report
MSc Complex Systems Engineering & Management

M.R. van de Ven
2020

TU Delft

Cover page: © Unsplash photos

Creating a taxonomy of business models for data marketplaces

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Complex Systems Engineering & Management**

Faculty of Technology, Policy and Management

by

Montijn van de Ven

Student number: 4351231

To be defended in public on September 4, 2020

Graduation committee
First Supervisor : Dr.ir. G.A. de Reuver, Information and Communication Technology
Second Supervisor : Dr.ir. Z. Roosenboom-Kwee, Economics of Technology and Innovation
Advisor : A.E. Abbas MSc, Information and Communication Technology

Preface

Dear reader,

In front of you lies my Master's thesis, the final product of six months of hard work. During this period of time, I have learned a lot about conducting scientific research, data marketplaces and business models. Despite the unprecedented circumstances in which I had to write my thesis, I really enjoyed the research process. This was partly due to the topic that I have found very interesting from the start of my research, but also due to the knowledge that I gained about myself and my interests during the process. In this preface, I would like to thank a number of people who have supported me during my research and throughout my entire studies at the TU Delft.

Firstly, I would like to thank Mark de Reuver, who has acted as the first supervisor of my graduation project. Mark has been a very devoted mentor who was always willing to schedule an online meeting or provide me with quick feedback. I am very thankful for how Mark guided me through the research process, helping me to deal with research challenges and pushing me to take my research to a higher level. I am also very happy to have been part of the TRUSTS team of the TU Delft. It was interesting to take part in the international meetings and to experience how a major European research project takes off and how progress is being made. Secondly, I would like to thank my second supervisor Zenlin Roosenboom-Kwee for her positivity and valuable feedback. After the meetings with Zenlin, I always felt happier about my research and I went back to my work with more confidence. Lastly, I would like to thank Antragama Abbas, who acted as an advisor during my graduation project. I really enjoyed our discussions and I am very thankful for our meetings during and sometimes outside of working hours. Your substantial feedback and critical comments have been very helpful to improve my thesis to the last detail.

Last but not least, I would like to thank my family who have supported me and encouraged me throughout this intensive period of hard work and throughout my entire studies in Delft. Thank you for always being there for me and providing mental support. I could not have done this without you. I would also like to thank my friends who have helped me stay motivated and supported me throughout the intelligent lockdown.

Montijn van de Ven

Delft University of Technology

Executive summary

For the past few years, the amount of available data has increasingly been growing. Because the data can be processed into comprehensible information, the large amounts of data have become an important source for innovation and economic growth for businesses and society in general. The rise of the Internet-of-Things and development of advanced data analytics techniques have made it easier for organizations to collect and to analyse data (Hürtgen & Mohr, 2018). Since newly collected and processed data can either be used internally or traded with external parties, data has involved into a strategic asset for firms (Opher et al., 2016). As a result of this, data economies have started to evolve. The data economy functions as a digital ecosystem where a network of data buyers, sellers and service providers come together to exchange data (European Commission, 2017). Data marketplaces can fulfil a key role in realizing the data economy. As an organization may not always possess the required data to carry out or improve their processes and services, they may wish to purchase this data from other organizations. A data marketplace can address this issue by providing a digital platform through which individuals and organizations can exchange data (Stahl et al., 2016; Schomm et al., 2013).

Despite the potential benefits of data marketplaces, in practice very little data is shared or traded via digital platforms (Koutroumpis et al., 2017). Only a small number of data marketplaces exists currently, and many data marketplaces that have been set up have failed or are shut down (Koutroumpis et al., 2017). Furthermore, it is found that data marketplaces adopt limited business models, focussed on hierarchical organization settings and bilateral trading arrangements (Koutroumpis et al., 2017; Stahl et al., 2017). A business model is a description of how a network of organizations creates and captures value (Bouwman et al., 2008). In general, little research has been conducted on data marketplaces (Thomas & Leiponen, 2016) and data marketplace business models in particular (Fruhwirth et al., 2020; Spiekermann, 2019).

In the literature, there are two taxonomies available that provide an overview of data marketplace business models (Fruhwirth et al., 2020; Spiekermann, 2019). A taxonomy is a classification scheme of a certain research object, from which wider generalizations are made (Lambert, 2015). However, the existing taxonomies are lacking in some areas that this study aims to improve. Firstly, the two studies mainly focus on multilateral data marketplaces, in which the data marketplace functions as neutral intermediary that matches multiple data buyers with multiple data sellers (Koutroumpis et al., 2017). However, in practice data is rarely traded via multilateral data marketplaces, and instead data trading often happens via bilaterally negotiated contracts (Koutroumpis et al., 2017). Secondly, the two existing taxonomies are structured using business model ontologies that view the business model from a single firm perspective (Remane et al., 2017; Gassman et al., 2014; Teece, 2010; Al-Debei et al., 2008). However, studies show that data marketplaces take part in a network of stakeholders, among which data analysts, application vendors, algorithm developers, data providers, consultants, licensing entities, platform providers (Spiekermann, 2019; Chakrabarti et al., 2018; Thomas & Leiponen, 2016; Muschalle et al., 2012). Thirdly, the data economy is emerging and more data marketplaces are being set up, and therefore new business model alterations may have been produced in practice that were not considered during the development of the existing taxonomies.

This study aimed to go beyond the state of the art by developing a taxonomy from a multi-stakeholder perspective on business models. The term data marketplace was broadly interpreted in this research, to also allow for the inclusion of atypical types of data marketplaces. New business model alterations that have been produced in practice were considered. The main research question of this research was: *How can the business model characteristics of different types of data marketplaces be classified into a taxonomy from a multi-stakeholder perspective?*

To develop the taxonomy a design science approach was adopted (Hevner, 2007) and a standard taxonomy development method by Nickerson et al. (2013) was employed. During the taxonomy development, both existing scientific theories and frameworks were considered, as well as concepts and knowledge from industry and practice. First, a literature review was conducted to discover existing business model characteristics in the academic literature. Furthermore, a database of existing data marketplaces was created based on desk research for empirical cases. The desk research process resulted in a final set of 178 existing data marketplaces. From the database, a set of 40 cases was sampled for further consideration of their business model characteristics. Because 60% of the cases in the database were data marketplaces active in the audience data industry domain, the cases were first divided into seven groups based on the type of data that was traded on the marketplace. Audience data is combined data about a certain target group of customers, the 'audience', that is often gathered by marketeers, to target the envisioned audience with highly personalized and relevant offers. Subsequently, a random sample was taken from the groups. In the final sample of 40 data marketplaces, the number of audience data marketplaces was drastically decreased to only 22,5% of the sample, to still account for the size of this category, but to not let this type of data marketplaces dominate the analysis.

To start the taxonomy development, first the meta-characteristics of the taxonomy were defined. The meta-characteristics function as overarching characteristic that provides the foundation for choosing further characteristics for the object of interest (Nickerson et al., 2013). Four domains that best describe multi-stakeholder business models were selected as the meta-characteristics of data marketplace business models: the Service domain, Technology domain, Organization domain and Finance domain (Bouwman et al., 2008). After specifying the meta-characteristics, the ending conditions of the taxonomy development process were defined. The eight ending conditions that were prescribed by the authors of the taxonomy development method were employed (Nickerson et al., 2013).

The design process of the taxonomy took off with the development of a conceptual framework, based on the data marketplace business model concepts found in the existing literature. Using this framework, the scientific concepts were checked with the business model characteristics of the set of 40 cases of existing data marketplaces. To retrieve information about the business models of the existing companies, websites and news articles were analysed to get to gather information about the business models of the respective data marketplaces. The discovered fragments of information were then analysed and compared with the concepts in the conceptual framework. If necessary, the framework was revised, and dimensions were merged or split. After that, the business models of the existing data marketplaces were considered again, to see whether new dimensions or characteristics could be added to the taxonomy. After every iteration, the pre-specified ending conditions were checked to see if the

design process could be ended. In total, four design iterations were conducted in which the taxonomy was revised and new dimensions and characteristics were added.

The final taxonomy comprises of four meta-dimensions, 17 business model dimensions and 59 business model characteristics of data marketplaces. The four meta-dimensions are the Service domain, Technology domain, Organization domain and Finance domain. The 17 identified business model dimensions are: the value proposition, enterprise data marketplace, data processing and analytics tools, marketplace participants, industry domain, geographic scope and time frame in the Service domain; the platform architecture, data access and data source in the Technology domain; the matching mechanism and platform sponsor in the Organization domain; and the revenue model, pricing model, price discovery, smart contract and payment currency in the Finance domain. The taxonomy was demonstrated on the basis of an empirical illustration of the use of the taxonomy on three cases of data marketplaces, Wibson, QueXopa and Advaneo. The three cases of data marketplaces varied in terms of their marketplace design, to showcase that the taxonomy is suitable to classify multiple types of data marketplaces. The demonstration illustrated that most of the business model characteristics of the three selected data marketplaces could be classified with the use of the taxonomy. A key take away from the demonstration is the taxonomy is useful to classify the business models of data marketplace when sufficient information is available about the respective business model characteristics of the companies.

This research makes a number of scientific contributions. Firstly, the results of this study improve the knowledge about data marketplace business models. And secondly, this study furthers the understanding of data marketplaces, by taking on a broad perspective on this type of marketplaces and by providing a new definition. Limitations of the research are that the information about data marketplace business models of the existing cases may be prone to the subjectivity of the researcher. Furthermore, some existing data marketplaces may have been missed when constituting the database of existing cases. Moreover, not all data marketplace companies provided sufficient information about their business model, and therefore not all business models of the existing data marketplaces could be classified on the taxonomy. Given the scientific contributions and the limitations of this research, a number of recommendations for further research are provided. Researchers may utilize the developed taxonomy to derive business model patterns and business model archetypes of data marketplaces, and in-depth case studies may be conducted on specific data marketplaces or in certain industry domains. Furthermore, future research may focus on providing a more concise definition of a data marketplace, by considering the various terms used in practice. Through interviews with relevant experts, the taxonomy may be validated, and new dimensions and characteristics may be added.

This research is relevant to managers and society, because managers can use the taxonomy during the exploration of setting up a data marketplace and designing its business model. Furthermore, the research results may raise awareness about data trading among consumers, and it provides insights in the emergence of harvesting data marketplaces, which may enable consumers to monetize their health and personal data.

Contents

List of figures

List of tables

1. Introduction

1.1. Problem identification

In recent years, the amount of available and generated data has staggered. DOMO (2019) estimated that by 2020 we will have generated 40 times more data bytes than the number of stars in the observable universe. As data can be processed into comprehensible information, the vast amounts of data have become an important resource for innovation and economic growth for businesses and society in general. Organizations have long been using data as input for decision-making and process optimization. However, the deployment of connected devices on the Internet of Things and advances in data analytics have made it easier for organizations to collect and analyse data (Hürtgen & Mohr, 2018). Because the newly collected and processed data can be monetized and traded, it can pose as a strategic asset to organizations (Opher et al., 2016). The emergence of data as a strategic asset for organizations lays the foundation for two important developments in the conduct of business: it enables firms to change from their conventional material-based business models to new data-driven business models, and the collected data can be sold to other organizations in raw or processed form, so that data becomes the product itself (Meisel et al., 2019).

With the amount of available data growing and data posing as a strategic asset to firms, data economies have started to evolve (European Commission, 2017). A data economy is a digital ecosystem in which a network of vendors gather, organize and exchange data (European Commission, 2017). The vendors trade their proprietary data with firms or individuals, often charging a fee for the exchanged products and services. By funding a number of ICT research projects through the H2020 programme, the European Commission aims to support the emergence of a European data economy (European Commission, 2019). To study the impact of the European data economy on the EU27 GDP, the EU monitors the number of data workers, data consuming enterprises, and vendors of data products and their respective revenues. The EU27 data economy was valued at a total of €400 billion in 2019, and it is expected to grow to a size of €550 billion by 2025 (European Commission, 2020).

Data marketplaces can fulfil a key role in realizing the data economy. As an organization may not always possess the required data to carry out or improve their processes and services, they may wish to purchase this data from other organizations. A data marketplace can address this issue by providing a digital platform through which individuals and organizations can exchange data (Stahl et al., 2016; Schomm et al., 2013).

In the simplest form, a data marketplace is an internet-based store that people visit to buy data (Carnelley et al., 2016). However, in a more sophisticated form, the data marketplace operates as an independent digital intermediary that provides value to data buyers, data sellers as well as third-party service providers (Carnelley et al., 2016). In the latter form, the data marketplace functions as a multi-sided platform that enhances the value of data products and services by enabling transactions and increasing interaction among platform participants (Koutroumpis et al., 2017).

Despite the potential benefits of data marketplaces, in practice very little data is shared or traded via platforms (Koutroumpis et al., 2020). Many of the data marketplaces that have been set up have failed or are shut down. Swivel.com, a commercial data platform that offered visualization services, closed because there were less than ten customers on the platform (Kosara, 2010). Kasabi, a data marketplace for published linked data, shut down in 2012 because at the time, the growth of the market for data was too slow for the business to be sustainable (Johnson, 2012; Dodds, 2012). Microsoft Azure DataMarket, one of the first movers to enter the data market, was closed six years after its launch due to “a lack of customer interest” (Ramel, 2016).

In this study, a data marketplace is defined as the digital infrastructure on which the commercial trading of data as a valuable good takes place. This loose definition of a data marketplace entails that there no assumption is made about the number of users on each side of the marketplace. Hence, this implies there may be multiple buyers and sellers active marketplace, but also leaves open the option that there is only one single seller or buyer present each side of the market. Secondly, it is assumed that the data exchanged on the marketplace is traded as a valuable good, rather than provided for free. Thirdly, this definition implies that the data is traded via a digital infrastructure that allows marketplace participants to exchange data in commercial transactions. Although technically data can be traded via a physical marketplace, this study focusses on the commercial exchange of data via an electronic marketplace.

Only a small number of data marketplaces currently exist, and most of the existing marketplaces adopt non-profit business models or sell data via bilateral negotiated contracts (Koutroumpis et al., 2017). Surveys among data marketplace providers show that existing data marketplaces are adopting limited business models, focussed on secure revenue streams and hierarchical organizational settings (Stahl et al., 2017). This contrasts business model developments elsewhere on the Internet. Therefore, the many failures of data marketplaces may be caused by a lack of understanding of business models of this type of marketplace.

1.2. Scientific problem

The way a data marketplace operates and conducts business can be mapped and managed using a business model. A business model is a description of how a network of organizations creates and captures value (Bouwman et al., 2008). Designing a business model for data marketplaces is challenging, because this activity requires making choices about different technical, user, organisational and financial arrangements, that need to be adopted and adjusted (Faber et al., 2003). For instance, choices need to be made about the value proposition of the data marketplace: the marketplace may decide to focus on the provision of easy data access and tooling, or it may choose to focus on making the exchange of data among marketplace participants as secure as possible.

In general, little research has been conducted on data marketplaces (Thomas & Leiponen, 2016) and data marketplace business models in particular (Fruhwirth et al., 2020; Spiekermann, 2019). Between 2012 and 2014, a number of surveys and interviews was conducted among data vendors and data marketplace providers to give an overview of the data market (Stahl et al., 2017; 2014a, 2014b; Schomm et al., 2013; Muschalle et al., 2012).

The studies analyse and discuss data market pricing strategies (Muschalle et al., 2012) dimensions of data marketplaces (Stahl et al., 2017; Stahl et al., 2014a; Stahl et al., 2014b; Schomm et al., 2013) and data market trends and scenarios (Stahl et al., 2017).

A number of frameworks and classification schemes of data marketplace business models is available in the literature. Stahl et al. (2016) established developed a classification framework of electronic marketplaces and provided a definition of data marketplace based on the results of the surveys. Koutroumpis et al. (2017) distinguish four types of data marketplaces on the basis of the number of participants on both sides of the market and outline market designs for both centralized and decentralized multilateral data marketplaces. Lastly, two taxonomies of data marketplace business models were developed, by Spiekermann (2019) and Fruhwirth et al. (2020) respectively. A taxonomy is a systematic classification of a research object of interest (Nickerson et al., 2013).

While the results of the data marketplace surveys and interviews provide insight in various dimensions of data marketplaces (Stahl et al., 2017; Stahl et al., 2014; Stahl et al., 2014b ; Schomm et al., 2013; Muschalle et al., 2012), they provide a limited overview of business model dimensions, and do not pay special attention to the characteristics of data marketplace business models. Furthermore, the proposed classification framework (Stahl et al., 2016) and market designs (Koutroumpis et al., 2017) may be useful for categorizing different types of marketplaces on a high level, but they do not go into detail about characteristics of data marketplace business models.

The two existing taxonomies of data marketplace business models (Fruhwirth et al., 2020; Spiekermann, 2019) do provide an overview of various dimensions and characteristics of the business models of data marketplaces. The two studies mainly focus on multilateral data marketplaces, in which the data marketplace functions as neutral intermediary that matches multiple data buyers with multiple data sellers (Koutroumpis et al., 2017). However, in practice data is rarely traded via multilateral data marketplaces, and instead data trading often happens via bilaterally negotiated contracts (Koutroumpis et al., 2017). Besides multilateral data marketplaces, three more types of data marketplaces are identified based on the number of data buyers and data sellers on each side of the market: bilateral data marketplaces (one-to-one matching), dispersal data marketplaces (one-to-many matching), and harvest data marketplaces (many-to-one matching) (Koutroumpis et al., 2017).

Furthermore, both taxonomy studies structure their taxonomies with the use of business model ontologies that take on a single-firm perspective on business models: Fruhwirth et al. (2020) utilizes a combination of Teece (2010) and Remane et al. (2017) to structure their taxonomy, and Spiekermann (2019) makes use of a combination of Gassman et al. (2014) and Al-Debei et al. (2008) respectively. In the literature, there are various ontologies of business models available. Many conceptualizations, Teece (2010) and Gassman et al. (2014) included, focus on business models from the perspective of a single company. Consequently, the existing taxonomy studies view data marketplace business models from a single firm perspective. According to Bouwman et al. (2008) the single firm perspective on business models is rather limited; they consider the business model from the level of an enterprise, a network of companies that collaborate together to offer the customer a joint value proposition. Data marketplace providers are part of a network of stakeholders, among which data analysts, application vendors, algorithm developers, data providers, consultants, licensing entities,

platform providers, that aim to capture and appropriate value in the data ecosystem (Thomas & Leiponen, 2016; Muschalle et al., 2012). Therefore, a multi-stakeholder perspective on business models suits data marketplaces better than the single firm perspective.

Lastly, taxonomies are seldom finished, and existing taxonomies only provide interim acceptance of the object of interest (McKelvey, 1982). To improve the understanding of data marketplace business models, new studies may be conducted that elaborate, refine or disconfirm the existing taxonomies (McKelvey, 1982). The taxonomy of Spiekermann (2019) is based on 16 empirical examples of data marketplaces and the taxonomy of Fruhwirth et al. (2020) is build based on the consideration of 20 existing data. Since the data economy is emerging and more data marketplaces are being set up, new business model alterations may have been produced in practice that were not considered during the development of the existing taxonomies.

1.3. Research objective and main research question

The identified problem in this research is as follows: in practice, many data marketplaces are shut down, and surveys among marketplace providers show that existing data marketplaces are adopting limited business models, focussed on secure revenue streams and hierarchical organizational settings (see Section 1.1). This may be caused by a lack of knowledge about data marketplace business models among practitioners. Currently, little scientific research is conducted on data marketplaces and data marketplace business models in particular (Section 1.2). Therefore, the object of interest in this study are the business models of data marketplaces. Two taxonomies of data marketplace business models exist (Spiekermann, 2019; Fruhwirth et al., 2020), but these mainly focus on multilateral data marketplaces and view data marketplace business models from a single-firm perspective. As new data marketplaces are being set up, new business model alterations may have emerged that have not yet been considered in the existing taxonomies.

In order to contribute to the knowledge and to improve the understanding about the object of interest, a novel classification of data marketplace business models in the form of a taxonomy is developed in this research. Classification is an activity that involves studying cases of the object of interest and ordering and grouping similar objects into certain ‘classes’ (Lambert, 2015). For centuries, biologists been classifying objects into general and widely accepted classification schemes, naming objects and providing a common terminology within the research domain (Lambert, 2015). Classification is also frequently used in other research areas, such as organizational sciences (McKelvey, 1982), social sciences (Bailey, 1994) and information systems research (Vessey et al., 2005).

Two main philosophies of classification can be distinguished: essentialist philosophy and empiricist philosophy (Lambert, 2015). In the essentialist philosophy, categories (types) are conceptually derived to form a typology, while in the empiricist philosophy the categories (“taxa”) of the object of interest are empirically derived into a taxonomy. Typologies are built with a specific end-goal or function in mind, and only few characteristics are considered in the development (McKelvey, 1982). On the contrary, taxonomies are built based on a consideration of many characteristics, leading to a general classification (Eisenhardt, 1989). From a taxonomy, wider generalizations can be made, to form a basis for hypotheses and

theories (Lambert, 2015). A taxonomy is a systematic classification of a research object of interest (Nickerson et al., 2013).

In this study, the taxonomy development method by Nickerson et al. (2013) is used to build a taxonomy of data marketplace business models. This method combines techniques from typology development (conceptual) and taxonomy development (empirical). The method by Nickerson et al. (2013) has previously been used in business model taxonomy development studies, in the fields of among others Blockchain-based business models (Weking et al., 2019), IoT platform business models (Hodapp et al., 2019), and carsharing business models (Remane et al., 2016). The final taxonomy provides a detailed overview of the business model characteristics of data marketplaces, based on conceptual and empirical knowledge, and can be used to classify the business models of existing data marketplaces.

Two taxonomies of data marketplace business models exist (Fruhwirth et al., 2020; Spiekermann, 2019). The existing taxonomies are lacking in some areas, that this study aims to improve:

- Firstly, the two existing taxonomies mainly focus on multilateral data marketplaces. This study aims to go beyond the state of the art, by taking on a broad interpretation of data marketplaces as the digital infrastructure on which the commercial trading of data as a valuable good takes place. This interpretation of a data marketplace allows for the inclusion of other types of data marketplaces besides multilateral data marketplaces. In this perspective, no assumption is made about the number of users on each side of the market, and it focuses on the commercial trade of data via a digital infrastructure.
- Moreover, the existing taxonomies are structured using business model ontologies that view the business model from a single firm perspective. However, as data marketplace providers are part of a data ecosystem that comprises of a network of stakeholders (Thomas & Leiponen, 2016; Muschalle et al., 2012), this study aims to go beyond the state of the art by developing a taxonomy from a multi-stakeholder perspective on business models (Bouwman et al., 2008).
- Furthermore, because the data economy is emerging and new data marketplaces are being set up, new business model alterations may have been produced in practice that are not considered in the existing taxonomies. By taking into account newly setup data marketplaces and novel alterations of data marketplace business models, the taxonomy updates and refines the knowledge about the object of interest.

A research objective is formulated to explicate the identified problem, knowledge gap and envisioned solution that helps solving the identified problem. The research objective of this study is formulated as follows:

Design a taxonomy of data marketplace business models, using the standard taxonomy development method by Nickerson et al. (2013), that can function as a tool to classify business model characteristics of data marketplaces. The taxonomy shall be developed based on both existing scientific concepts and characteristics found in empirical cases. The term data marketplace shall be broadly interpreted during the taxonomy development process and the business models of data marketplaces shall be viewed from a multi-stakeholder perspective. Newly set up data marketplaces that have not yet been considered in existing studies shall be

considered to take into account novel alterations of data marketplace business model characteristics.

The main research question that summarizes the academic challenge and structures the research is:

How can the business model characteristics of different types of data marketplaces be classified into a taxonomy from a multi-stakeholder perspective?

1.4. Research approach

The selected research approach for this study is presented in this section. The research approach provides guidance for determining the organization of research activities, that together will provide a logic answer to the main research question of this study. Using the select approach, the main research question is divided into relevant sub question that structure the research process.

Section 1.3. explicated that in this study, the object of interest are the business models of data marketplaces, and that a taxonomy will be developed to improve the understanding about this object. To guide the taxonomy development, the design science research approach by Hevner (2007) is employed in this study. The methodology was developed in the field of Information Systems research, with the goal to tackle organizational challenges by creating and evaluating (IT) artifacts.

A design science research study comprises of three complementary cycles of research activities: the rigor cycle, relevance cycle and design cycle (Hevner, 2007). In the rigor cycle, scientific theories, methods and expertise are examined to provide a theoretical foundation for the research. In the relevance cycle, design requirements are derived from problems and opportunities in the real-world environment. Central in design science research is the design cycle, that comprises of an iterative process of building and evaluating design artifacts. Artifacts built in design science research include but are not limited to: models, methods, processes, tools and theories (Hevner, 2007).

In this research, the design science research cycle is performed to create an artifact in the form of a business model taxonomy. To design the taxonomy, the taxonomy development method by Nickerson et al. (2013) is employed. This method was designed in such a way that it is analogous to the design science approach (Nickerson et al., 2013), and therefore includes a rigor, relevance and design cycle.

The identified problem that this design science research aims to solve is explicated in section 1.3: in practice, many data marketplaces are shut down, and surveys among marketplace providers show that existing data marketplaces are adopting limited business models, focussed on secure revenue streams and hierarchical organizational settings. This may be caused by a lack of knowledge about data marketplace business models among practitioners. Currently there is little scientific research available on data marketplaces and data marketplace business models in particular (section 1.4). Therefore, there seems to be a lack of knowledge about the data marketplace business models in both science and practice.

Because data marketplaces can fulfil a key role in realizing the data economy, the problem is significant. The taxonomy can improve the understanding of data marketplace business models by providing an overview of business model characteristics found in existing cases. The taxonomy may provide new knowledge to practitioners and may encourage them to set up new data marketplaces or to alter business models of existing data marketplaces.

An advantage of the design science approach is that the designed artifact may provide a better understanding of an identified problem (Hevner, 2007). The approach consists of iterative cycles of design and evaluation of the artifact, which may enhance the quality of the final product of the research. A disadvantage of the approach is that the subjectivity of the researcher may cause ethical issues (Iivari, 2007). The researcher may be biased to work towards a certain outcome, based on the researchers' values or the interests or power of the client or other stakeholders in the research field. In order to overcome subjectivity in the artifact development, measures will be taken in this study during the multiple design iterations in the artifact development process.

1.5. Research process and sub questions

The design science research approach can now be utilized to divide the main research question into relevant sub questions. This section outlines the research process that is followed to answer the main research question. The research is broken down into five sub questions that together will provide an answer to the main research question. The selected methods, tools and required data per sub question are discussed.

In order to answer the main research question, a taxonomy of business models of data marketplaces is created, using a standard taxonomy development method (Nickerson et al, 2013). The taxonomy development method consists of seven steps for building a taxonomy. A visualization of the method is presented in Figure 1 (Nickerson et al., 2013, p. 345).

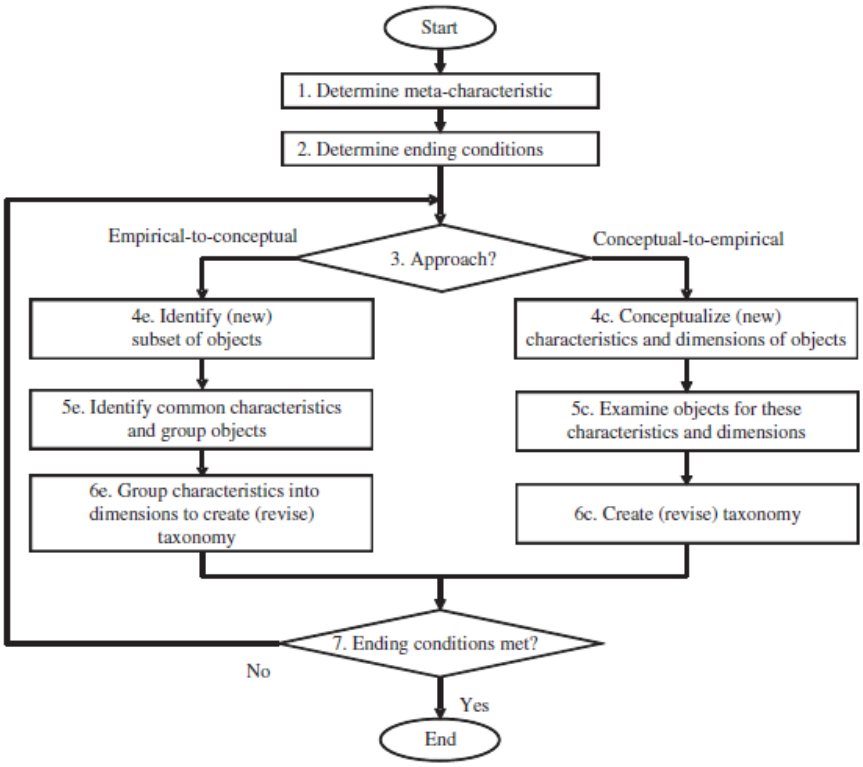


Figure 1: the taxonomy development method (Nickerson et al., 2013) p. 345

Before the taxonomy development method takes off from 'Start', it is important to first gain a profound understanding of the object of interest (Nickerson et al., 2013). In this study, the object of interest are the business models of data marketplaces. Therefore, the notion 'data marketplace business model' is first explained on the basis of a number of key concepts from the literature: business models, digital platforms, electronic marketplaces and data marketplaces. The corresponding question to this first research step is:

SQ 1: How can a data marketplace business model be defined?

A literature review is conducted to discover existing theories and artifacts about the object of interest (Webster & Watson, 2002). State of the art classifications and taxonomies of data marketplaces and data marketplace business models are analysed and discussed. Furthermore, existing artifacts that are relevant to the object of interest are examined. The characteristics of data marketplace business models that result from the literature review serve as input for input for the taxonomy design process. The second sub question is:

SQ 2: What theories and artifacts are currently available in the state of the start research that classify or are relevant to the classification of data marketplace business models?

Having established a profound understanding of the object of interest and having reviewed existing solutions in the state of the art research, the first step of the taxonomy development process is to define the meta-characteristics of the taxonomy (Nickerson et al., 2013). These set the basic conditions on which the taxonomy will be built upon, and also set the boundaries to which the object of interest will be considered. Then, requirements are set that define the ending conditions of the taxonomy development process. In order to overcome subjectivity in

design research, objective and subjective ending conditions are pre-defined for the taxonomy development method, that researchers need to consider during and at of the end of the multiple design iterations in the taxonomy development process (Nickerson et al., 2013). Therefore, the third sub question in this research is:

SQ 3: What are the meta-characteristics of data marketplace business models and what ending conditions terminate the taxonomy development?

When the meta-characteristics and ending conditions are set, the design phase can start. In the design phase of the taxonomy development method, researchers can choose between two possible design approaches: the conceptual-to-empirical approach and the empirical-to-conceptual approach (Nickerson et al., 2013). In this study, both approaches are employed to build the taxonomy. In the conceptual-to-empirical design approach of the taxonomy development method, characteristics and dimensions of the object of interest are conceptualized and empirical objects are examined to detect these characteristics and dimensions (Nickerson et al., 2013). In the empirical-to-conceptual approach, existing or new cases of the object of interest are identified. Common characteristics among these objects are identified and objects that have similar characteristics are grouped. Characteristics that exhibit similarities are grouped into dimensions to create or revise the taxonomy.

Therefore, desk research is conducted to constitute a database of empirical cases of existing data marketplaces. From the database, a representative sample of data marketplaces is taken that will be considered for a within case analysis on their business model characteristics. This phase of the research is equivalent to the relevance cycle in design science research (Hevner, 2007). The results of the within case analysis serve as input for the design iterations.

In this study, the design phase starts with a conceptual-to-empirical approach. In these design iterations, the preliminary business model taxonomy is applied to the sample of empirical cases to see if the conceptual dimensions and characteristics correspond with the empirical dimensions and characteristics. To gather information about the business model characteristics of the sampled data marketplaces, a within case analysis is conducted in which websites and publications about the companies are scanned for relevant information about the business model characteristics of the respective case. The identified characteristics that result from the within case analysis are specified in a comprehensive table for each case. If the identified characteristics of a data marketplace are not yet specified in the preliminary taxonomy, the taxonomy is revised by adding the new characteristics to the existing dimensions.

The conceptual-to-empirical design iterations are followed by empirical-to-conceptual design iterations. In this phase, the newly found business model dimensions and characteristics that result from the within case analysis of empirical cases are iteratively added to the preliminary taxonomy, to construct the final taxonomy of business models. The taxonomy design is evaluated by checking the framework with the pre-defined ending conditions. When the results of the evaluation of ending conditions are not satisfactory, researchers may choose to go back to the drawing table and pick one of two approaches to revise and improve the taxonomy. This way, multiple iterations may be conducted. When the ending conditions are met, the taxonomy development process ends. The leading question in the design phase of the taxonomy development is:

SQ 4: What business model characteristics can be derived from the business models of existing data marketplaces and how can these revise or be added to the preliminary taxonomy?

After having established the final taxonomy, the use of the taxonomy is demonstrated by applying the framework to three empirical cases. The demonstration of the taxonomy can help researchers and practitioners to understand how to the taxonomy for the classification of data marketplace business models. Thus, the final sub question of this research is:

SQ 5: How can the business model taxonomy be used to classify business models of different types of data marketplaces?

Finally, conclusions will be drawn from the final taxonomy and the taxonomy development process. The observed results will be evaluated, and limitations of the research will be addressed.

1.6. Research flow diagram

The research activities of this study and the corresponding data requirements and methods that were discussed in the previous section can now be pictured in a comprehensive research flow diagram. The sequential phases of the research process and the major research activities are visualized in the research flow diagram in Figure 2.

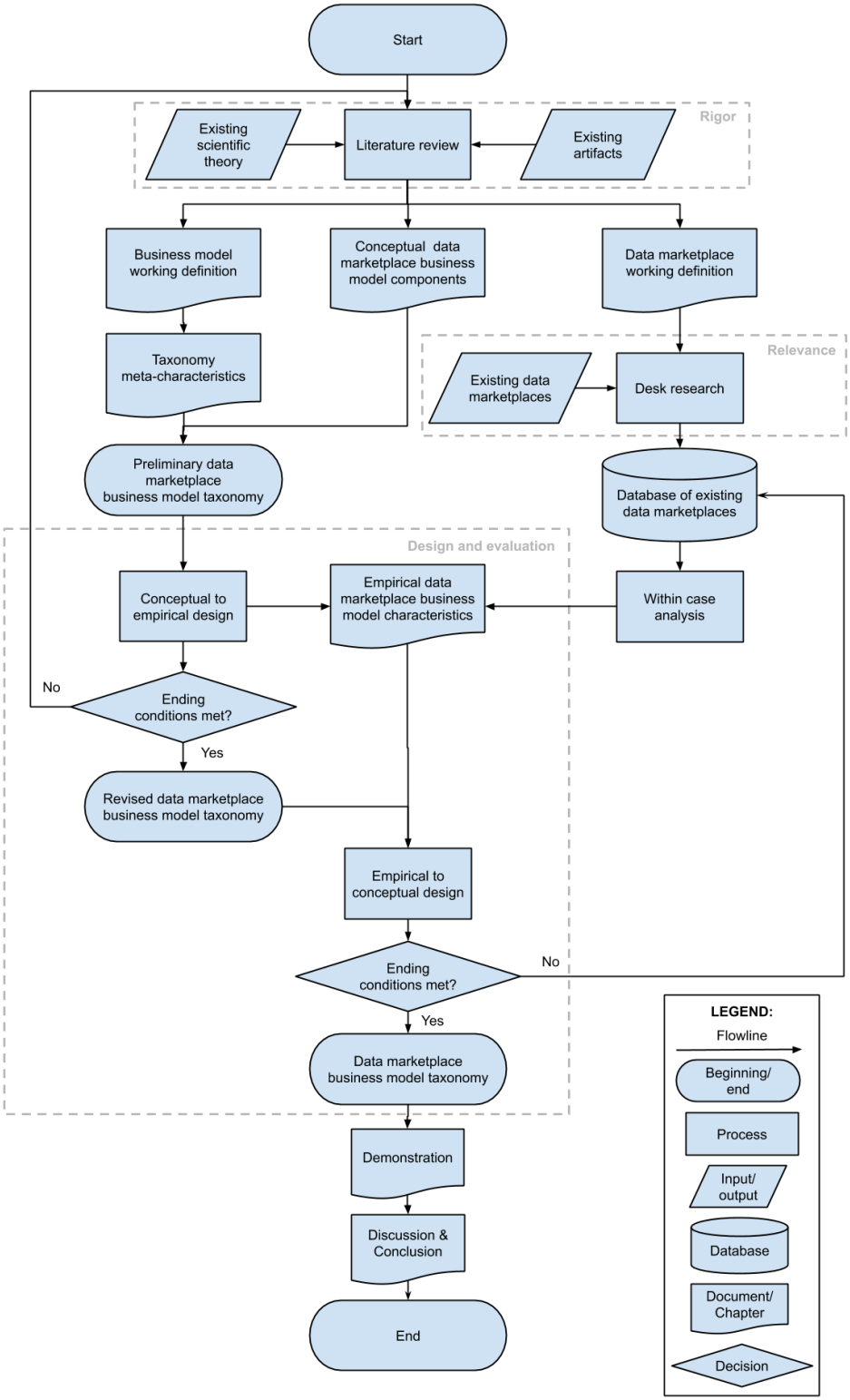


Figure 2: Research flow diagram (own illustration)

1.7. Thesis structure

The remainder of this thesis is structured as follows: in Chapter 2 the object of interest of this study, data marketplace business models, is introduced on the basis of key concepts from the literature. In Chapter 3, the results of a literature review on existing scientific theories and artifacts of data marketplaces and data marketplace business models are presented. The concepts derived from the literature serve as input for the taxonomy development. Chapter 4 describes the desk research process and sample selection of empirical cases. The taxonomy development process is described in Chapter 5, including the selection of meta-characteristics, selection of ending conditions, the multiple design iterations and the checking of ending conditions. In Chapter 6, the final taxonomy is presented, and the various dimensions and characteristics of data marketplace business models are explained. In Chapter 7, the use of the taxonomy is demonstrated on the basis of three mini case studies of empirical cases. The research is concluded by answering the main research question on the basis of the answers to the sub questions in Chapter 8. Finally, the final taxonomy and taxonomy development process are reflected on in Chapter 9.

2. Business models of data marketplaces

The main object of interest in this study are the business models of data marketplaces. Before classifying the characteristics of this object of interest on a taxonomy, it is important to define the object of interest. In this chapter, the notion 'data marketplace business model' is explained on the basis of a number of key concepts from the scientific literature. First, the business model concept is introduced, and an argumentation is given for the multi-stakeholder perspective on data marketplace business models in this research. Then, the key characteristics of digital platforms are discussed, that can enable data marketplaces to create value for multiple user groups. Next, the economic notions of markets and marketplaces are introduced, and the distinction between general and electronic markets and marketplaces is discussed. Lastly, the definition of a data marketplace that is employed in this study is presented. The different players that are part of the platform ecosystem of a data marketplace are discussed. On the basis of the various key concepts, a sound interpretation and definition of the data marketplace business model is given.

2.1. Business models

As the object of interest in this study are the business models of data marketplaces, this section introduces the concept 'business model'. The different business model research streams in the academic literature are presented and an argumentation is given for the multi-stakeholder perspective on business models that is employed in this research.

2.1.1. Business model research streams

Three main research streams can be distinguished in the business model literature: Information Systems, Strategy, and Innovation and Technology Management (Bouwman et al., 2019). In this section, the three business model research streams are discussed.

The business model became a popular research object in Information Systems research during the rise of the internet and digital business, as it enabled new ways of value creation and distribution (Krcmar et al., 2011). The emergence and expansion of e-commerce resulted in a large number of new business models, which led researchers to classify business models in the e-commerce domain (Applegate, 2001; Weill & Vitale, 2001; Tapscott et al., 2000; Timmers, 1998). Traditionally, information systems research has focussed on design approaches and the creation of artifacts (Hevner, 2007). Consequently, this research stream has produced a number of business model ontologies and tools for business model design. Prominent examples of business model ontologies and tools from the information systems domain are: the Business Model Canvas (Osterwalder & Pigneur, 2010) and the E3-value methodology (Gordijn & Akkermans, 2001).

The Strategy school of thought perceives the business model as the bridge between strategy formulation and strategy implementation (Casadesus-Masanell & Ricart, 2010; Richardson, 2008). Firstly, strategy scholars analyse the effects of business models on the performance of the firm, to distinguish successful business models from unsuccessful business models. Strategy research on business models is characterized by empirical studies, such as interviews with business leaders (Giesen et al., 2007; Cantrell & Linder, 2000), analyses of specific industries (Zott & Amitt, 2008), and stock market analyses (Weill et al., 2011).

Secondly, strategy scholars analyse the use of business models to explain value creation in networks of stakeholders, where value is created beyond the boundaries of the firm (Hedman & Kalling, 2003; Amit & Zott, 2001).

In the Innovation and Technology Management research stream, business models are perceived as a mechanism for the commercialization of innovation and technology, and as a means of innovation itself (Krcmar et al., 2011). Major authors in this research stream are Chesborough and Rosenbloom (2002), who conducted an in-depth case study at Xerox, a global printing company, to investigate the role of business in models in capturing value from technological innovation. In this school of thought, the business model is viewed as a means to convert technology development into economic outputs (Chesborough & Rosenbloom, 2002). In line with this perspective, Teece (2010) states that a "business model articulates the logic, the data and other evidence that support a value proposition for the customer and a viable structure of revenues and costs for the enterprise delivering that value" (p. 173).

An important difference between the various business model definitions and ontologies is that some authors view the business model from the perspective of a single company (Teece, 2010; Casadesus-Masanell & Ricart, 2010; Chesborough & Rosenboom, 2002), while other authors view the business model from a multi-stakeholder perspective, where a network of companies collaborates together (Bouwman et al., 2008; Gordijn & Akkermans, 2001; Amit & Zott, 2001; Timmers, 1998). While business model definitions that adopt a single firm perspective either describe the position of an individual company in the value chain or the role that individual firms play in value networks, the network view on business models emphasizes the cross-company collaboration that takes place to offer the product or service to the customer (Faber et al., 2003).

2.1.2. Multi-stakeholder perspective on business models

When referring to the term business model in this study, we refer to the business model definition by Bouwman et al. (2008), who define a business model as "a blueprint for a service to be delivered, describing the service definition and the intended value for the target group, the sources of revenue, and providing an architecture for the service delivery, including a description of the resources required, and the organizational and financial arrangements between the involved business actors, including a description of their roles and the division of costs and revenues over the business actors" (p. 3). In this definition, the perspective on business models is extended from a single company to the level of an enterprise: a network of companies that collaborate together to offer consumers a joint value proposition. The service that is offered to the customer and delivered by the network of stakeholders is central in this definition.

The multi-stakeholder perspective on business models in the definition of Bouwman et al. (2008) fits with the concept of data marketplace providers, as these companies are involved in a network of business actors that aim to capture and appropriate value in the data ecosystem (Spiekermann, 2019; Chakrabarti et al., 2018; Thomas & Leiponen, 2016; Muschalle et al., 2012). Furthermore, the focus on the service offering in the business model ontology suits data marketplaces, as data marketplaces aim to provide a marketplace where data buyers and data sellers meet for the commercial trading of data (Koutroumpis et al., 2020). This entails that the main service provided by the data marketplace provider is the

provision of an infrastructure for the buying and selling of data (Stahl et al., 2016). To enhance the value of the service offering, the marketplace provider may offer additional data-related services on top of the infrastructure. These services can range from data analysis tools to visualisation and preparation services (Spiekermann, 2019). These additional services on top of the data may also be offered by third-parties, such as software and application developers that develop data-related services (Muschalle et al., 2012).

The adoption of the business model definition by Bouwman et al. (2008) as the leading perspective on business models implies that in this study, the business model of a data marketplace is considered to be a joint effort of multiple stakeholders (data marketplace provider and possible third-party service providers) and that designing a viable data exchange service and other data-related services is central in the business model approach.

2.1.3. The STOF ontology

The business model definition by Bouwman et al. (2008) lays the foundation for the STOF ontology. The STOF ontology is a business model ontology that was originally developed for the design of business models for mobile ICT services (Faber et al., 2003). As the ontology is design-oriented and functions as tool for business model design, it can be classified as part of the Information Systems research stream in business model research. The ontology takes service as unit of analysis, and takes consideration of the network of stakeholders that collaborate together to offer a joint value proposition to customers (Bouwman et al., 2008). Central in the STOF approach to business model design is the customer value, and the organizational, technological and financial arrangements that are necessary to offer a service that provides value to both customers and service providers (Bouwman et al., 2008). Based on this view, the framework comprises of four business model domains: the service domain (S), technology domain (T), organization domain (O) and finance domain (F), of which the initial letters of the domains make up the name of the ontology: ‘STOF’ (Bouwman et al., 2008).

The service domain represents the demand side of the service offering, and the organization, technology and finance domains represent the supply side of the service offering (Faber et al., 2003). In the STOF approach, business model design starts with the definition of the demand side, and therefore the service offering is specified first in the service domain of the STOF model (Bouwman et al., 2008). In the service domain, the focus lies on the value proposition that is offered to the customer. The service definition is central in the STOF ontology, and serves as reference point to the other domains in the model. Next, the technical functionality that is needed to actualize the product or service offering is defined in the technology domain (Bouwman et al., 2008). In the STOF ontology, technology is viewed as an enabler of customer value, and therefore user requirements play a major role in the technology domain. After having specified the service and the required technology, the way resources are made available in the form of organizational arrangements is specified in the organization domain (Bouwman et al., 2008). Finally, the revenue model and pricing strategies are defined in the finance domain (Bouwman et al., 2008).

The four business model domains of the STOF ontology provide the meta-dimensions of the business model taxonomy that is developed in this study. Consequently, the STOF ontology provides the lens through which the various business model characteristics of data marketplaces are classified.

2.2. Digital platforms

The data marketplace provider offers a digital infrastructure that enables marketplace participants to buy or sell data goods (Stah et al., 2016). In an advanced form of a data marketplace, the platform functions as a digital intermediary that provides value to data buyers, data sellers, and third-party service providers (Carnelley et al., 2016). In this form, the data marketplace operates as a multi-sided digital platform that allows value creation in the form of transactions and innovation. This section addresses the two value creation mechanisms that data marketplaces may enable, and discusses two key characteristics of digital platforms that data marketplaces may hold.

2.2.1. Digital platforms as multi-sided intermediaries

In economic theory, a platform functions as a mediator between different user groups (Rochet & Tirole, 2003). Economic value is created on the digital platform through interactions between pairs of end users, such as buyers and sellers, in the form of transactions. Platforms that mediate between multiple groups of users are also called multi-sided platforms (Boudreau & Hagiu, 2009). By connecting different user groups, digital platforms create network externalities (also referred to as network effects) that represent the increase in usefulness of a good for a user that arises when the usage of the good by other users increases (Katz & Shapiro, 1985). A distinction is made between direct and indirect network effects, sometimes also referred to as same-side and cross-side network effects respectively (Eisenmann et al., 2006). Direct network effects arise when an increase in usage of a platform by one user in a certain user group increases the utility of the platform for all other users in that same user group. Indirect network effects emerge when the usage of one product or service on the platform increases the value of other products or services offered via the platform.

In principle, a data marketplace provider may take on the role of a neutral intermediary that matches one or multiple data buyers with one or more sellers in a two-sided market. If the marketplace decides to also allow third-party service providers to interact with buyers or sellers on the marketplace, the data marketplace functions as a multi-sided platform that mediates between multiple user groups. Examples of third-party service providers that may capture and create value on a data marketplace are application vendors and algorithm developers (Muschalle et al., 2012).

2.2.2. Digital platforms as breeding grounds for innovation

From a technical perspective, a digital platform functions as a code base that can be extended with third-party modules, such as software and applications (Tiwana et al., 2010). The platform has interfaces through which third-party complementors can interact and exchange data and information with the platform. The extendibility of the digital platform is enabled by its technical architecture: digital platforms comprise of a modular and stable set of core components with a variable periphery of complementary components (Baldwin & Woodward, 2009). The two sets of components interact and are governed via interfaces at the boundaries of the platform. This combination of stability and variety in the technical architecture allows modular innovation: while the interfaces of the platform remain stable, the core components and complementary components may change over time (Henderson & Clark, 1990). This way, digital platforms can allow third-party complementors to develop new innovations on the

boundaries of the platform, without them having to build an entire new system. The newly developed applications may in turn provide value to other participants on the platform.

In the context of data marketplaces, the platform provider may allow third-party service providers to have access to some parts of the code of the platform, to develop and offer services to other platform participants. The data marketplace may decide to open 'shelf' space on the marketplace to third-parties, by renting out parts of the marketplace infrastructure in return for a certain fee (Muschalle et al., 2012). Different modes of access to the data marketplace may be offered, for instance in the form of APIs (Fricker & Maksimov, 2017). If the data marketplace takes on the form of a multi-sided platform, third-party service providers may provide value to both data buyers and data sellers on the marketplace. The external parties can provide value to data sellers by offering services to leverage their data offerings, and they may offer value to data buyers by uploading applications and algorithms that ease data access and data usage (Spiekermann, 2019). The user data that is generated by usage of the third-party service by marketplace participants may be utilized by the service providers to develop and improve their offerings.

2.2.3. Data marketplaces as digital platforms

It is important to note that not all data marketplaces function as multi-sided intermediaries or breeding grounds for innovation. While matchmaking between one or more buyers and sellers is an important requirement for electronic marketplaces and data marketplaces (see Section 2.3), *multi-sided* matchmaking between a multitude of parties (buyers, sellers and third-party service providers) is not. Hence, some data marketplaces may take on the form of a data 'store', where data buyers go to buy data goods and services from a single data and marketplace provider (Carnelley et al., 2016). In this simple form of a data marketplace, the goods and services are developed and provided by the data marketplace provider itself. In this form, the data marketplace cannot be classified as a digital platform. However, in a more advanced of data marketplace, the marketplace may function like a multi-sided digital platform, where the marketplace provider takes on the role of a neutral intermediary that matches multiple buyers, sellers and external service providers (Carnelley et al., 2016). This way, the data marketplace opens up as a breeding ground for innovation, where third-party service providers are invited to develop and improve their service offerings.

2.3. Electronic marketplaces

In practice, the terms market and marketplace are often confused. In order to understand data marketplace business models, it is important to establish a common understanding of these two terms. In this section, the notions markets and marketplaces are explicated from an economics perspective. Secondly, an explanation is given how electronic markets and marketplaces differ from general markets and marketplaces.

2.3.1. Markets and marketplaces

From an economics perspective, markets are viewed as an abstract place where actors (often: buyers and sellers) meet to exchange goods and services at certain price and quantity that they agree on (Stahl et al., 2016). Conversely, a marketplace is an explicit place in terms of time and location, where the exchange of goods or services among actors takes place (Stahl

et al., 2016). In other words, a marketplace provides the physical or virtual infrastructure where products are traded, which enables the abstract concept of a market. Hence, the difference between a market and a marketplace lies in the level of abstraction of both concepts.

In general, markets fulfil three functions: they match buyers and sellers, facilitate transactions, and establish an institutional infrastructure (Bakos, 1998). Firstly, the process of matching of buyers and sellers consists of three components: determining available product offerings, searching for potential buyers and sellers, and the discovery of prices (Bakos, 1998). Price discovery entails the determination of the price at which a trade occurs between the demand and supply side of the market (Bakos, 1998). Secondly, markets facilitate transactions (Bakos, 1998). After both market parties agree on the transaction conditions, the marketplace provider ensures that the product is transported to the buyer and that the transaction is settled by transfer of payment to the seller (Bakos, 1998). To establish market transactions, a certain degree of trust among buyers, sellers and the marketplace provider is required, to prevent opportunistic behaviour of market participants. The marketplace provider offers the infrastructure on which the transfer of information, products, services and payments takes place. Lastly, a market provides an institutional infrastructure in the form of legal arrangements and regulations, to govern the behaviour of marketplace participants and to enable the efficient functioning of the market (Bakos, 1998).

2.3.2. Electronic markets and marketplaces

Digital markets and marketplaces are also referred to as electronic markets and marketplaces (Stahl et al., 2016). The distinction between electronic markets and marketplaces is analogous to the difference between general markets and marketplaces, as explained in the previous section 2.3.1; an electronic market is an abstract concept, and an electronic marketplace is an explicit place where transactions take place (Stahl et al., 2016). For a market to be classified as an electronic market, at least the negotiation between the buyer and the seller of a product offering needs to be carried out electronically (Stahl et al., 2016). In the same way, an electronic marketplace provides a digital infrastructure on which market participants interact online (Stahl et al., 2016). It is important to note that the digitality of electronic markets and marketplaces does not necessarily imply that only digital goods are traded. Both physical and digital goods may be traded on an electronic marketplace, under the condition that at least the negotiation phase is takes place in an electronical way (Stahl et al., 2016).

Electronic marketplaces have a major impact on the functions of a market, compared to general markets. Electronic marketplaces make use of information technologies to improve the matching of buyers and sellers (Bakos, 1998). Advantages that the use of information technologies in electronic marketplaces may offer are: increased personalization, cost-effective customization of product offerings, decreased search costs for buyers, lower communication costs for sellers, and new ways of price discovery (Bakos, 1998). Furthermore, electronic marketplaces decrease facilitation costs by enabling online information sharing to decrease logistics costs and by offering online payment methods to lower transaction costs (Bakos, 1998).

2.4. Data marketplaces

In recent years, data marketplaces have started to emerge and both science and industry have begun to investigate the potential of these data trading platforms. Data marketplaces are still a novel field of research, and little research on this type of marketplaces has been conducted yet (Thomas & Leiponen, 2016). This section provides the definition and interpretation of a data marketplace that is employed in this study, based on the examination of a number of different definitions in the literature. Furthermore, an overview of the players in the platform ecosystem of a data marketplace is given, to provide an argument for the multi-stakeholder perspective on data marketplace business models.

2.4.1. Definition of a data marketplace

Table 1 gives an overview of definitions of data marketplaces in the literature. In early research, data marketplaces are viewed as a platform on which any individual or organization is invited to upload and maintain datasets, and where data access and usage is regulated via a variety of licensing models (Schomm et al., 2013). Building on the earlier conducted data marketplace surveys and interviews, Stahl et al. (2016) define data marketplaces as "electronic marketplaces where the commodity data is traded" (p. 141). The authors defined two criteria for electronic marketplaces to qualify as a data marketplace: (1) The main business model of the electronic marketplace should be the provision of data and/or data related services (Stahl et al., 2016), and (2) the marketplace should offer an infrastructure for users to upload, search, retrieve, buy and sell machine-readable data. The marketplace provider is responsible for the data and the origin of the data needs to be clear (Stahl et al., 2016).

In more recent research, the platforms described by Schomm et al. (2013) and Stahl et al. (2016) are denoted as multilateral data marketplaces, where the marketplace provider functions as a neutral intermediary that matches multiple data buyers with multiple data sellers (Koutroumpis et al., 2017). However, in practice data is rarely traded via multilateral data marketplaces, and instead data trading often happens via bilaterally negotiated contracts between a single data buyer and a single data seller (Koutroumpis et al., 2017). Furthermore, it is important to note that in practice, data marketplaces often exchange access to data and data-related services rather than explicitly selling data goods (Koutroumpis et al., 2020).

Data marketplaces are more than just repositories of data sets or providers of cloud services; they function as market makers, enabling exchange of data between data providers and data consumers (Carnelley et al., 2016). In contrast with most other platforms, where data is utilized to improve services or manage customer relationships, on data marketplaces data is actually the product itself (Spiekermann et al., 2018). This way, data marketplaces enable new data-driven business models for data buyers, data sellers, and third-party service providers (Hartmann et al., 2014).

Table 1: Definitions of a data marketplace in the literature

Definition	Source
"[…] a platform on which anybody (or at least a great number of potentially registered clients) can upload and maintain datasets. Access to and use of data is regulated through varying licensing models."	Schomm et al. (2013), p. 16
"1. Having established that markets and marketplaces are shaped by the goods they focus on, a provider's primary business model needs to be providing data and/or related services to be a data marketplace. *2. Data marketplace providers need to offer an infrastructure that allows customers to upload, browse, download, buy, and sell machine-readable (e.g., RDF or XML) data. The data have to be hosted by the providers and it needs to be clear whether the specific data come from the community or the operator to classify as an electronic marketplace in the narrow sense."*	Stahl et al. (2016), p. 141
"[…] a third party, cloud-based software platform providing Internet access to a disparate set of external data sources for use in IT systems by business, government or non-profit organizations. The marketplace operator will manage payment mechanisms to reimburse each dataset owner/provider for data use, as necessary. Optionally, the marketplace provider may provide access to analysis tools that can operate on the data."	Carnelley et al., (2016), p. 5
"A data marketplace can be understood as a digital platform on which data products are traded (Lange et al., 2018; Koutroumpis et al., 2017; Fricker & Maksimov, 2017). These platforms must act like a neutral intermediary and allow anyone (or at least a large number of potentially registered customers) to upload and sell their data products. Data marketplaces can provide both static data or (dynamic) data streams and allow access via various access types, such as individual fi le downloads, APIs or customised web interfaces (Fricker & Maksimov, 2017). Therefore, data marketplaces provide standardised licensing models, as well as regulations regarding data access and usage."	Spiekermann (2019), p. 210. Based on Lange et al. (2018), Koutroumpis et al. (2017), Fricker and Maksimov (2017)

In this study, the term data marketplace is broadly interpreted on the basis of two assumptions:

1. A data marketplace is interpreted as a marketplace that enables the commercial trading of data as a valuable good (Koutroumpis et al., 2020).
2. It is assumed that the trading of data takes place on an electronic marketplace, a digital infrastructure that is provided by the data marketplace provider (Stahl et al., 2016).

Thus, in this study a data marketplace is defined as *the digital infrastructure on which the commercial trading of data as a valuable good takes place*.

This interpretation of a data marketplace has a number of important implications:

- Firstly, this definition entails that there is no assumption about the number of users on each side of the marketplace. Hence, it is assumed that there may be multiple buyers and sellers active on the marketplace, but also leaves open the option that there is only one single seller or buyer present each side of the market.
- Secondly, it is assumed that the data exchanged on the marketplace is traded as a valuable good, rather than provided for free. This implies that 'data marketplaces' hosted by government agencies and NGOs that provide free or open data are not considered to be data marketplaces, as on this type of platform data is not traded with a commercial purpose (Carnelley et al., 2016). Besides, these open data marketplaces often do not consider themselves to be a 'marketplace' since they *share* data as a public good rather than *trading* it, and because they do not adopt for-profit business models (Carnelley et al., 2016).

- Thirdly, this definition implies that the data is traded via a digital infrastructure (Stahl et al., 2016). Although technically it is possible to trade data via a physical marketplace, this study focusses on the commercial exchange of data via electronic marketplaces.

2.4.4. The players in a data marketplace ecosystem

A data marketplace may decide to function as a multi-sided digital platform by allowing third-party service providers to interact with data buyers and data sellers on the marketplace. In principle, all platforms comprise of an ecosystem with the same fundamental structure, that comprises of four different players: platform owners, providers, producers, and consumers (Van Alstyne et al., 2016). In the case of data marketplaces, these translate to the data marketplace owner, data providers, third-party service providers and data buyers (Spiekermann, 2019). An overview of the four roles is provided in Figure 3 (own illustration based on Van Alstyne et al., 2016, p.6). In this section, the different players in the data marketplace ecosystem are discussed per role.

The **platform owner** designs and holds the intellectual property rights of the platform (Eisenmann et al., 2009). For instance, T-Mobile is the owner of the Data Intelligence Hub data marketplace. The main goal of the data marketplace owner is to provide a common digital infrastructure for the exchange of products and services between the different marketplace participants (Chakrabarti et al., 2018). The marketplace provider is mainly responsible for hosting the data offerings on the data marketplace and enabling the commercial exchange of data. The marketplace provider may host the data of external data providers, but it may also host data that they created or bought themselves (Stahl et al., 2016). The provider of the digital infrastructure can influence the behaviour of participants in the ecosystem by imposing platform governance mechanisms (Tiwana, 2014).

Providers in the data marketplace platform ecosystem operate at the interface between customers and the platform, and provide value adding services to marketplace participants. They may comprise of a variety of third-party service providers (Spiekermann, 2019; Thomas & Leiponen, 2016), but the services may also be provided by the data marketplace owner itself. Examples of third-party service providers are application vendors and data algorithm developers (Muschalle et al., 2012). Application vendors develop apps and services to simplify data access and usage for data buyers. Data algorithm developers provide algorithms to improve data integration for data buyers and other service providers. Through interaction between producers, providers and consumers on the platform, new products and services may developed through a value co-creation process at the boundaries of the data marketplace (Yoo et al., 2012).

Producers in platform ecosystems are the creators of the platform's offerings (Van Alstyne et al., 2016). In the case of data marketplaces, the producers are data providers: organizations or individuals that 'own' data and use the data marketplace to store and sell their data (Spiekermann, 2019). A differentiation can be made between commercial and non-commercial data providers (Muschalle et al., 2012). Government agencies and NGOs such as the World Bank are non-commercial data providers that share their data via data marketplaces for free (Muschalle et al., 2012). Commercial data providers are organizations such as Reuters and Bloomberg that aim to commercialize their data by selling it for a certain fee (Muschalle et al.,

2012). Data providers can use the services offered by the marketplace provider or third-party service providers to leverage their data offerings (Spiekermann, 2019).

Data buyers take on the role of **consumers** in the data marketplace ecosystem, that buy the data goods and services offered via the platform (Spiekermann, 2019). Analysts form an important group of data buyers, as they create a demand for data products and data-related services through ad-hoc queries (Muschelle et al., 2012). Examples of analysts are financial analysts, sales agents, product managers or other domain experts, that buy and use data from the marketplace for the conduct of data exploration and business intelligence activities (Muschalle et al., 2012).

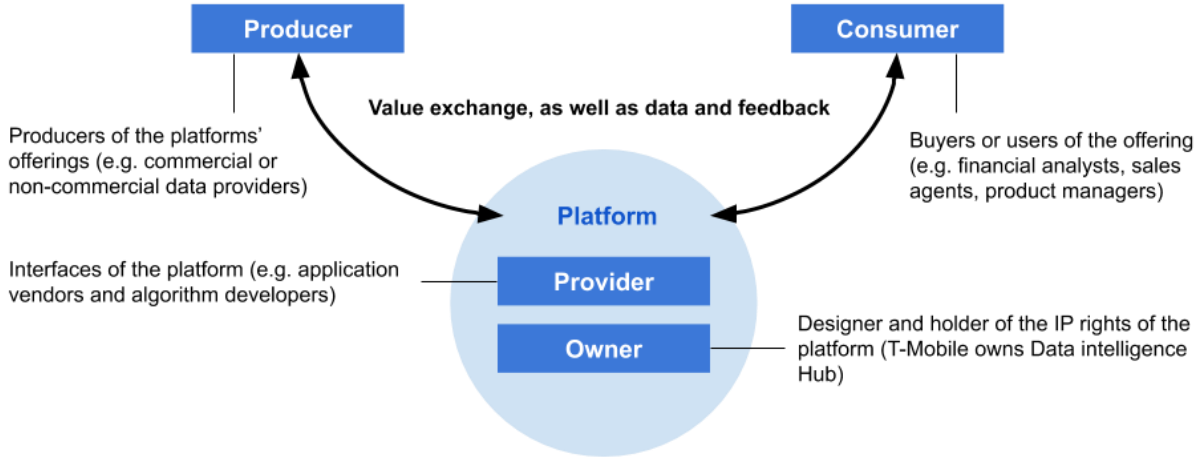


Figure 3: Players in the platform ecosystem with data marketplace examples (own illustration based on Van Alstyne et al., 2016, p.6)

It is important to note that although the players in the ecosystem can be divided into four different roles, they may switch from one role to another throughout the interactions on the platform (Van Alstyne et al., 2016). For instance, the data marketplace owner may also take on the role of data buyer, data seller or service provider in the platform ecosystem. As an additional practical example, data algorithm developers may take on the role of provider when offering their intelligent algorithms as a service to data buyers on the platform. However, they may also take on the role of consumer when renting a space on the platform to offer their services (Muschalle et al., 2012).

From the description of the different players in the data marketplace ecosystem, it can be concluded that data marketplaces comprise of a platform ecosystem in which multiple stakeholders collaborate to offer and improve their products and services. Therefore, data marketplaces require a multi-stakeholder perspective on business models, rather than a single firm perspective.

2.5. Defining the notion of a data marketplace business model

In this chapter, a number of key concepts were introduced and discussed. This section summarizes the main conclusions about each key concept and synthesizes the main findings to define the notion of a data marketplace business model.

The STOF ontology by Bouwman et al. (2008) was selected as the leading perspective on data marketplace business models in this study. In the STOF ontology, a business model is viewed as the mapping of how a network of organizations aims to create and capture value (Bouwman et al., 2008). The STOF approach takes service as a unit of analysis and employs a multi-stakeholder perspective on business models. This approach suits data marketplaces, as a network of business actors are involved in and around data marketplaces, among which data buyers, data sellers and external service providers (Spiekermann, 2019; Chakrabarti et al., 2018; Thomas & Leiponen, 2016; Muschalle et al., 2012). Moreover, the approach is well-suited for data marketplace business models because the main aim of data marketplace companies is to provide a marketplace as a service for the commercial trade of data between data buyers and data sellers (Koutroumpis et al., 2020). Additional value adding services such as data processing and visualization services may be offered on top of the data (Spiekermann, 2019).

Based on the theories on electronic marketplaces, a data marketplace was defined as the digital infrastructure on which the commercial trading of data as a valuable good takes place in this chapter. The four main roles of players in the data marketplace ecosystem were defined as: the data marketplace owner, data providers, third-party service providers and data buyers. Combining this definition with the STOF approach to business models, a data marketplace business model is defined as: *The mapping of how a data marketplace enterprise aims to create and capture value by providing a marketplace and additional value adding services for the commercial trade of data between data providers and data buyers.* In this definition, the data marketplace enterprise may comprise of the single data marketplace provider or a combination of a data marketplace provider and external service providers.

3. State of the art classifications and taxonomies

The business models of data marketplaces are the main object of interest in this study. In order to discover existing scientific knowledge and artifacts about data marketplace business models, a literature review was conducted (Webster & Watson, 2002). This section presents the scientific studies and existing artifacts that resulted from the literature review. The existing knowledge and artifacts provide the conceptual basis for developing a preliminary taxonomy. The preliminary taxonomy that resulted from the concepts found in the literature is presented at the end of this chapter, and serves as input for the conceptual-to-empirical design phase in the taxonomy development (Nickerson et al., 2013).

3.1. Literature review process

A literature review was conducted to identify academic literature relevant to the classification of data marketplace business models (Webster & Watson, 2002). The scientific database Google Scholar was consulted to find relevant academic sources, using the search string “Data marketplaces” AND (“Business models” OR “Digital platform” OR “Digital marketplace” OR “Data trading” OR “Data economy”). This string resulted in a total of 359 articles.

The articles were scanned based on their title, abstract and relevance, which resulted in a preliminary selection of 17 articles. After making this pre-selection of articles, the full text of the articles was read. Special attention was paid to whether the studies discussed dimensions and characteristics of data marketplaces and data marketplace business models. The full reading of the text resulted in the exclusion of 7 articles, that did not explicitly discuss dimensions or characteristics. The articles that were excluded from the list of relevant articles were not fully omitted, but they were used to provide background information in Chapters 1 and 2 of this research and to complement the information from the relevant sources. Based on the literature mentioned in the selected articles, 4 additional articles that presented topic-relevant business model taxonomies were added to the list.

The literature review resulted in a final set of 14 articles, which are presented in Table 2. For every article, an indication is given about the type of research.

Table 2: Overview of classifications and taxonomies relevant to data marketplace business models

Author(s) (Year)	Title	Type	Citations (14.04.2020)
Schomm et al. (2013)	Marketplaces for data: an initial survey	Dimensions of data providers and data marketplaces	73
Stahl et al. (2014a)	Data Marketplaces: An Emerging Species.		14
Stahl et al. (2014b)	The data marketplace survey revisited		16
Stahl et al. (2017)	Marketplaces for digital data: Quo vadis?		12
Stahl et al. (2016)	A classification framework for data marketplaces	Classification of electronic marketplaces	30
Koutroumpis et al. (2017)	The (unfulfilled) potential of data marketplaces	Market designs for data marketplaces	19
Muschalle et al. (2012)	Pricing approaches for data markets	Pricing models for data marketplaces	74
Fricker & Maksimov (2017)	Pricing of data products in data marketplaces		8
Spiekermann (2019)	Data marketplaces: Trends and monetisation of data goods	Taxonomy of data marketplace business models	9
Fruhwirth et al. (2020)	Discovering Business Models of Data Marketplaces		1
Bock & Wiener (2017)	Towards a Taxonomy of Digital Business Models-Conceptual Dimensions and Empirical Illustrations	Taxonomy of digital business models	22
Täuscher (2016)	Business Models in the Digital Economy: An Empirical Classification of Digital Marketplaces	Taxonomy of digital marketplace business models	6
Täuscher & Laudien (2018)	Understanding platform business models: A mixed methods study of marketplaces	Taxonomy of marketplace business models	153
Hartmann et al. (2014)	Big data for big business? A taxonomy of data-driven business models used by start-up firms	Taxonomy of data-driven business models	131

In the following sections, the content of relevant articles that resulted from the literature is discussed and synthesized. The dimensions and characteristics of data marketplaces and data marketplace business models serve as input for the development of the conceptual framework, the preliminary taxonomy.

3.2. Dimensions of data marketplaces and data providers

Studies on data marketplaces first started when researchers observed that data vendors, data warehouse providers and software developers began to offer their data, platforms and software via data marketplaces (Muschalle et al., 2012). Between 2012 and 2014, three surveys were conducted among data vendors and data marketplaces providers to give an overview of the data market (Stahl et al., 2017, 2014a, 2014b; Schomm et al., 2013). In the surveys, a data marketplace is defined as a platform on which anybody (most of the times: registered users) can upload and maintain datasets, and data access is regulated via standardized or negotiated licensing models (Schomm et al., 2013). In addition, data vendors are entities that own datasets and offer it to others (Schomm et al., 2013). They may choose to charge a certain fee for the exchanged data or provide the data for free. Data vendors can offer the datasets by themselves or choose to exchange the data via a data marketplace.

In total, 46 companies were questioned over the course of 3 years. Based on the survey results, 11 dimensions were derived to classify data marketplaces and data providers. An overview of the findings of the studies is presented in Table 3 (own representation based on Stahl, 2016, p. 170).

Table 3: Dimensions of data marketplaces (own representation based on Stahl, 2016, p. 170)

	Dimension	Categories
Objective	Type*	Web crawler, customizable crawler, search engine, pure data vendor, complex data vendor, matching vendor, enrichment tagging, enrichment sentiment, enrichment analysis, data marketplace
	Time frame*	Static, up-to-date
	Domain*	All data, finance/economic data, bio medicine data, social media data, geo data, address data
	Data origin*	Internet, self-generated, user, community, government, authority
	Pricing model*	Free, freemium, pay-per-use, flat rate
	Data access*	API, download, specialized software, web interface
	Data output*	XML, CSV/XLS, JSON, RDF, Report
	Language*	English, German, More
	Target audience*	Business, customer
	Pre-purchase tesability**,m	None, restricted access, complete access
	Ownership***,m	Private, consortium, independent
Subjective	Trustworthiness*	Low, medium, high
	Size of vendor*,m	Start-up, medium, big, global player
	Maturity*,m	Research project, beta, medium, high
	Pre-Purchase Information**,m	Barely any, sparse media information, rich media information

* = dimension first introduced in the first survey
** = dimension first introduced in the second survey
*** = dimension first introduced in the third survey
m = mutually exclusive dimension

In the dimensions that resulted from the data market survey, a distinction is made between objective and subjective dimensions of data providers. The number of stars next to the various dimensions indicates in which of the surveys the dimension was first introduced. Not all categories within the dimensions are mutually exclusive, since some of the participating companies offered multiple data-related services or provided data from multiple domains, origins, outputs, and languages (Stahl, 2016). The objective dimensions are based on the empirical evidence from a sample of 46 companies (Stahl, 2016). Most of the dimensions relate to the datasets offered by the providers, such as the timeframe, industry domain, origin of the data, data access, data output, language, and pre-purchase testability. Some of the dimensions can directly be related to the business model of data marketplaces: the type of company, pricing model, target audience, and ownership of the company. The subjective criteria are based on the qualitative judgement of the researchers, and comprise of the dimensions: trustworthiness of the provider, size of the vendor, maturity and pre-purchase information (Stahl, 2016).

3.3. Classification framework of electronic marketplaces

Stahl et al. (2016) provide a classification framework of electronic marketplace business models (Figure 4). The framework comprises of three dimensions: orientation, ownership, and business model.

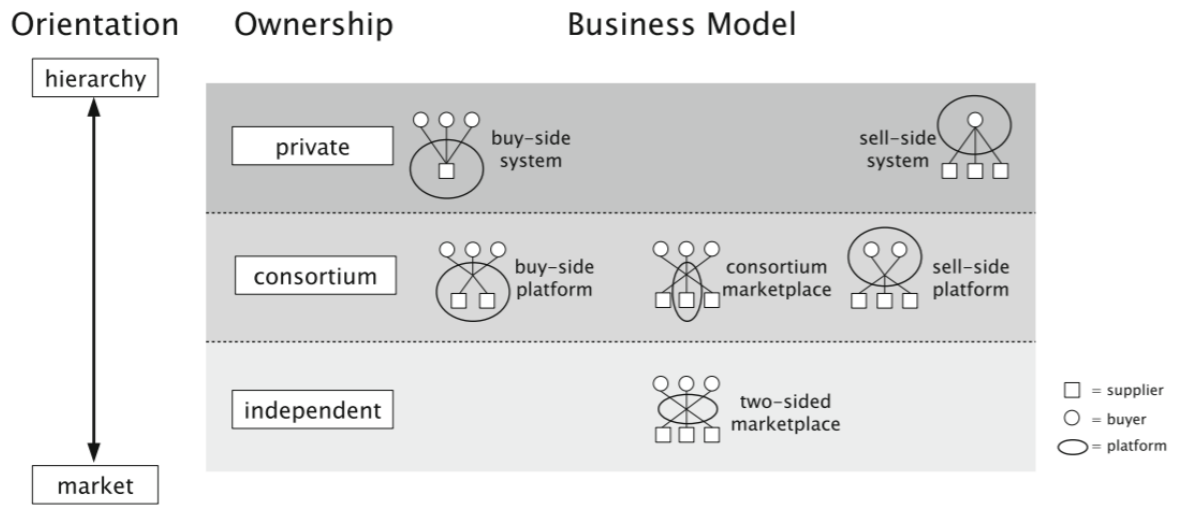


Figure 4: A model of electronic marketplaces discerning between three ownership types (Stahl et al., 2016), p. 141

Firstly, the marketplace providers are placed on a scale that indicates the orientation of transactions on the marketplace. A distinction is made between two types of transactional relations: hierarchical relations and market-based relations (Stahl et al., 2016). Hierarchic transactional relations are characterized by predetermined price levels for buyers or suppliers, while in the market orientation competition among marketplace participants determines the price and quantity of goods (Stahl et al., 2016). Marketplaces are categorized based on their ownership model: private (a single company owns the marketplace), consortia-based (a small number of companies owns the marketplace, either buyers or sellers), or independent (the platform owner has no connection with buyers or sellers) (Stahl et al., 2016). Based on the dimensions in the framework, six business model types for electronic marketplaces are distinguished: buy-side system, sell-side system, buy-side platform, consortium marketplace, sell-side platform, and two-sided marketplace (Stahl et al., 2016).

3.4 Marketplace designs for data marketplaces

Data is traded as a common-pool resource on a data marketplace because of the legal and technical governance challenges that are associated with data transactions (Koutroumpis et al., 2017). To address the common-pool issue related to data trading, Koutroumpis et al. (2017) distinguish four different marketplace designs and propose three marketplace designs for multilateral data marketplaces. The marketplace designs are presented and discussed in this section in a point-by-point manner.

3.4.1. Classification of data marketplaces by matching mechanism

Data marketplaces can be classified based on the number of parties on each side of the marketplace, that are linked by the marketplace provider via a matching mechanism (Koutroumpis et al., 2017). The four type of matching models and their corresponding marketplace design, terms of exchange and examples are presented in Table 4 (Koutroumpis et al., 2017, p. 16).

Table 4: Types of data marketplaces by matching mechanism (Koutroumpis et al., 2017), p. 16

Matching	Marketplace design	Terms of Exchange	Examples
One-to-one	Bilateral	Negotiated	Data brokers; Acxiom
One-to-many	Dispersal	Standardized	Twitter API
Many-to-one	Harvest	Implicit Barter	Google Services
Many-to-many	Multilateral	Standardized or negotiated	Microsoft Azure Marketplace

In the following points, the four distinguished types data marketplaces by Koutroumpis et al. (2017) are discussed:

- A one-to-one matching mechanisms is used in the *bilateral data marketplace design*, that can be characterized by negotiated terms of exchange (Koutroumpis et al., 2017). Data markets that employ bilateral trading agreements may not be very efficient, because of the high transaction costs needed for search, negotiation and relationship management (Koutroumpis et al., 2017).
- In the *dispersal marketplace design*, one-to-many matching is used to mediate between a single data seller and many data buyers (Koutroumpis et al., 2017). In this marketplace design standardized terms of exchange are usually employed to individually negotiate each transactional relationship. Many examples of such marketplaces exist, including platforms such as Twitter and Facebook that distribute data to advertising companies through APIs (Koutroumpis et al., 2017).
- Many-to-one matching is employed in *harvest marketplace designs*, where many sellers are simultaneously trading data with one single buyer (Koutroumpis et al., 2017). On such data marketplaces, users often trade their data in exchange for access to services that are free of charge. Google Waze and Google Search are examples of marketplaces that adopt this marketplace design, and many online social network platforms adopt similar designs (Koutroumpis et al., 2017).
- Data marketplaces that adopt a many-to-many matching mechanism follow a *multilateral marketplace design* (Koutroumpis et al., 2017). These marketplaces function as intermediaries, on which anybody can upload and maintain datasets, and where data access is managed via a variety of licensing models (Schomm et al., 2013). This type of marketplaces allows marketplace participants to interact, enabling transactions and innovation.

3.4.2. Marketplace designs for multilateral data marketplaces

Three emerging data market designs are identified for multilateral marketplaces: the centralized multilateral marketplace, decentralized multilateral data marketplace and collective multilateral platform (Koutroumpis et al., 2017). The three proposed designs are discussed in the following three points:

- *Generic multilateral data marketplaces* are designed as a centralized marketplace (Koutroumpis et al., 2017). This type of marketplaces functions as a multi-sided platform, where data providers, data buyers and complementary service providers are connected by a digital intermediary (Koutroumpis et al., 2017).

- *Decentralized multilateral data marketplaces* make use of distributed ledger technologies to enable the direct execution and verification of transactions by participants in the data market (Koutroumpis et al., 2017). Decentralized multilateral data marketplaces show similarities with the centralized marketplace design, but also come with some differences and limitations. In decentralized this design, the marketplace is now a communication structure that facilitates the operation of the decentralized data marke (Koutroumpis et al., 2017). The use of distributed ledger technology can enhance the transparency of transactions, but the scalability of decentralized systems is still unclear (Koutroumpis et al., 2017).
- The design of *collective multilateral data marketplaces* is based on Ostrom's collective governance principles: the adoption of strong boundaries via contractual agreements, clear rules and regulations and effective monitoring of the enforced policies (Koutroumpis et al., 2017). This design could help to overcome the challenges in common-pool resource markets. However, the collective platform design also faces some issues: the establishment of complex contracts and need for functional monitoring are very costly (Koutroumpis et al., 2017). Therefore, this type of platform design may be a good option to consider for a stable consortium of partners for which a high degree of trust is already established.

3.5. Pricing models for data marketplaces

The datasets offered on a data marketplace come in a variety of forms, such as raw or aggregated data, and static or streaming data (Spiekermann, 2019). The marketplace provider may provide different modes of access to the data, such as access via a data repositories, API licencing or through subscriptions (Fricker & Maksimov, 2017). Selecting an appropriate pricing model for data goods and services is a key challenge, and an important part of the business model of a data marketplace (Fruhwirth et al., 2020; Spiekermann, 2019; Fricker & Maksimov, 2017; Schomm et al., 2013; Muschalle et al., 2012). In this section, the identified pricing models for data goods offered via data marketplaces are presented point-by-point:

- The data goods on a data marketplace may be offered for *free* (Schomm et al., 2013; Muschalle et al., 2012). Many hosted by governmental organizations and NGOs do not charge any money for their datasets, and share their data for free (Muschalle et al., 2012). The marketplace provider may decide to offer parts of the data goods and services on the data marketplace for free, to attract new customers or to test new features (Schomm et al., 2013; Muschalle et al., 2013).
- In *pay-per-use pricing* or *usage based pricing*, marketplace customers are charged based on how much data or service they use on the marketplace (Fricker & Maksimov, 2017; Schomm et al., 2013; Muschalle et al., 2012). For instance, data buyers may be charged for the amount of data they buy on the marketplace measured in MB or GB.
- *Package based pricing* refers to the pricing models in which data goods are bundled in packages of a certain size, for a fixed price (Muschalle et al., 2012). Depending on the size of the package, the price may be discounted to stimulate data buyers to buy larger packages of data goods.
- In the *flat rate* or *flat fee tariff* pricing model, customers are charged a fixed amount of money to gain unlimited access to a certain service offered on the marketplace

(Schomm et al., 2013; Muschalle et al., 2012). The time span at the service may be used is often limited, to for instance a month or year of usage.

- The *freemium* model is a pricing model in which basic data goods are offered for free, and a fee is charged for premium services (Schomm et al., 2013; Muschalle et al., 2012)

3.6. Business model taxonomies

In the selected research approach for this study, design science research, examining existing artifacts in the application domain is an important activity in the rigor cycle of the research (Hevner, 2007). The literature review resulted in the discovery of a number of business model taxonomies (See Table 1 in section). Two articles presenting taxonomies of data marketplace business models were found in the literature, by Spiekermann (2019) and Fruhwirth et al. (2020). Furthermore, four business model taxonomies in fields related to data marketplaces were discovered, in the fields of digital business models (Bock & Wiener, 2017), digital marketplace business models (Täuscher, 2016), marketplace business models (Täuscher & Laudien, 2018) and data-driven business models (Hartmann et al., 2014). These artifacts offer a broader view on business models, and may comprise of dimensions and characteristics that can may be generalized and translated to data marketplace business models. In the following two sub sections, the taxonomies are presented and discussed.

3.6.1. Taxonomies of data marketplace business models

Spiekermann (2019) developed a taxonomy that presents important conceptual elements of data marketplaces in a morphological box. The taxonomy is based on dimensions found in the literature, that are compared with the business models of 16 empirical examples of data marketplaces. 5 of the 16 data marketplaces that were considered were terminated during the time of the study. The widely used business model description by Gassmann et al. (2014) was used as a starting point of analysis, to derive the taxonomic meta-dimensions: value proposition, value added and business model. Furthermore, the meta-dimension ‘value architecture’ was derived from the V4 framework for business model design in digital business (Al-Debei et al., 2008), to account for the technological and organisational infrastructure that delivers the product and service to the customer. Conceptual attributes for the taxonomy are derived from the platform-key factors for setting up digital trading platforms by Von Engelhardt et al. (2017). 8 attributes of data marketplaces are derived from the meta-dimensions and platform-key factors, and the final taxonomy is presented in a morphological box. The final taxonomy consists of 29 characteristics of data marketplace business models.

Fruhwirth et al. (2020) identify data marketplace business dimensions and characteristics from a business model perspective. The paper is based on a Master’s thesis by Prlja (2019). The authors use the taxonomy development method by Nickerson et al. (2013) to develop a business model taxonomy for data marketplaces. To provide structure to the taxonomy development process, a combination of the business model definition by Teece (2010) and the business model pattern database by Remane et al. (2017) is used to establish the meta-characteristics of the taxonomy: value creation, value proposition, value delivery and value capture. Based on concepts found in the literature and an analysis of the business models of 20 existing data marketplaces, 16 business model dimensions are derived. A frequency analysis is conducted using the taxonomy and the set of 20 existing platforms to distinguish

archetypal business models patterns of data marketplaces. From the final results, four data marketplace business model archetypes are distinguished: centralized data trading, centralized data trading with smart contract, decentralized data trading, and personal data trading (Fruhwirth et al., 2020).

3.6.2. Business model taxonomies in related fields

A number of business model taxonomies have been developed in research fields related to data marketplaces. This section provides a discussion of four taxonomies from the fields digital business models (Bock & Wiener, 2017), digital marketplace business models (Täuscher, 2016), marketplace business models (Täuscher & Laudien, 2018) and data-driven business models (Hartmann et al., 2014) respectively.

Bock and Wiener (2017) developed a digital business model taxonomy by employing the taxonomy development method by Nickerson et al. (2013). Guiding in the taxonomy development are the business model dimensions by Al-Debei and Avison (2010). In the conceptualization of taxonomic dimensions, the business model dimensions are adapted to better fit digital business models: digital offering (value proposition), digital experience (value proposition), digital platforms (value architecture/network), data analytics (value architecture), digital pricing (value finance) (Bock & Wiener, 2017). To demonstrate the use of the taxonomy, four mini-cases of companies are presented and classified along the dimensions of the taxonomy.

Täuscher (2016) constructed an empirical taxonomy of different digital marketplace business models (DMBMs), based on an existing framework (Täuscher et al., 2017). The conceptual DMBM framework used in the study provides an overview of the different business model attributes and specifications in digital marketplaces. The DMBMs framework is structured based on the five value dimensions by Abdelkafi et al. (2013): value proposition, value communication, value creation, value delivery & customer segments, and value capture. Based on business model elements in existing research, expert interviews with start-up investors and entrepreneurs, and in-depth analyses of a number of digital marketplaces, 20 business model dimensions are derived. Using a standard approach for business model classification (Lambert, 2015), six different types of digital marketplace business models were identified: efficient product transactions, product community, product aficionados, offline services on-demand, online services, and peer-to-peer offline services (Täuscher, 2016).

Täuscher and Laudien (2018) developed a taxonomy framework of marketplace business models. The authors use a mixed methods approach to develop the framework based on both conceptual and empirical concepts (Tashakkori & Teddlie, 2003). The taxonomy is structured by three dimensions, value creation, value delivery and value capture, derived from the business model definition of Teece (2010). Principles of morphological analysis were used to select the elements and specifications in the taxonomy, resulting in a taxonomy in the form of a morphological box. The taxonomy is constituted based on a review of the literature on business models and platforms, expert interviews and framework validation based on coding. In total, 14 business model attributes were derived from the literature and taxonomy development iterations. A mixed methods approach is used to identify clusters of common business models in 100 existing marketplaces. The research results reveal six types of

marketplace business models, which according to the authors shows that there is no one-size-fits all business model for marketplaces and platforms in general (Täuscher & Laudien, 2018).

Hartmann et al. (2014) developed a taxonomy of data-driven business models used by start-up firms. The business model descriptions of 100 randomly sampled start-ups were coded to develop the taxonomy. The start-ups selected for the taxonomy development rely on data as a key resource for conducting business. The framework comprises of six dimensions and 35 features that were derived from the literature. The six business model dimensions of the framework were based on a literature review of existing business model frameworks: key resources, key activities, value proposition, customer segment, revenue model and cost structure (Hartmann et al., 2014). For each of the identified dimensions, a set of features was determined from disciplines related to data-driven business, such as business intelligence and cloud-based business models. The final taxonomy is presented in a morphological box. A notable difference between this taxonomy and the taxonomies presented earlier is that in this framework, the business model can have more than one feature in each of the dimensions. Using the developed framework, a qualitative analysis was conducted on the publicly available data of the 100 sampled start-up companies. A k-medoids clustering algorithm was used to identify business model types across the sample of 100 data-driven start-up business models. The cluster analysis resulted in six different types of data-driven business models for start-ups. To validate the findings, case studies were conducted on four companies of the sample.

3.7. Conceptual framework: Preliminary business model taxonomy

The existing theories, classifications and taxonomies of the object of interest and taxonomies of business models related to the object of interest comprised of a number of business model dimensions that were found to be relevant for data marketplaces. A conceptual framework in the form of a preliminary taxonomy was developed on the basis of the dimensions and characteristics in the literature (Table 5). Below the table, a legend is provided.

When deriving the concepts from the literature, the preference was given to concepts directly related to data marketplaces over concepts from taxonomies from topic-related fields. Only when a certain dimension or characteristic was not yet present in the two existing data marketplace taxonomies, concepts from the topic-related taxonomies were selected. For each dimension, the main source or sources are noted.

A number of considerations were made when deriving dimensions and characteristics from the existing literature and artifacts:

- The industry domains specified by Schomm et al. (2013) were selected as relevant industry domains that data marketplaces are active: any data, finance/economy data, bio medicine data, social media data, geo data, and address data. The dimension was complemented with the characteristic ‘sensor data’ as depicted by Fricker and Maksimov (2017).
- The characteristic *complete access* in the *pre-purchase testability* dimension by Schomm et al. (2013) was not considered, since Fruhwirth et al (2020) found that none of the data marketplaces offered complete access to the data offered on the

marketplace. This finding also resonates Arrow's paradox, that states that it is challenging for data buyers to evaluate the quality and value of the data before purchasing and fully accessing the datasets (Arrow, 1972). This is often not possible because the buyer would then receive the data for free.

- The dimension *platform owner* by Stahl et al. (2017) was renamed to *platform sponsor,* as this is a more commonly used term in the platform literature (Eisenmann et al., 2009).
- The characteristics *no info* in the dimensions *pre-purchase testability* and *review system* resulting from the taxonomy of Fruhwirth et al. (2020) were removed. This characteristic was incorporated in the taxonomy of Fruhwirth et al. (2020) because some empirical cases did not provide enough information with regard to the respective business model dimensions. In this research, the *no info* characteristic was removed because it did not seem an appropriate business model characteristic. Rather, if an empirical case does not provide enough information about the respective business model characteristics, the cells in the classification table will be blank and the case will not be classified on the respective business model dimension (see the results in Appendix II).
- For the dimension *data origin*, the characteristic *government* was removed, as data marketplaces that offer open data from governmental organizations are not considered in this research, because they often do not consider themselves to be a 'repository' rather than a 'marketplace' and because they often do not adopt for-profit business models (Carnelley et al., 2016).
- The characteristic *multiple* was added to the dimensions *data output* and *data access*, since Fruhwirth et al. (2020) found that data marketplaces also offer multiple types of data output and data access during their empirical research on data marketplaces. Furthermore, Spiekermann (2019) found that some data marketplaces adopt *multiple* pricing models to price the data goods on the marketplace, and therefore this characteristic was added to the *pricing model* dimension.

Table 5: Preliminary business model taxonomy

	Dimension	Characteristics						Main source(s)
Service domain	**Value proposition**	Transaction-oriented		Data-oriented				Spiekermann (2019)
	Marketplace participants	B2B		C2B		Any		Fruhwirth et al. (2020)
	Industry domain	Any data Finance data	Bio Medicine data	Social Media data	Geo data	Address data	Sensor data	Fricker and Maksimov (2017), Schomm et al. (2013)
	Geographic scope	Global		Regional		Local		Täuscher and Ludien (2018), Täuscher (2016)
	Pre-purchase testability	None		Restricted access		Complete access	No info	Fruhwirth et al. (2020)
	Time frame	Static		Dynamic		Both		Schomm et al. (2013)
	Review system	User reviews		Reviews by marketplace		None	No info	Fruhwirth et al. (2020)
Technology domain	**Platform architecture**	Centralized				Decentralized		Fruhwirth et al (2020), Spiekermann (2019), Koutroumpis et al. (2017)
	Data access	API	Download	Specialized software		Web interface	Multiple options	Schomm et al. (2013), Fruhwirth et al. (2020)
	Data origin	Internet	Self-generated	User	Community	Authority	Govern-ment	Schomm et al. (2013)
	Data output	XML	CSV/XLS	JSON	RDF	Report	Multiple	Schomm et al. (2013), Fruhwirth et al. (2020)
Organization domain	**Matching mechanism**	One-to-one		Many-to-Many		One-to-many	Many-to-one	Koutroumpis et al., (2017)
	Platform sponsor	Private		Consortium		Independent		Stahl et al. (2017, 2016)
	Main revenue partner	Seller		Buyer		Third party	None / other	Täuscher and Laudien (2018), Täuscher (2016)
Finance domain	**Revenue model**	Commissions		Subscriptions		Advertising	Service sales	Täuscher and Laudien (2018), Täuscher (2016)
	Pricing model	Freemium	Pay-per-use	Flat fee tariff		Package based pricing	Multiple	Fruhwirth et al. (2020), Spiekermann (2019), Fricker & Maksimov (2017), Schomm et al. (2013)
	Price discovery	Fixed prices	Set by sellers	Set by buyers		Auction	Negotiation	Täuscher and Laudien (2018), Täuscher (2016)
	Key costs	User acquisition and retention		Platform infrastructure and development		Service capacity	Other	Täuscher (2016)
	Smart contract	Yes				No		Fruhwirth et al. (2020)
	Payment currency	Fiat money		Cryptocurrency		Both		Fruhwirth et al. (2020)

Legend:

Unchanged	Renamed	Additional characteristic	Removed

4. Desk research and sample selection

To ensure the practical relevance of the to be designed artifact, desk research is conducted to constitute a database of empirical cases of data marketplaces (Hevner, 2007). A representative sample of data marketplaces is sampled from the database to conduct a within case analysis on the business models characteristics of the set of cases. The results of the within case analysis serve as input for the iterative taxonomy development process. In this chapter, the desk research process for empirical cases is described and the sample selection method is elaborated on.

4.1 Desk research process

To ensure the relevance of the to be designed artifact (Hevner, 2007), desk research was conducted to compile a database of existing data marketplaces. A number of different sources that link to data marketplace websites were considered:

- 65 websites of data marketplaces that were mentioned and analysed in existing studies of data marketplaces were included in the database (Koutroumpis et al., 2020, 2017; Prlja, 2019; Spiekermann, 2019, Stahl et al., 2016; Carnelley et al., 2016).
- The data discovery platform datarade.ai was consulted, a website that provides an overview of 1800+ data providers, 200+ data platforms and 200+ data categories. In the database of data platforms on the website, the categories ‘audience data marketplace’, ‘data marketplaces & exchanges’, ‘personal data marketplaces’, ‘IoT data marketplaces’, ‘alternative data marketplaces & platforms’, ‘financial data marketplaces’, ‘second party data marketplaces’ and ‘B2B data marketplaces’ were consulted to discover data marketplaces to add to the database of this research. In total, the search in the categories of datarade.ai resulted in the discovery of 187 data marketplaces.
- To complement the database with data marketplaces that were not considered in the existing studies or part of the datarade.ai database, the search engine Google was utilized to conduct a desk research. The keywords “data marketplace”, “data market” and “data trading platform” were applied during the search. This resulted in an additional 15 data marketplaces that were added to the database.

In total, the desk research resulted in an initial list of 267 data marketplace websites. Since 187 of the 267 cases were derived from the datarade.ai database, equal to about 70% of the total number of cases, this study highly relies on the cases in the datarade.ai website. It is therefore important to be critical towards this major source of empirical cases. In the following paragraph, a reflection is given on the company datarade.ai.

Datarade.ai is a profit-based German IT company with headquarters in Berlin. The company was set up in 2018 by a number of professionals that had worked in the data industry for years. Datarade.ai is a website that functions as a data platform that matches data buyers with data providers. The goal of datarade.ai is to provide access to and foster the exchange of external data. The matchmaking between data buyers and data providers is automated, and data buyers are offered tools to assess the data offerings of data providers. The company was set up with the help of the SAP.io foundry and Techstars Accelerator, and major investors in the company are the High-Tech Gründerfonds (€1,000,000 seed investment), Techstars

Accelerator ($120,000 pre-seed) and the Hasso Plattner Institute (€50,000 pre-seed). The use of the matching service offered by datarade.ai is free, though users are asked to sign up for a basic subscription. Furthermore, the website also offers a premium subscription and managed services option. Billing is conducted on a subscription based for a periodic fee. Data providers can offer their data offerings on the website for free, but a referral fee is charged for each deal that is closed between the data provider and a data buyer.

To ensure that the sample of empirical cases contained relevant data marketplaces, a number of criteria were applied to the companies that resulted from the desk research:

- Data marketplaces that turned out to be terminated after inspection of the website were excluded from the database.
- The websites were inspected and analysed to make sure that the companies fit the definition of a data marketplace that is employed in this study. The companies that did not fit the definition were excluded.
- Data marketplaces that did not have an English version of their website or of which the English version seemed outdated compared to the webpage in the native language were excluded from the database.
- Data marketplaces that only provided open data, such as governmental organizations and NGOs were excluded from the database.
- Data marketplaces that were still in the construction phase were excluded.

The application of these five criteria to the cases resulting from desk research lead to the exclusion of 89 cases. The final database consisted of 178 cases of data marketplaces for further consideration.

4.2. Sample selection

To analyse the business models of existing data marketplaces, a sample was taken from the database of cases. The empiricist philosophy of classification prescribes to build a taxonomy based on the consideration of many characteristics (Lambert, 2015). Therefore, the 'population' of data marketplaces in the database was first divided into groups based on the similarity of their characteristics, to ensure that the sample size included data marketplaces with varying characteristics. The goal of this division is to ensure mutual exclusivity between cases, so that the final sample of cases consists of a set of data marketplaces that have a variety of characteristics that can serve as input for the taxonomy development.

4.2.1. Exploration of the empirical cases in the database

To explore the variety between cases in the database, the 178 cases in the database were grouped and labelled based on the different types of data that is traded on the data marketplaces. The data type was selected as the main segmentation variable to label the cases, because information about this variable was available in the sources of cases, the datarade.ai database and the scientific papers. The database of datarade.ai provided an indication of the type of data marketplace based on the labelling per category of data marketplaces ('audience data marketplace', 'data marketplaces & exchanges', 'personal data marketplaces', 'IoT data marketplaces', 'alternative data marketplaces & platforms', 'financial data marketplaces', 'second party data marketplaces' and 'B2B data marketplaces').

Therefore, 138 cases that originated from datarade.ai database could be labelled. The remaining 40 cases were labelled based on the classification of data marketplaces in existing scientific classification studies (Fruhwirth et al., 2020; Spiekermann, 2019), and through the inspection of the companies' website. Table 6 provides an overview of the number of cases per identified data type.

As can be seen in table 6, a number of labels in the database occur multiple time in the table. For instance, the database comprises of five data marketplaces that focus on the trading of alternative data, and five data marketplaces that trade financial data as well as alternative data. The first five data marketplaces are labelled as alternative data marketplace, because these companies solely focus on the trading of alternative data. The second group of five data marketplaces offers both financial data and alternative data. Hence, although the ten cases overlap in their trade of alternative data, they are labelled separately because not all of the cases provide financial data besides the alternative data. This logic of overlapping labels occurs for data marketplaces in the following industry domains: alternative data, financial data, market data, parking data, traffic data, and petrol price data.

Table 6: Labelling of data marketplaces based on data type and number of cases in the database

Labelling of cases by type(s) of data traded on the marketplace (based on labelling of datarade.ai, Fruhwirth et al., 2020 and Spiekermann, 2019)	Number of cases in the database (N)
Agriculture data	2
Alternative data	5
Any data	8
Audience data	112
B2B data	9
Connected car data, automotive data	4
Data for AI and machine learning	3
Environmental data	1
Financial data	2
Financial data, alternative data	5
Financial data, market data	2
Healthcare data	8
Location data	6
Parking data	1
Personal data	4
Real estate data	1
Sensor data	2
Satellite data	1
Traffic data, petrol price data, parking data	2
Total	**178**

After labelling the data marketplaces in the database based on the type of data traded on the platform (Table 6), it became clear that some data marketplace types in the database were overrepresented compared to others. This was especially the case for audience data marketplaces, that made up over 60% of the cases (N=112) in the database. Audience data is combined data about a certain target group of customers, the 'audience', that is often gathered by marketeers, to target the envisioned audience with highly personalized and relevant offers. All of the audience data marketplaces in the database originated from the datarade.ai database. It is unclear why datarade.ai has included such a large number of

audience data marketplaces in their database. It may be because the marketing and advertisement industry, the main consumer of audience data, is highly data-driven. The audience data market may therefore be more advanced than other industry domains, with a fragmented market and a large number of audience data marketplace providers.

If random sampling were to be applied to sample a set of cases from the database, the high number of audience data marketplaces in the database may cause a bias in the sample towards characteristics of this type of data marketplace. The goal of the taxonomy development is to consider many different characteristics of data marketplace business models. Therefore, instead of random sampling, the disproportionate stratified sampling method was applied to compensate for the overrepresentation of some types of data marketplaces in the database (Daniel, 2011).

4.2.2. Overcoming overrepresentation of segments in the population: disproportionate stratified sampling

Disproportionate stratified sampling, a variant of stratified sampling, is a method for sampling a number of cases from a population that consists of subgroups of cases that greatly differ in size (Daniel, 2011). Stratified sampling is a sampling method that consists of eight major steps, that also apply to disproportional stratified sampling (Daniel, 2011). In short, stratified sampling entails that a population is first grouped into meaningful and exclusive segments of cases ('strata'), and cases are then randomly sampled from the pre-defined strata (Daniel, 2011). In the disproportionate variant of stratified sampling, the proportion of cases in each of the pre-defined strata of the sample is not proportional to the proportion of cases of the strata in the population. This way, the strata that are overrepresented in the database can be compensated by increasing the proportion of cases of smaller strata in the sample.

The proportion of cases per stratum in disproportionate stratified sampling can be defined in a number of ways, depending on the purpose of the research. The goal of this study is to consider many different business model characteristics of data marketplaces, by conducting a detailed analysis of a variety of data marketplaces. Therefore, *the disproportionate allocation for within strata analyses* variant is picked as the proportion allocation method that best suits the purpose of this study. In this subtype of disproportionate stratified sampling, smaller strata in the population are oversampled to ensure that a sufficient number of cases in the smaller strata are available for within-strata analysis (Daniel, 2011). In this type of allocation, the proportions of cases per strata in the sample is based on the judgement of the researcher, in order to conduct detailed and meaningful analysis of specific and preferred strata (Daniel, 2011).

Stratified sampling and its subtype disproportionate stratified sampling are different from quota sampling. In stratified sampling, random sampling of cases is applied after having specified the different segments of cases, whereas in quota sampling cases are sampled in a non-probabilistic way until the predefined quotum is reached (Daniel, 2011).

The disproportionate stratified sampling method has a number of advantages and disadvantages. A disadvantage of the method is that the manipulation of proportions of segments in the sample may cause the sample to be less reliable compared to the population. As the determined proportions of segments are disproportional from the population, the end

result of the within-case analysis may be biased compared to the actual characteristics in the population. Furthermore, a misclassification of cases into segments may increase the variability of the results of the analysis (Daniel, 2011). In science, disproportionate stratified sampling is often used when subgroups in the population greatly differ in sizes but it is important to include a higher number of cases from smaller segments of the population (Daniel, 2011). An advantage of the disproportionate stratified sampling method is that it improves the representation of specific groups in the sample. This way, disproportionate stratified sampling can ensure that some groups in the sample are not overrepresented compared to others. Therefore, this sampling method is a way to ensure that a variety of data marketplaces is included in the sample, that possibly have varying business model characteristics.

4.2.3. Disproportionate stratified sampling process

The disproportionate stratified sampling method consists of eight steps (Daniel, 2011). In this section, the sampling process that was followed in this study is described:

1. First, the target population needs to be defined. In this study, the target population are data marketplaces.
2. The stratification variables of the population need to be identified and the number of desired segments should be defined. The type of data traded on the data marketplace was picked as the leading stratification variable of the population (see section 4.2.1. for the labelling process). An important requirement in disproportionate stratified sampling is that at least two cases need to be taken from each segment (Daniel, 2011). This requirement implies that the size of each segment needs to be at least two. The labelling of the cases resulted in some data marketplace segments consisting of a single case only (e.g. satellite data marketplaces), and therefore these cases needed to be merged with other segment in the next step of the sampling process.
3. An existing sampling frame needs to be picked or a novel sampling frame is developed that includes relevant information on the stratification variables of the population. The data domains of data marketplaces as specified by Schomm et al. (2013) were selected as relevant segments for the sampling frame in this study: all data, finance/economy data, bio medicine data, social media data, geo data, and address data. The frame was complemented with the segment ‘sensor data’ as depicted by Fricker and Maksimov (2017).
4. In the next step, the sampling frame needs to be evaluated, and adjustments should be made when necessary. The existing samples frame segments were adjusted based on the data labels as specified by datarade.ai. The sampling frame is divided into relevant segments. The goal is to minimize differences between segments; cases in the segments should not be overlapping. After further inspection of the cases, it was found that *AI and machine learning* data marketplaces could better be classified under *personal data* and *any data* respectively. Furthermore, the same was true for data marketplaces in the *B2B data* class, that were split up and divided among the *any data* and *audience data* categories. Based on the existing frame segments and the requirement of minimum two cases per segment into account, a novel frame was developed for this study that consists of the following sample segments: any type of data, financial & alternative data (financial, alternative, market, and real estate data), audience data, sensor & mobility data (connected car, automotive, traffic, petrol price,

parking data and sensor data), geo data (agriculture, environmental, industry location, and satellite data) and health & personal data (healthcare data, personal data)

5. Every case should be assigned a unique ID number. Thus, every data marketplace case in the database was placed in a segment that corresponded to the type of traded on the platform, and was assigned an ID within its segment.
6. Before sampling, the sample size for each segment needs to be determined. In this study, the disproportionate stratified sampling method is applied, which entails that the proportion of segments in the sample is disproportionate to their actual proportion in the population. In general, a sample size of N>30 is appropriate for statistical analysis. Based on the total number of cases in the database (N=178), a sample size of N=40 was found to be a representative sample size. A sample of 40 cases would account for 20% of the 178 total cases, which was found to be an appropriate size. Having established that the total sample size should be N=40 and following the selected disproportionate allocation for within strata analyses method, the proportion of cases for each segment were defined based on the judgement of the researcher (Daniel, 2011). The disproportionate sample size and respective percentage of the sample are presented in Table 7. As audience data marketplaces were overrepresented in the population compared to other cases, the proportion for this type of data marketplaces was drastically lowered to 22,5% of the sample. This way, the larger size of this segment was still taken into account compared to other categories, but it would not dominate or be overrepresented in the sample.
7. For each segment the predefined proportion of cases is randomly sampled. For the disproportionate stratified sampling method to be valid, it is important that at least one case is sampled from each segment and that at least two cases need to be chosen from each segment to be able to calculate the error rate from the collected data (Daniel, 2011). Thus, for each segment of cases a random sample was taken from the numbered list of cases, based on the predefined proportions and using the inter set function of the website Random.org. The final sample of selected cases per segment is presented in Appendix I. If, during the analysis of the selected cases, a data marketplace turned out to not provide enough information about its business model, the case was omitted and the case with the next number on the list of the segment was selected for further analysis. The list of cases in Appendix I comprises of that final cases that were considered in this research.

Table 7: Disproportionate sample sizes per segment of data marketplaces

Data marketplace segments	Types of data marketplaces in the segment	Number of cases in the database (N)	Disproportionate sample size	Percentage of sample (%)
Any	Any type of data	14	10	25%
Financial & Alternative	Financial data, alternative data, market data, real estate data	15	4	10%
Audience	Audience data	115	9	22,5%
Sensor & Mobility	Connected car data, automotive data, Traffic data, petrol price data, parking data, sensor data	10	6	15%
Geo	Agriculture data, environmental data, location data, satellite data	10	4	10%
Health & Personal	Healthcare data, personal data	14	7	17,5%
	Total	**178**	**40**	**100%**

5. Taxonomy development process

In this chapter, the iterative taxonomy development process that was followed in this study is described (Nickerson et al., 2013). First, the meta-characteristics of the taxonomy are defined. The meta-characteristics function as overarching characteristic of the object of interest and provide the basis of further selection of characteristics. Then, the ending conditions are specified that terminate the taxonomy development process. After that, a description is given of the multiple design iterations that were conducted to develop the taxonomy of data marketplace business models. Lastly, the checking of ending conditions per design iteration is presented and discussed. Figure 5 provides an overview of the taxonomy development process on the basis of the selected meta-characteristics and dimensions of data marketplace business models.

5.1. Determination of the meta-characteristic

During the development of a taxonomy, a central problem is the determination of the characteristics of the object of interest. In this study, the object of interest are the business models of data marketplaces. To provide a starting point for the identification of characteristics, a meta-characteristic is specified at the beginning of the taxonomy development process (Nickerson et al., 2013). The meta-characteristic functions as an overarching characteristic that provides the foundation for choosing characteristics of the object of interest (Nickerson et al., 2013). Each determined characteristic follows logically from the meta-characteristic. The determination of the meta-characteristic is based on the contemplated purpose of the taxonomy, which in turn is based on the anticipated use of the taxonomy by the target user group (Nickerson et al., 2013). Choosing the meta-characteristic is a thoughtful process, as it has major impact on the further taxonomy development.

5.1.1. Identification of user group: researchers and practitioners

In this research, the intended users of the taxonomy are scientific researchers and practitioners. These user groups can use to taxonomy classify, study and design business models for data marketplaces. The primary function of the taxonomy for both user groups is to classify the business models of data marketplaces. Furthermore, the taxonomy can function as a knowledge map for researchers that displays the available knowledge from both scientific research and empirical examples from practice. It may expose certain areas that require further research. Practitioners that have set up or are in the phase of setting up a data marketplace may use the taxonomy for designing and improving the business model of their platform. To this user group, the taxonomy can pose as a visual aid that displays best practices in business models of existing data marketplaces. Furthermore, the concepts from scientific research that have not yet been adopted in practice may inspire practitioners to alter their business model. This way, the research contributes to both science and society. Hence, the purpose of the taxonomy is to function as a classification and communication tool for both user groups.

5.1.2. Meta-characteristic: STOF ontology business model domains

The object of interest in this study are the business models of data marketplaces. As discussed in section 2.1.1, a large number of business model definitions and representations are available in the literature. In this study, the four business model domains of the STOF ontology

provide the meta-characteristics of the business model taxonomy: the service domain, technology domain, organization domain and finance domain (Bouwman et al., 2008; Faber et al., 2003):

- *Service domain*: a description of the value that the network of actors aims to deliver to the customer, specifically the service offering.

- *Technology domain*: a description of the technical architecture that the value network utilizes to deliver the proposed service offering as explicated in the service domain.

- *Organization domain*: a description of the organization of actors in the network, the value network, explicating the roles they take on and the value activities they perform to deliver the service and create value for the customer.

- *Finance domain*: a description of how the value network aims to make money from the service offering and how the costs and revenue streams are split among the different actors in the network.

Besides providing the four overarching dimensions of the taxonomy, the STOF ontology provides the logic by which the various business model characteristics of data marketplaces are classified.

5.2. Determination of ending conditions

The taxonomy development method follows an iterative design process, and therefore conditions need to be specified that terminate the process. The selected taxonomy development method by Nickerson et al. (2013) provides eight objective ending conditions and five subjective ending conditions.

The objective ending conditions of the taxonomy development are (Nickeron et al., 2013):

1. All objects, or at least a representative sample of objects have been examined

2. In the last design iteration no object was merged with a similar object or split into multiple objects

3. For every characteristic of every dimension in the taxonomy at least one object is classified

4. In the last design iteration no new dimensions or characteristics were added to the taxonomy

5. In the last design iteration no dimensions or characteristics were merged or split

6. There is no duplication of dimension in the taxonomy; no dimension is repeated and every dimension is unique

7. There is no duplication of characteristics within the dimensions; every characteristic is unique within its dimension

8. There is no combination of characteristics and there is no cell duplication in the taxonomy; each cell is unique and there is no repetition of cells

Before terminating the development process, subjective ending conditions also need to be considered. The subjective ending conditions determine whether the taxonomy is useful from the perspective of the researcher. When terminating the taxonomy development process, an argumentation needs to be given why all subjective ending conditions are met. The subjective ending conditions of the taxonomy development are (Nickerson et al., 2013):

1. The taxonomy is concise; the number of dimensions make the taxonomy to be meaningful, but not overwhelming

2. The taxonomy is robust; the dimensions and characteristics of the taxonomy enable researchers to sufficiently differentiate the objects of interest

3. The taxonomy is comprehensive; the taxonomy can be used to classify a sample or all objects within the research domain

4. The taxonomy is extendible; new dimensions or characteristics of an existing dimension can easily be added

5. The taxonomy is explanatory; the identified dimensions and characteristics can be utilized to explain about an object

5.3. Design iterations

After having specified the meta-characteristics of the taxonomy and the ending conditions of the development process, the iterative design of the taxonomy can start (Nickerson et al., 2013). The preliminary taxonomy that resulted from the literature reviews on existing theories and artifacts (Table 5) served as input for the iterative design process. In total, four design iterations were conducted, of which two conceptual-to-empirical iterations and two empirical-to-conceptual iterations. Table 9 provides an overview of the changes that were made to the initial preliminary taxonomy. In this section, the changes that were made to the taxonomy throughout the iterative design process are discussed.

5.3.1. Conceptual-to-empirical design

In the conceptual-to-empirical design phase, the dimensions and characteristics in the preliminary taxonomy (Table 5) were applied to the 40 sampled data marketplaces from the empirical database. To map the business model characteristics of the selected cases on the preliminary taxonomy, information on the business models of the cases was collected from publicly available sources. Main sources of information were company websites and blogs and news articles, to which the companies commonly link to from their website. Many companies often provided a whitepaper on their website with the vision and mission of the company, that sometimes provided information about the business model of the data marketplace.

The information sources were thoroughly analysed to classify the business model characteristics of the selected data marketplaces. A within case analysis was conducted to gather relevant information in the form of text fragments, pictures, screenshots and other informational elements. The discovered information fragments were coded using the dimensions and characteristics of the preliminary taxonomy as a guideline. Table 8 provides an example of the coding logic. A database of the coding and information sources that was developed during the taxonomy development process is available upon request.

Table 8: Coding examples for the value proposition dimension

Characteristic	Case	Quote
Easy data access and/or tooling	Open:Factset Marketplace	*"FactSet creates data and technology solutions for investment professionals around the world, providing instant access to financial data and analytics that investors use to make crucial decisions."*
	Knoema	*"Knoema is a cloud-based data technology platform that makes data accessible and delivers intelligent data tools to enable data access and discovery."*
Secure data sharing	DAWEX	*"With Dawex Global Data Marketplace providers can highlight the value of their data while retaining full control over the distribution and configuration of usage rights."*
	Snowflake	*"Unlike other data marketplaces, Snowflake Data Marketplace leverages Snowflake's Secure Data Sharing technology, which means no data transfer and no need to squeeze data through APIs or use cloud storage."*
High quality and unique data	Amazon DSP	*"Use exclusive Amazon audiences to reach your ideal audience on and off Amazon."*
	Datax	*"Quality business data for better sales leads - Any campaign is only as good as the data it's built on – so make sure yours is the best.*

During the conceptual-to-empirical iterations, the identified business model characteristics that resulted from the within case analysis were specified in a comprehensive table for each case. If the identified business model characteristic of a data marketplace was not yet specified in the preliminary taxonomy, the newly discovered characteristic was added to the existing dimension and considered in the subsequent design iterations. The classification of the business models characteristics of all empirical cases is presented in Appendix II.

Following the iterative nature of the taxonomy development process (Nickerson et al., 2013), the pre-specified ending conditions were checked after every iteration. If the ending conditions were not met, a new iteration was conducted with the dimensions and characteristics that were

specified in the previous iteration. As the design of the taxonomy improves towards its final form throughout the taxonomy development process, the first few iterations (based on the preliminary taxonomy) are more extensive than the last few iterations.

1st iteration: conceptual-to-empirical design

In the first conceptual-to-empirical iteration, a large number of dimensions and characteristics were revised or removed. The following changes were conducted:

Service domain

- The characteristics *transaction-oriented* and *data-oriented* (Spiekermann, 2019) in the *value proposition* dimension were removed and replaced with four characteristics: *easy data access and/or tooling*, *secure data sharing*, *all services in a single platform*, and *high quality and unique data*. During the desk research, it became clear that the two characteristics as proposed by Spiekermann (2019) did not cover all the value propositions of the different types of data marketplaces (bilateral, dispersal, harvesting and multilateral data marketplaces). Therefore, the dimension value proposition was kept, but the two characteristics were changed into four characteristics: *easy data access and/or tooling*, *secure data sharing*, *high quality and unique data*, and *all services in a single platform*, based on value proposition statements made on the website of the data marketplace companies.
- The frame of types of data marketplaces on the basis of the data type traded on the marketplace that was adopted in the sample selection was employed to alter the business model characteristics in the *industry domain dimension*. Therefore, the following changes were made: the *finance data industry domain* was renamed to *finance & alternative data*, the characteristics *bio medicine data* and *personal data* were merged into the *health & personal data* characteristic, the *social media data* and *address data* characteristics were added to the new characteristic *audience data*, and the *sensor data* characteristic was renamed into *sensor & mobility data*.
- Removal of the dimension *pre-purchase testability*, since after the analysis of empirical cases, it turned out that not many companies provide information about whether or not buyers can test the data before purchasing it.
- The characteristic *dynamic* in the *time frame* dimension was split up into two characteristics: *up-to-date* data and *(near) real time* data. Data marketplaces that offer up-to-date data regularly check the data offered on the platform, updating the data when necessary or asking the sellers on the data marketplace to do so. Updates are carried out in an ad hoc manner, or after certain points in time, for instance on a monthly basis. Data that is not up to date anymore may be removed from the marketplace. Real-time data or (near) real time data is often offered by e.g. financial, mobility and sensor data marketplaces, where data providers offer data that is generated by online trackers of financial information or ‘smart’ assets that are equipped with sensors and that are connected to the Internet-of-Things. The data providers can direct the generated data directly to the data marketplace, to offer the real-time or near real-time data for sale.

- Because the time frame characteristic *dynamic* data was split into two characteristics, the characteristic *both* was renamed to data in *multiple* time frames, to cover data marketplaces that offer static, up-to-date and (near) real-time data.
- The dimension *review system* and its corresponding characteristics were deleted, as after analysing the empirical cases, it became clear that the question whether or not a data marketplace offers a review system does not seem very relevant to characterize and distinguish different types of data marketplaces. None of the data marketplaces in the sample offered the option of *user reviews* and only 18% of the sampled data marketplace companies provided information about whether or not they review the data before offering the data on the data marketplace. Therefore, this dimension was deleted.

Technology domain

- At first, the dimension *data origin* and its corresponding characteristics by Schomm et al. (2013) was utilized as a starting point to classify the origin of the data offered on data marketplaces. However, during the analysis of empirical cases, it was found that not all the characteristics were specific enough. For instance, the characteristic *self-generated* was found to be unclear, since some data marketplace offer proprietary data that they generated themselves, while other data marketplaces invite external data providers to offer their sensor generated data on the data marketplace. Therefore, the data origin dimension and characteristics as described by Schomm et al. (2013) was dropped and replaced by the characteristics of the *data source* dimension, as described by Hartmann et al. (2014) in their taxonomy for data-driven business models. This dimension and characteristics offered a distinction between internal and external data, and a more precise definition of the different characteristics. The characteristic *none* was added to the two dimensions, to account for data marketplaces that either offer only internal data, or external data
- During the empirical analysis of cases, it became clear that most of the data marketplaces offer multiple types of *data output* instead of providing their data solely in XML or CSV format. Therefore, this dimension and its characteristics by Schomm et al. (2013) seemed kind of outdated. The data output type did not seem distinctive for different types of data marketplaces, it was therefore dropped.
- The characteristic *web interface* in the *data access* dimension was removed, because it was found that none of the data marketplaces offered access to the data via a web interface. This was also found in the empirical research by Fruhwirth et al. (2020).

Organization domain

- During the browsing of data marketplace company websites and other sources for information about their *main revenue partner*, it was found that many companies do not often openly state who is the key player from which they gain the most revenue. Therefore, this dimension was removed from the taxonomy.

Finance domain

- When searching for information about the *revenue model* of data marketplaces, it became clear the *advertising* and *service sales* revenue model was not adopted by the companies in the sample. Therefore, these characteristic was removed.

- The characteristic *asset sales* was derived from Osterwalder and Pigneur (2010), to account for the revenue that data marketplaces make by selling datasets and data analytics algorithms and tooling as assets on their marketplace.
- As sometimes the data marketplace is the seller of data, and sometimes external data providers are invited to the marketplace to sell data, the two characteristics *fixed prices* and *set by sellers* in the *price discovery* dimension were altered to *set by marketplace provider* and *set by external sellers*, to make a clear distinction between those two dimensions.
- Furthermore, none of the data marketplaces in the sample seemed to provide price discovery by an auction price discovery model. Therefore, this characteristics was removed from the taxonomy.
- As with the dimension *main revenue partner*, data marketplace companies did not provide much information about their *key costs*. Therefore, this dimension was removed from the taxonomy.
- From the classification of data marketplaces based on their *payment currency*, it became clear that none of the data marketplace provided the possibility to pay or get paid in *both* fiat and cryptocurrency money. Therefore, the characteristic *both* was removed from this dimension.

2nd iteration: conceptual-to-empirical design

Technology domain

- Not many data marketplaces provided data from an *internal data source* (N=5), and the ones that did mostly provided *self-generated data* (N=4). Therefore, the two dimensions *internal data source* and *external data source* that were specified in the previous domain were merged into the new dimension *data source*, and the characteristic *self-generated* was transferred to the new dimension.

5.3.2. Empirical-to-conceptual design

During the empirical-to-conceptual design, new dimensions that are found during the empirical analysis are designed into conceptual dimensions and added to the taxonomy.

3rd iteration: empirical-to-conceptual design

- During the analysis of empirical cases, it became clear that some data marketplaces are more 'advanced' than other data marketplaces since they offer a number of additional services on top of the data. Therefore, the dimension 'additional services' was added with sub-dimensions with common additional services offered by data marketplaces, namely: *data processing and analytics tools* and *enterprise data marketplace*. The dimensions comprise of a binary characteristic *yes* and *no*, to indicate whether the data marketplace offers this additional service.

4th iteration: empirical-to-conceptual design

- In the last iteration, the characteristics and dimensions specified in the previous iterations were evaluated. During this final iteration, no new characteristics and dimensions of data marketplace business models could be identified. This implies that all the sampled could be distinguished using the final set of characteristics.

Based on the multiple design iterations, the preliminary taxonomy was revised. The revised taxonomy that resulted after taxonomy development process is presented in Table 9. Below the table, a legend is provided. The dimensions and characteristics that were presented in the preliminary taxonomy that remained unchanged are presented in white. Dimensions and characteristics that were merged, refined or renamed are presented in light grey. New dimensions and characteristics are presented in blue, and dimensions and characteristics that were removed are presented in dark grey.

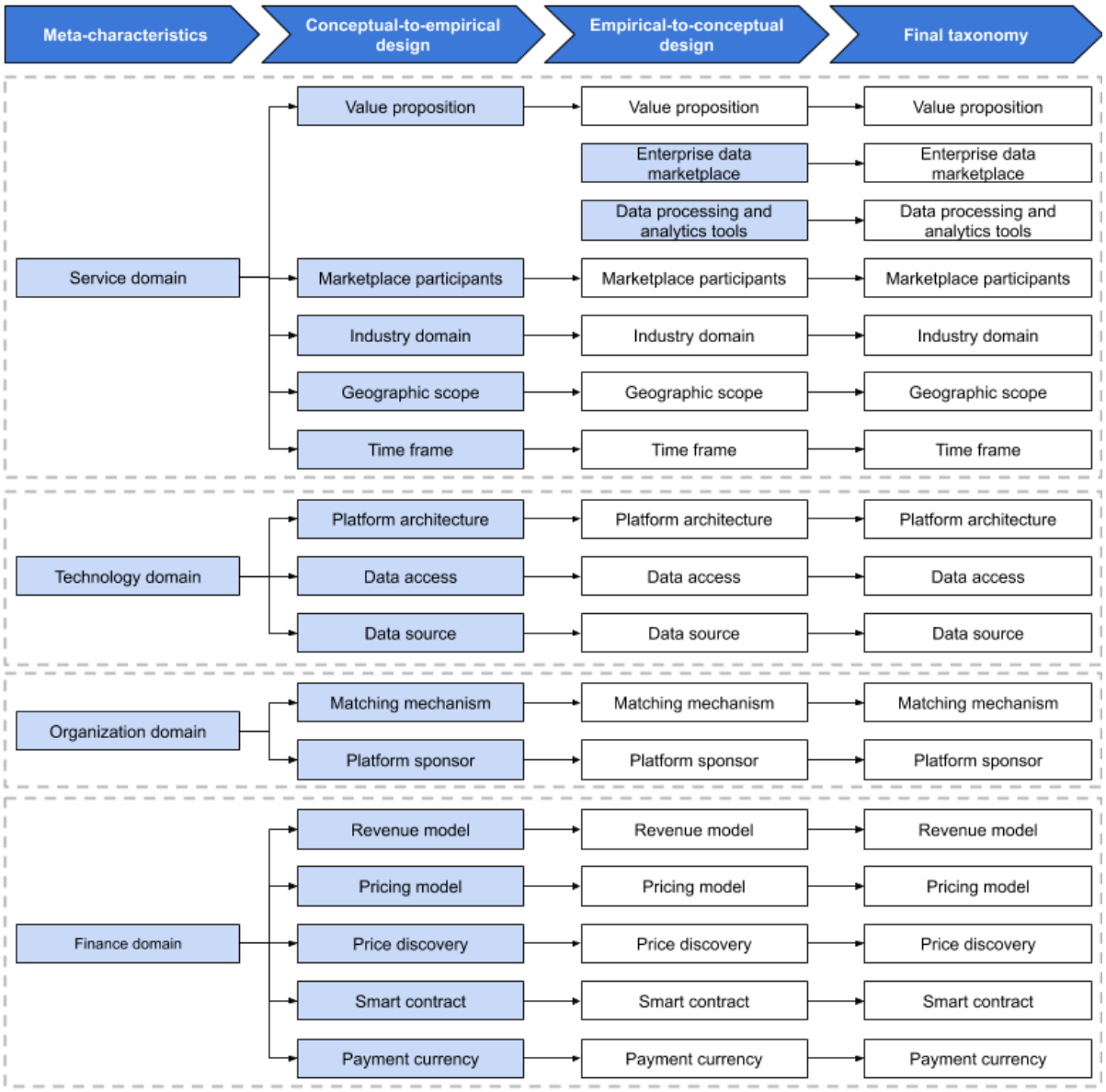


Figure 5: Overview of the taxonomy development process (own illustration)

Table 9: Revised taxonomy after design iterations

	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 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	
	Pre-purchase testability	None		Restricted access			
	Time frame	Static	Up-to-date	(Near) real-time	Multiple		
	Review system	User reviews		Reviews by marketplace		None	
Technology domain	**Platform architecture**	Centralized		Decentralized			
	Data access	API	Download	Specialized software	Multiple options	Web interface	
	Data source	Self-generated	Customer provided data	Acquired data	Multiple sources		
	Data output	XML	CSV/XLS	JSON	RDF	Report	Multiple
Organization domain	**Matching mechanism**	One-to-one	Many-to-Many	One-to-many	Many-to-one		
	Platform sponsor	Private		Consortium		Independent	
	Main revenue partner	Seller	Buyer	Third party	None / other		
Finance domain	**Revenue model**	Commissions	Subscriptions	Usage fees	Asset sales	Service sales	Advertising
	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	Auction	
	Key costs	User acquisition and retention	Platform infrastructure and development	Service capacity	Other		
	Smart contract	Yes		No			
	Payment currency	Fiat money		Cryptocurrency		Both	

Legend:

Unchanged	Merged / refined characteristic	Newly added	Removed

5.4. Checking ending conditions

After every design iteration, the ending conditions were checked. The selected taxonomy development method by Nickerson et al. (2013) provides eight objective ending conditions and five subjective ending conditions. In this section, the iterative checking of the ending conditions throughout the design iterations is discussed point by point per objective and subjective ending condition. An overview of the results of the iterative ending condition checks is presented in table 10.

The checks of the objective ending conditions developed as follows:

1. In the first iteration, all objects (the 40 sampled data marketplaces) were examined. Therefore, this objective ending condition was fulfilled for all iterations. During the following iterations, the same objects were considered and analysed multiple times.
2. The sampled data marketplaces were never split; they were analysed as impartial objects, therefore the second condition was fulfilled for all iterations.
3. A number of characteristics were removed from the taxonomy in the first iteration, because no object could be classified for that respective characteristic. For instance, the characteristic *advertisements* in the *revenue model* dimension was removed, because no data marketplace was found which used advertisements as a main source of revenue. Therefore, this condition is not fulfilled for the first iteration. In the following iterations, dimensions and characteristics were only revised or added, and not removed.
4. In the first design iteration, a large number of dimensions were added when applying the preliminary taxonomy that resulted from the literature review to the empirical cases. In the second conceptual-to-empirical iteration no new dimensions were added. In the third iteration and first empirical-to-conceptual iteration, new dimensions were added in the service domain of the taxonomy. In the final iteration, no new dimensions were added.
5. Throughout the iterative design process, dimensions and characteristics were merged and split during the first three iterations. In the final iteration, no dimensions were merged or split, and the taxonomy was evaluated with the existing dimensions.
6. The design process started with the consideration of dimensions and characteristics from the existing literature. The considered dimensions were mutually exclusive, and therefore there was no duplication of dimensions throughout the taxonomy development process.
7. After the first design iteration, a number of characteristics that showed similarities in the empirical cases were merged, to ensure that there was no duplication of characteristics within the dimensions. Therefore, this characteristic was fulfilled throughout the entire design process.
8. In the first iteration, dimensions and characteristics from the existing literature were considered. All cells that were added for consideration were unique. In the following iterations, only unique cells were added by either splitting existing characteristics or adding new characteristics that were not considered before. Therefore, this objective was fulfilled throughout all iterations.

The subjective ending conditions developed as follows throughout the design iterations:

1. The taxonomy developed process started with the consideration of 20 conceptual dimensions that were derived from the literature. This was found to be an appropriate amount for the taxonomy to still be precise. After the first iteration, a number of dimensions was deleted, among which key costs and the main revenue partner dimensions. In later iterations, new dimensions were added. After the fourth iteration, the final taxonomy comprises of 18 dimensions, which is 2 less than the initial number of dimensions. As the number of total dimensions did not increase excessively, the taxonomy was found to be concise throughout the entire development process.
2. In the first two iterations, dimensions from the existing literature and artifacts were considered. During the empirical analysis, it was found that some data marketplaces are more advanced than others in terms of the additional services they offer. Therefore, in the third iteration, the additional services dimension was added, with sub-dimensions *enterprise data marketplace* and *data processing and analytics tools*. After the addition of these new dimensions the taxonomy was found to be more robust and able to differentiate between regular data marketplace and more advanced intermediary platforms.
3. Throughout the entire iterative design process, the taxonomy proved to be extendible by allowing the addition of new dimensions in the multiple iterations. In the empirical analysis, 40 objects were considered, as this was found to be a representative sample. If more data marketplaces were to be considered in future research, or new data marketplaces emerge, the taxonomy would still allow the addition of new dimensions and characteristics. For instance, if new data marketplaces emerge in a specific industry, a new characteristic can be added to the industry domain dimension.
4. The dimensions derived from the literature ensured the taxonomy to be explanatory. The revision and addition of new dimensions and characteristics improved the ability to use the taxonomy to explain the different data marketplace cases. In this research, the explanatory power of the taxonomy will be demonstrated by the illustrating the use of the taxonomy on a number of existing cases.

Table 10: Checking of ending conditions after every iteration (own representation based on Nickerson et al., 2013)

	Ending conditions	1st: c2e	2nd: c2e	3rd: e2c	4th: e2c
Objective	All objects, or at least a representative sample of objects have been examined	X	X	X	X
	In the last design iteration no object was merged with a similar object or split into multiple objects	X	X	X	X
	For every characteristic of every dimension in the taxonomy at least one object is classified		X	X	X
	In the last design iteration no new dimensions or characteristics were added to the taxonomy		X		X
	In the last design iteration no dimensions or characteristics were merged or split				X
	There is no duplication of dimension in the taxonomy; no dimension is repeated and every dimension is unique	X	X	X	X
	There is no duplication of characteristics within the dimensions; every characteristic is unique within its dimension	X	X	X	X
	There is no combination of characteristics and there is no cell duplication in the taxonomy; each cell is unique and there is no repetition of cells	X	X	X	X
Subjective	The taxonomy is concise; the number of dimensions make the taxonomy to be meaningful, but not overwhelming	X	X	X	X
	The taxonomy is robust; the dimensions and characteristics of the taxonomy enable researchers to sufficiently differentiate the objects of interest			X	X
	The taxonomy is comprehensive; the taxonomy can be used to classify a sample or all objects within the research domain	X	X	X	X
	The taxonomy is extendible; new dimensions or characteristics of an existing dimension can easily be added	X	X	X	X
	The taxonomy is explanatory; the identified dimensions and characteristics can be utilized to explain about an object	X	X	X	X

6. Taxonomy of data marketplace business models

In this chapter, the taxonomy of data marketplace business models that resulted from the iterative taxonomy development process is presented and discussed. The final taxonomy consists of four meta-dimensions, 17 dimensions and 59 characteristics, and is presented in Table 11. In the following sections, the business model dimensions and characteristics are discussed per meta-dimension. Per characteristic, examples are provided of existing data marketplaces that have applied the business model characteristic. An overview of all identified business model characteristics of the sampled empirical cases are presented in Appendix II.

6.1. Service domain

The business model dimensions in the service domain are the starting point of the STOF approach to business model design (Bouwman et al., 2008). 'Value' is the central issue in the service domain: the data marketplace provider and other service providers intend to collaborate as an enterprise to deliver a certain value proposition to the customer, that in turn expects or perceives a certain value from the value offering. The intended value of the service offered by the data marketplace is summarised in the *value proposition*, and sets requirements for the technical architecture and value network in other domains of the business model (Bouwman et al., 2008). After having specified the value proposition, the *marketplace participants* are specified. Marketplace participants are the customers or end-users of the service. Participants on the data marketplace are active in a certain *industry domain* and within a certain *geographic scope*, which may affect the perceived and expected value of the service. The data marketplace may offer a number of *additional services* to enhance the value of the service offering for marketplace participants. The data marketplace may offer an *enterprise data marketplace* as a service, which is a private data marketplace that offers organizations the opportunity to securely share data within the organization or with selected external partners in the value chain. Lastly, the marketplace may offer *data processing and analytics tools* as an additional service on top of the data, for marketplace participants to process and analyse their proprietary data or data offered on the marketplace.

6.1.1. Value proposition

The value proposition is a statement that indicates the proposed value that an enterprise intends to deliver to the customer (Bouwman et al., 2008). It often describes how customers can benefit from using the service and how the enterprise aims to set itself apart from the competition. Organizations usually present their value proposition on the first page of their website, to clearly communicate their intended value to the customer. Data marketplaces can be characterized by five value propositions: *easy data access and/or tooling*, *secure data sharing*, *high quality and unique data* and *all services in a single platform*.

Data marketplaces that propose to deliver value by offering *easy data access and/or tooling* often have an easy registration process that enables any company or individual to sign up for the data marketplace to start exchanging data. The and/or part of this characteristic indicates that not all data marketplaces offer tooling or easy access to tooling on top of the data offering on the marketplace. An example of a data marketplace that adopts this value proposition is Datahub, which have the goal to make the use and sharing of data and insights faster, easier and more reliable.

Secure data sharing is a value proposition often proposed by data marketplaces that make use of distributed ledger technologies, personal data anonymization and smart contracts to ensure the consensual sharing of data. For instance, the value proposition of the automotive data marketplace Otonomo is to enable car data sharing that complies to strict privacy and security standards. In order to fulfil this value proposition, Otonomo has adopted a neutral server that enables connected car owners to take control over the sharing of their vehicle data with third-party service providers.

High quality and unique data is proposed by data marketplaces that own or have access to unique datasets, that may be acquired or self-generated. Examples of data marketplaces that adopt this value proposition are SimilarWeb and Red Lion Data, that use web crawlers to gather business and geospatial data about companies, such as the number of clicks on the companies' website or the locations of a companies' distribution centres. The crawled datasets are then aggregated to ensure a high data quality.

Data marketplaces that aim to provide value by offering *all services in a single platform* present themselves as a marketplace that incorporates multiple services, often for different types of stakeholders in the industry domain. For instance, the geo data marketplace CARTO present themselves as a one-stop shop for data buyers and sellers. On the data buyer side, CARTO offers a wide range of different datasets and analytics tools for data scientists, data analysts and developers, and on the data seller side the company partners with leading data providers such as Mastercard and Vodafone. Furthermore, audience data marketplaces such as BidTheatre offer a demand side platform (DSP) for marketeers that incorporate a data management platform to handle companies' proprietary and enable automatic buying of media and audience data on an ad exchange.

6.1.2. Enterprise data marketplace

Some data marketplaces offer an *enterprise data marketplace* as an additional service. An enterprise data marketplace, sometimes also referred to as 'data exchange', functions as a private data marketplace that enables organizations to share data within the company or with external partners, such as suppliers, customers and other players that are invited to the platform by the focal organization. Marketplace participants are able to present data sets in a shared environment for other participants to use. This way data sharing is shifted from a demand-based model, in which departments and partners have to make requests for data, to a supply-based model, wherein the datasets available for sharing are presented on the platform. For some companies, the enterprise data marketplace is their main offering, and they incorporate external data in the data marketplace for data enrichment and analytics activities within the marketplace environment. Prominent companies that offer an enterprise data marketplace are DAWEX, Snowflake and Data Republic.

6.1.3. Data processing and analytics tools

The data processing and analytics tools characteristic refers to the tooling that is offered on top of the data, often in a workspace environment, where data and tooling buyers can perform analytics activities on their proprietary data or data bought from the platform. Some companies offer a large variety of tools on top of their data, such as Data Intelligence Hub (by T-Systems),

while other companies do not offer tooling, such as Red Lion Data, focussing solely on the data offering on their marketplace.

6.1.4. Marketplace participants

The specification of users or customers is a key element in business models (Bouwman et al., 2008). Data marketplaces can choose to direct their platform to individual consumers or businesses on both the supply-side and the demand-side (Fruhwirth et al., 2020). Three types of variations of marketplace participants are distinguished: business-to-business (B2B), consumer-to-business (C2B) or any combination of business and consumers (Fruhwirth et al., 2020).

B2B data marketplaces direct themselves specifically to organizations and businesses that are willing to become more data-driven or possess a large amount of data that they wish to monetize or commercialize. For instance, the data marketplace Veracity (by DNV) offers a B2B data marketplace for the exchange of data among companies in the maritime, oil and gas, and energy and renewable industries.

Many C2B data marketplaces act as harvesting data marketplaces, that gather the personal data of users in exchange for rewards. Datax is a C2B data marketplace that gathers label data by asking users to annotate images, collect recordings and classify dialogues and sentiment in exchange for rewards. The crowdsourced data is then sold to businesses for data labelling, AI model training and automation purposes.

Some data marketplaces are open for any party, business or consumer, to register and exchange data on the marketplace (Schomm et al., 2013). For example, Knoema offers a free version of its statistical database, but also has options for professionals and enterprises with more features.

6.1.5. Industry domain

Based on the analysis of empirical cases, data marketplaces are providing their data goods and services in the following industry domains: any data, geo data, financial and alternative data, health and personal data and sensor and mobility data.

A number of data marketplaces allow the exchange of *any data* on their marketplace. An example of a data marketplace that provides the exchange of any type of data is Databroker (rebranded from Databroker DAO in 2019), that extended its scope from IoT data to all types of data, with the goal to be the to-go-to marketplace for data.

Geo data refers to data that has a link with a location on the Earth. Geo data is sometimes also referred to as geospatial or geographical data. This type of data is often stored and used in geographical information systems (GIS). Examples of *geo data* include but are not limited to environmental data, housing data, road data, weather data, business locations and static or dynamic maps. Prominent examples of data marketplaces that have a main focus on geo data are the HERE Platform and the CARTO Location Intelligence platform.

The *finance and alternative data* industry domain refers to data marketplaces active in the financial industry. Finance data comprises of datasets that provide information about the financial state of a company, such as data about a companies' assets, liabilities and equity. Alternative datasets provide information about a company that is published by sources outside of the company. Examples of alternative data are credit card transactions, website usage, product reviews and price trackers. Alternative data may provide unique insights about investment opportunities. Both financial and alternative data are used by investment professionals such as hedge fund managers, venture capitalists, private equity funds and investment bankers to make investment or divestment decisions. It was found that the data offering of financial data marketplaces often overlap with the offering of alternative data marketplaces, and therefore these two data industry domains were merged.

Data marketplaces in the *health and personal data* industry domain often function as harvesting data marketplaces that provide rewards to customers for providing their health or personal data. Health data refers to e.g. patient names, birth dates, medical treatments and health conditions of individuals or the population. Examples of personal data are name, sex, age, home address and income. Some dimensions of health and personal data overlap, and as many function as harvesting data marketplaces, these two types of data industry domains were merged.

Audience data is combined data about a certain target group of customers, the 'audience'. Marketeers aim to gather as much data about their envisioned audience as possible, to target the audience with highly personalized and relevant offers. In many cases, audience data is gathered by a data provider company through the automatic or manual scanning of user behaviour on websites and mobile applications. For instance, users that accepting the browser cookies of a companies' website agree to share their clicking behaviour, timestamps and geographic location. This gathered data is then sold to marketers through an audience data marketplace. Examples of audience data marketplaces are the Amazon DSP, Salesforce Audience Studio and BidTheatre.

Data marketplaces in the *sensor & mobility data* industry provide sensor data gathered by Internet-of-Things sensors, such as smart city data, traffic data, parking data and automotive data. The data offered in this industry domain is often (near) real-time, because the sensor data is directly sent to the data marketplace by transferring the streaming data via APIs. Two well-known data marketplaces in this industry domain are Otonomo and Caruso, that provide data for B2B service providers and the automotive after-market, such as workshops, vehicle part manufacturers and insurance companies (Bounie et al., 2018).

6.1.6. Geographic scope

The geographic scope describes the regions in which the data marketplace is operating and available to users. A distinction is made between global data marketplaces, regional data marketplaces and local data marketplaces (Täuscher & Laudien, 2018; Täuscher, 2016). Global marketplaces serve clients across two or more continents. Examples of global data marketplaces are DAWEX, BattleFin Ensemble, SimilarWeb and CARTO. Regional data marketplaces focus on multiple countries in a single continent or region. For instance, BidTheatre has a focus on the European media landscape, Data Republic is the leading platform in the Asia Pacific region, and Red Lion Data is providing location data in North

America. Lastly, local marketplaces solely focus on a single country. For example, Mobility Data Marketplace is a platform for the exchange of German mobility data, and Marketscan and oneTRANSPORT focus on trading data in the United Kingdom.

6.1.7. Time frame

The data traded on the data marketplace may have a certain temporal context in a *time frame*, that describes whether or not the data needs frequent updates to maintain the relevancy of the data (Schomm et al., 2013). A distinction is made between static datasets, up-to-date datasets, (near) real time datasets, and data marketplaces that offer datasets with multiple time frame relevancies. An example of a static data are the labelled datasets sold by Datax, that crowdsources data by asking consumers to label images, recordings and dialogues. Up-to-date datasets are essentially static datasets, that are repeatedly updated by the marketplace provider or the external data sellers on the data marketplace. For instance, Datahub offers a list of 'core datasets', that are regularly updated by the marketplace provider. A number of data marketplaces offers real-time or near-real time data. This type of data is often generated by IoT sensors or online data trackers, such as website and stock market trackers. IOTA is a data marketplace that enables the exchange of (near) real-time sensor data between devices and machines.

6.2. Technology domain

The requirements specified in the service domain determine the identification and specification of the technical architecture in the technology domain of the business model (Bouwman et al., 2008). The *platform architecture* dimension describes whether the platform makes use of a centralized or decentralized architecture. Both architectural designs have pros and cons with regard to platform control and data provenance. Furthermore, the marketplace provider may provide *data access* to in different ways, such as via direct download or API. Lastly, the datasets offered via the data marketplace originate from a certain *data source*.

6.2.1. Platform architecture

Data marketplaces may adopt two types of platform architectures: centralized and decentralized (Koutroumpis et al., 2017). In the centralized approach, data providers offer their data products via a predefined location central on the platform, such as a cloud repository. This type of platform architecture provides better control over data access and enables data buyers to directly process the data. In decentralized platforms, the data products remain at the data provider and the data is traded using distributed ledger technologies (Koutroumpis et al., 2017).

The decentralized approach enhances data provenance, but makes data processing and storage more challenging for the platform users (Koutroumpis et al., 2017). Examples of centralized data marketplaces that store the data in a central cloud repository are Opendatasoft, Mobility Data Marketplace and oneTRANSPORT. Decentralized data marketplace architectures are emerging. Two examples of companies that have adopted such an architecture are the Data Intelligence Hub and Snowflake Data Marketplace. The Data Intelligence Hub has implemented the security standards by the International Data Spaces Association (IDSA). The platform has adopted a decentralized approach, in which data is

transferred directly from data sellers to data buyers through a secured line. This way, the data never passes through the Data Intelligence Hub platform itself. A similar approach is adopted by Snowflake Data Marketplace: by implementing Snowflake's Secure Data Sharing technology, the data offered on the marketplace does not move from data provider to data consumer, but remains at the data provider. No data is transported, pushed through APIs or stored in a cloud. Instead, data consumers have direct 'read-only' access to the datasets of the data provider. The data providers in turn have control over who can access their data.

6.2.2. Data access

Platform providers may provide access to the data in a number of different ways, (Schomm et al, 2013): via APIs, direct download options, specialized software or via multiple of the aforementioned options. Data marketplaces that offer data access via APIs develop a predefined software protocol to establish an interface that enables access and interaction with the platform. In the download option of data access, the data is accessed via a download file and there is no need for developing a software component. Some data marketplaces develop specialized software to provide access to the data on the marketplace. A large number of data marketplaces in the sample offered multiple options to access the data, either via APIs, direct download options and specialized software.

6.2.3. Data source

The data source dimension describes the origin where the data was gathered or collected by the data marketplace platform (Hartmann et al., 2014). The following data sources are distinguished for data marketplaces: self-generated data, customer provided data, acquired data or data from multiple of the aforementioned sources. Data marketplaces may have generated data themselves, by for instance gathering data manually or automatically from the internet. For instance, the company Factual gathers data from many sources about places, businesses and landmarks that data consumers may want to buy. All the gathered data is sucked in an agile machine and cleaned and restructured into valuable datasets. Furthermore, the data marketplace may also invite customers to provide their proprietary datasets on the platform. An example of a data marketplace that exclusively offers data provided by external data providers is Veracity. This company aims to function as a neutral intermediary between companies for the sharing of data and applications. Moreover, Marketscan is an example of a data marketplace that acquires data from external data providers. The company integrates data from the feeds of five large UK data suppliers, and then verifies and aggregates the data in the central database to ensure high data quality and coverage. Lastly, some data marketplaces retrieve data from multiple types of sources. Data Intelligence Hub, a data marketplace that aims to offer all sorts of data, retrieves data from open data portals and publishes them on the platform, but also invites commercial data providers to sell their data offerings on the marketplace.

6.3. Organization domain

The technologies that are used to deliver the service to the customer depend on the organization design of actors that take ownership and invest in these technologies (Bouwman et al., 2008). Central in the organization domain of the business model is the value network of actors that is needed to realize the service offering. The data marketplace is sponsored by a

platform sponsor, that designs and holds the intellectual property rights of the platform (Eisenmann et al., 2009). Actors perform value activities in the network to deliver the proposed service to customers and end-users. The value activities put requirements on the technical architecture, ask for specific investments and may generate organizational costs. One of the main value activities that performed by the data marketplace provider is the matching of marketplace participants to foster transactions and innovation. The *matching mechanism* determines the number of parties on each side of the platform that are matched by the data marketplace.

6.3.1. Matching mechanism

The matching mechanism of a data marketplace determines the number of parties on each side of the platform (Koutroumpis et al., 2017). With regard to data marketplaces, the following variations of matching exist: one-to-one matching (bilateral marketplace design), one-to-many (dispersal marketplace design), many-to-one (harvest marketplace design), and many-to-many (multilateral marketplace design).

One-to-one matching mechanisms can be characterized by negotiated terms of exchange (Koutroumpis et al., 2017). Marketscan is an example of a data marketplace that adopts a one-to-one matching mechanism. Marketscan provides customized lists of audience data, by combining unique data from multiple sources. For the customized data lists, the marketplace has adopted multiple pricing models for data buyers to choose from. Furthermore, Marketscan aims to set itself apart from the competition by providing excellent one-on-one customer service.

Data marketplaces that adopt a one-to-many matching mechanism mediate between a single seller and many buyers (Koutroumpis et al., 2017). This type of data marketplaces are also called dispersal data marketplaces. An example of a data marketplace that adopts a one-to-many matching mechanism is CARTO, a location intelligence platform. CARTO aims to make location data accessible for individuals and enterprises, and offers both free and premium access to their platform. CARTO integrates data from multiple sources in a comprehensive data catalog and offers tooling such as Jupyter notebooks and web mapping tools on top of the data. The company has partnerships with some of the leading location data providers, such as mastercard, Vodafone and TomTom.

In many-to-one matching, where many sellers are trading data with a single buyer at the same time, is used in harvest marketplace designs (Koutroumpis et al., 2017). A company that has adopted the harvest marketplace design is BIGtoken, a platform that aims to empower users by providing rewards for the secure and consensual sharing of their personal data. BIGToken collects the personal data of users, with the goal to sell the data to advertisers. Users can choose which company has access to their data.

Data marketplaces that adopt the many-to-many matching model, allow any user to upload and maintain datasets on the platform (Schomm et al., 2013). A major data marketplace that adopts this matching mechanism is DAWEX, a global data marketplace that acts as a neutral intermediary between data buyers and data sellers. The company invites individuals and organizations to buy and sell their data in a direct and secure way.

6.3.2. Platform sponsor

The platform sponsor constitutes and holds the property rights of the platform components, rules and ecosystem (Eisenmann et al., 2009). The platform can be sponsored by a private individual or group, a consortium of buyers or sellers on the supply or demand side of the platform, or an individual or group that is independent of other market players (Stahl et al., 2016).

An example of a data marketplace with a private sponsor is Informatica B2B Exchange. Informatica is a software development company, with its proprietary software platform as main resource. In 2015, the company was acquired for $5.3B by Permira, a European private equity firm (Permira, 2015).

The HERE Platform is an example of a data marketplace that is sponsored by a consortium of data buyers. HERE is a provider of location data and platform provider, with a service offering comprising of a development workspace, data marketplace, and map creation and visualization tools. The technology company is invested in by some main shareholders in the automotive industry, such as Audi, BMW and Daimler, and other engineering and service suppliers such as Bosch, Continental, Intel and Pioneer. Last year, HERE welcomed Mitsubishi as a major new shareholder, that took 30% ownership of the technology company (HERE, 2019). By investing in HERE, the big automotive and engineering companies profit from the R&D abilities of the company and its experience with and knowledge of location data, to develop solutions such as autonomous driving.

An example of an independent platform sponsor is oneTRANSPORT. The company aims to function as a neutral infrastructure provider in the UK data market. ONE transport is a private limited company, and positions itself as a neutral handler of data on the marketplace, and as a neutral party that facilitates the exchange of data and services between organizations on the platform.

6.4. Finance domain

In the finance domain of the STOF model, the financial arrangements between the different participants in the value network are specified (Bouwman et al., 2008). The value activities and technological architecture in the organization and technology domain are costs sources from the supply side of the service that affect the financial domain. Viable business models contain a balance between financial risks and benefits for the stakeholders involved in the value network. The finance domain therefore provides a description about how the network of actors intends to capture value. The *revenue model* depicts whether financial revenue comes directly from the buyers, or whether there are also other sources of revenue for the value network. The final price for the data good or service is specified by the *pricing model* of the data marketplace. The *price discovery* function of a data marketplace describes how and by who the prices of the goods and services on the data marketplace are set. To provide safe payment, data marketplace may offer *smart contracts* that are enabled by blockchain. This may enhance privacy and trust among marketplace participants. Lastly, cryptocurrencies are emerging as an alternative *payment currency* to fiat money, as a way for marketplace providers to securely handle payments between data buyers and data sellers.

6.4.1. Revenue model

Financial revenue may come directly from the buyer of the good or service, but there are also other main sources of revenue for an enterprise (Bouwman et al., 2008). Five revenue models for data marketplaces are distinguished: the commission model, subscription model, usage fee model and asset sales model. In the commission or transaction fee model, the data marketplace receives a certain fee for every transaction that takes place on the platform (Spiekermann, 2019; Täuscher & Laudien, 2018; Täuscher, 2016). Examples of data marketplaces that adopt commission fees as a main revenue model are Data Intelligence Hub, Otonomo and Caruso. In the subscription model, the data marketplace signs a contract with platform users to provide a specific service for a recurring fee (Täuscher & Laudien, 2018; Täuscher, 2016). The subscription revenue model is adopted by among others Intrinio, HERE Marketplace and SimilarWeb, often in combination with a freemium pricing model. In the service sales model, the data marketplace sells services that are not standardly offered to all users (Täuscher & Laudien, 2018; Täuscher, 2016). Data marketplaces may charge are fee for the usage of their platform or services. For instance, Snowflake charges usage fees based on the cloud storage and computational resources used by customers. A similar revenue model is adopted by Opendatasoft and Datum Data Marketplace. In the asset sales revenue model (Osterwalder & Pigneur, 2010), the main source of revenue comes from the sales of data goods. For instance, the revenue of the location data marketplace Red Lion Data depends on the sales of their proprietary packages of data lists.

6.4.2. Pricing model

The pricing model specifies how the final price for the data good or service is composed. From the empirical analysis, it was found that data marketplaces employ seven types of pricing models: freemium, pay-per-use, flat fee tariff, package based pricing, and a combination of multiple of the aforementioned pricing models.

In the freemium model the data marketplace provides basic functions for free, but marketplace users will need to pay a fee to make use of the premium functions (Fruhwirth et al., 2020; Spiekermann, 2019; Täuscher & Laudien, 2018; Täuscher, 2016). For instance, Knoema offers a free version of their data solutions for potential customers to try out simple functions of the platform, but also hosts a professional and enterprise version of their data platform solution, with premium functions such as data enrichment and visualization tools.

In pay-per-use or usage based pricing, customers pay a price that is proportional to the amount of units consumed by the data marketplace user (Fruhwirth et al., 2020; Spiekermann, 2019). For instance Rollworks, an account-based platform and audience data marketplace, charges users credits for each contact they redeem via the platform. Marketplace users can buy credits with fiat money to buy more contacts. Unused credits will still be usable the next month after purchase. If marketplace users do not buy any contacts in a given month, no extra fee is charged.

The flat fee tariff or flat rate pricing model provides marketplace participants full access to the marketplace for a recurring fee (Fruhwirth et al., 2020; Schomm et al., 2013). The alternative data marketplace BattleFin Ensemble adopts a flat fee tariff pricing model. For a monthly fee, technical users of the marketplace may test and evaluate a certain amount of data. The flat

fee tariff is as follows: testing and valuating datasets up to 100 GB for technical users at $2.500 per month, and over 200 GB of dataset testing and evaluation for premium technical users at a price of $10.000 per month.

In the package based pricing model, data goods or services are bundled in certain packages, of which the price may decrease by a certain discount rate when the size of the package increases (Fruhwirth et al., 2020; Spiekermann, 2019; Schomm et al., 2013). As an example, Red Lion Data sells packages of data lists against a certain discount level. The larger the data package (e.g. 10, 25 and 100 lists), the higher the discount.

On some data marketplaces, the pricing of the data products and services are based on multiple pricing models. For instance, users of Marketscan Online can choose between three types of pricing models to pay for the audience data on the marketplace: a 12 month data package (package based pricing), via data credits from pre-specified credit packages (package based pricing) or pay as you go (pay-per-use pricing).

6.4.3. Price discovery

A price discovery function allows buyers and sellers on the marketplace to determine a transaction price which they both agree on (Bakos, 1998). Data marketplaces make use of price discovery mechanisms to determine the price of a dataset before it is transacted on the platform: prices set by data buyers, discovery by negotiation, prices set by the marketplace provider and prices set by external sellers.

The data marketplace may decide to let data buyers set the prices for the datasets they wish to buy. The harvesting data marketplace BIGToken aims to gather user data and information with the goal to selling it to advertisers. In this case, the data marketplace functions as buyer. BIGToken asks users to participate in brand and product surveys in return for rewards in the form of points, that can be exchanged for PayPal money or gift cards. The rates at which the earned points are exchanged is specified by the data marketplace.

In the negotiation model, data marketplaces may allow data buyers and sellers to negotiate about the price before coming to an agreement. For example, Datum Data Marketplace allows data buyers to send a data purchase request to users to buy a copy of their encrypted personal data. A purchase request comprises of details about the purchaser and the proposed price set by the data buyer. Users can agree to the proposed purchase price or send a counter offer to the data buyer. This way, the negotiation process about the data price takes place.

The data marketplace provider may also decide to take charge of setting prices for the data goods and services on the platform. For instance Intrinio, a finance and alternative data marketplace, sets the prices of datasets offered on the data marketplace based on the usage of the data feed by data buyers. The prices for individuals and startups are listed on the platform, and enterprise clients may contact the service team to request the pricing for a certain data feed.

Lastly, the data marketplace may allow external sellers to set the prices for their own data offering on the marketplace. Streamr, a decentralized data platform for the exchange of real-time data, invites external data providers to monetize their real-time data. The third-party data

sellers can integrate their data and create offerings on the marketplace to start earning the DATA cryptocurrency. The data providers are free to set their own product description and prices. Streamr has created instruction videos for the onboarding of external parties.

6.4.4. Smart contract

Data marketplaces may implement smart contracts to enhance transparency and to enforce trust among marketplace participants (Fruhwirth et al., 2020). A smart contract comprises of an contractual agreement that is coded into a script that is automatically executed when the terms in the contract are met. The use of smart contracts by data marketplaces is emerging as a way to introduce transparency and to automatically handle payments made on the marketplace (Lawrenz et al., 2019). DAWEX is an example of a company that has implemented a smart contract in the Ethereum blockchain for the exchange of data on its data marketplace.

6.4.5. Payment currency

The payment currency dimension explicates which currencies are accepted by the data marketplace for the payments that are made by data buyers on the platform (Fruhwirth et al., 2020). Data marketplaces may handle their payments via cryptocurrencies or fiat money. Data marketplace companies that use cryptocurrencies as a payment method are emerging. Examples of marketplaces that offer payment in cryptocurrency are IOTA and Streamr, that both have developed their own coin, the IOTA and DATA.

The final taxonomy of identified dimensions and characteristics of data marketplace business models is presented in Table 11.

Table 11: Taxonomy of data marketplace business models

	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			

7. Demonstration of the use of the taxonomy

In this chapter, the use of the taxonomy is demonstrated by applying the derived business model dimensions and characteristics to three empirical examples of data marketplaces. Demonstration on the basis of empirical illustration is employed in a number of other taxonomy development studies (Azkan et al., 2020; Bock & Wiener, 2017). In this study, three mini case studies are conducted to showcase how to use the taxonomy for the classification of business models of data marketplace companies. The three selected cases are part of the database of data marketplaces that was established in this study, but they were not included in the taxonomy development process. Therefore, the demonstration proves that the taxonomy can be used for data marketplaces outside of the sample of cases. The demonstration of the taxonomy can help researchers and practitioners to understand how to use the taxonomy for the classification and design of data marketplace business models. At the end of the chapter, the key take-aways from the demonstration are discussed.

7.1. Selected empirical cases

Three data marketplace companies are selected from the sample of empirical cases to demonstrate the usefulness of the taxonomy. The three cases are part of the database of data marketplace companies that was established during the taxonomy development process, but they were not part of the sample of 40 cases.

As the two existing taxonomies of data marketplaces mainly focussed on the classification of multilateral data marketplaces (Fruhwirth et al., 2020; Spiekermann, 2019), this study aimed to go beyond the state of the art by developing a taxonomy that is also suitable for classifying other types of data marketplaces, based on their matching models and marketplace design. Therefore, the three selected companies differ in terms of their marketplace design: a bilateral data marketplace, harvest data marketplace and multilateral data marketplace (Koutroumpis et al., 2017).

The first selected company is Wibson, a bilateral marketplace (one-to-one matching) in the personal data industry. QueXopa, an alternative data marketplace focussed on the Latin American market that has that has adopted a dispersal marketplace design (one-to-many matching), is the second selected company for the empirical illustration of the taxonomy. The third selected company is Advaneo, a multilateral data marketplace (many-to-many matching). Information on the business models of the selected data marketplaces was derived from desk research that was conducted to gather information on the empirical cases. Main sources of information were the company websites of the selected cases, white papers, terms and conditions and news articles about the data marketplace companies. For the company Wibson, a scientific whitepaper by Fernandez et al. (2020) was available. Table 11 presents a summary of the business model characteristics of the selected cases.

7.2. Bilateral data marketplace: Wibson

Wibson is an example of a bilateral data marketplace that enables personal data trading between individuals and organizations. The company offers a decentralized data marketplace that makes use of smart contracts to enable inviduals to securely and anonymously share data

in a trusted environment (Fernandez et al., 2020). Wibson is active in the personal data industry domain. The company provides an infrastructure for individuals to share information with data buyers. The company has implemented smart contracts to arrange the secure selling of data between data buyers and data sellers (Fernandez et al., 2020). Hence, data is provided by customers on the data marketplace. The individuals are in control of their personal data, and are able to monetize their personal data by giving organizations data access in return for money. This way, Wibson adopts a one-to-one matching mechanism, where the company itself functions as neutral intermediary that provides the blockchain infrastructure for the data exchange. Prices on the marketplace are set by data buyers, and buyers are matched with individual data sellers that are willing to sell their personal data for the set price. Wibson has implemented smart contracts to handle payments, and transactions are paid in cryptocurrencies, that can be exchanged with Wibson in return for fiat money (Fernandez et al., 2020).

7.3. Dispersal data marketplace: QueXopa

QueXopa presents itself as the single source for Latin American alternative data. The company aims to set itself apart from the competition by finding, sourcing and aggregating alternative data, to provide exclusive, high quality and accurate alternative data. Customers of QueXopa are finance professionals such as investors, hedge funds, market analysts, retailers and corporations. The company is active in the alternative data industry domain, and provides credit card transactions, mobile location data, insurance policies, real estate listings, mobile app metrics, price monitoring, email receipts and maritime and port data from Latin American sources. This type of data is mostly static, and provides a snapshot of the moment of measurement of the data. On the one hand QueXopa sources data from major Latin American data providers, and on the other hand the company generates data themselves by conducting equity research and data scraping from websites. QueXopa makes use of the one-to-many matching mechanism, as it aims to sell its proprietary high quality alternative data to a multitude of governmental and industrial finance professionals. The company offers alternative data via its website, but also provides on custom tailored on demand data on request. Customers can subscribe to the data offerings of the company, and pricing is dependent on the frequency, history, quantities and regions of the data. The prices of the data are set by QueXopa itself, and the company handles payment in fiat money.

7.4. Multilateral data marketplace: Advaneo

Advaneo is a data marketplace that aims to provide easy data access and tooling. The company offers a data science workbench on top of their data offering, that consists of a Jupyter Notebook operated via the Advaneo cloud. Furthermore, Advaneo offers an enterprise data marketplace solution in the form of Closed User Groups. This function of the data marketplace allows marketplace participants to control who has access to the proprietary datasets of participants, and allows organizations to invite both internal as well as external users to participate in projects. The Advaneo marketplace is open for any individual or organization to join, and offers four different member ship models: free, premium, small business and enterprise. The company aims to foster cross-domain innovations, and therefore any type of data is traded on the platform. The website is available in 16 languages, which shows that Advaneo is open for global reach. The marketplace offers open data as well as

commercial data, and while the number of datasets and portal on the platform is increasing, Advaneo conducts regular updates and maintenance of the datasets offered on the platform. Advaneo has adopted a decentralized architecture design, in which the data traded on the marketplace is transferred directly from the data seller to the data buyer through a secured line, without touching the platform. This way, Advaneo aims to take on a neutral intermediary position in the data market. The company offers multiple forms of access to the data, such as acces via API and access via specialized software (IDS connector). The main source of revenue for Advaneo are subscriptions, that vary depending on the type of membership of the marketplace. The company also offers a freemium option, that allows users to test data, build basic visualizations, trial the IDS-connector and use the workbench for up to 15 GB data. Prices for the data offerings on the marketplace are set by external data sellers, paired with a data license agreement. The marketplace also includes open data that is offered for free. Advaneo offers the possibility to pay in fiat money by credit card. Other digital options such as Apple pay are currently being built.

7.5. Take-aways from the demonstration

The objective of this study was to design a taxonomy of data marketplace business models that could function as a tool for the classification of business model characteristics of data marketplaces. The taxonomy development method by Nickerson et al. (2013) was used to develop the taxonomy. The final taxonomy is presented in Table 11 in Chapter 6, and comprises of four meta-characteristics and 17 business model dimensions to classify data marketplace business models. On the basis of three mini-cases studies, the usefulness of the taxonomy is demonstrated in this chapter. Table 12 provides an overview of the business model characteristics that were identified for the selected cases with the use of the taxonomy. The three cases varied based on their marketplace design.

As can be seen in Table 12, most of the business model characteristics of the three selected data marketplaces could be classified with the use of the taxonomy. The cases were classified based on publicly available information on the websites of the companies, white papers, news articles and academic studies. If the information sources did not provide sufficient information about the business model characteristics of a selected case, this is denoted by *no info* in the table. For Advaneo, sufficient information was available to classify its business model characteristics on all business model dimensions of the taxonomy. The other two data marketplaces could not be classified on all characteristics because no sufficient information was available about the *geographic scope* and *revenue model* of Wibson, and the *platform architecture* and *data access* of QueXopa. The fact that no information was available about certain business model characteristics does not entail that there is no information about these characteristics at all. The information may be retrievable when questioning the owners of the respective data marketplaces. However, in this demonstration of the taxonomy, only publicly available sources were considered. Thus, it can be concluded that the taxonomy is useful to classify the business models of data marketplace when sufficient information is available about the respective business model characteristics of the companies.

In this study, the term data marketplace was broadly interpreted, as the digital infrastructure on which the commercial trade of data takes place, and where the data is traded as a valuable good. While existing taxonomies mainly focus on multilateral data marketplaces (Fruhwirth et

al., 2020; Spiekermann, 2019), this studies' interpretation of a data marketplace allows for the inclusion of other types of data marketplaces, such as bilateral data marketplaces, harvest data marketplaces and dispersal data marketplaces.

The demonstration of the use of the taxonomy illustrates that the taxonomy is useful for indicating the differences between the business models of the three cases. For instance, the three cases all adopted a different matching mechanism, that indicates the number of players that are matched on each side of the market. Furthermore, the three cases differ based on the marketplace participants that are active on the platform. Wibson is focussed on C2B transactions, QueXopa focusses on B2B transactions, and Advaneo allows transactions between any marketplace participant, either consumers or business participants. Moreover, the three different types of data marketplaces vary in price discovery. On the Wibson data marketplace, prices are set by data buyers, whereas on QueXopa and Advaneo prices are set by the marketplace provider and external data providers, respectively. Hence, with the use of the taxonomy, a clear distinction can be made between different types of data marketplaces, on the basis of the business model dimensions and characteristics that were incorporated in the taxonomy. This study goes beyond the state of the art by developing a taxonomy that is not only suitable to classify multilateral data marketplaces, but that can also be used to classify data marketplaces with other matching models and other business model characteristics.

Table 12: Illustration of the use of the taxonomy by application to three empirical cases

		Company		
		Wibson	QueXopa	Advaneo
Service domain	Value proposition	Secure data sharing	High quality and unique data	Easy data access and tooling
	Enterprise data marketplace	No	No	Yes
	Data processing and analytics tools	No	No	Yes
	Marketplace participants	C2B	B2B	Any
	Industry domain	Health & personal data	Alternative data	Any data
	Geographic scope	*No info*	Regional	Global
	Time frame	Static data	Static data	Up-to-date
Technology domain	Platform architecture	Decentralized	*No info*	Decentralized
	Data access	Specialized software	*No info*	Multiple
	Data source	Customer provided	Multiple	Multiple
Organization domain	Matching mechanism	One-to-one	One-to-many	Many-to-many
	Platform sponsor	Independent	Private	Independent
Finance domain	Revenue model	*No info*	Subscriptions	Subscriptions
	Pricing model	Pay-per-use	Pay-per-use	Freemium
	Price discovery	Set by buyers	Set by marketplace provider	Set by external sellers
	Smart contract	Yes	No	No
	Payment currency	Crypto	Fiat	Fiat

8. Conclusions

In this chapter, conclusions are drawn from the research results. The main research question that was specified in the Introduction of this research in Chapter 1 is repeated. In order to answer the main research question, a number of sub questions were formulated. In the first section, the objective of this study is repeated and the answers to the sub questions are presented. Based on a synthesis of the answer to the sub questions, the main research question is answered in the second section. Finally, the societal and managerial relevance of the research are discussed, and the relevance to the CoSEM programme is argued for.

8.1 Answering the sub questions of this research

In this section, the answers to the sub questions that were formulated in Chapter 1 are presented. For every sub question, first the question is introduced, followed by the answer to the sub question that resulted from the research process and results of this study.

With the amount of available data growing and data posing as a strategic asset to firms, the data economy has started to evolve (European Commission, 2017). Data marketplaces can fulfil a key role in realizing the data economy. The way a data marketplace operates and conducts business can be mapped and managed using a business model. Data marketplaces are a new area of research, and therefore not much research has been conducted on this type of digital platforms yet (Thomas & Leiponen, 2016), nor on the business models of data marketplaces (Fruhwirth et al., 2020; Spiekermann, 2019). To investigate the business models of data marketplaces, a design science approach was employed (Hevner, 2007) and a standard taxonomy development method was followed to develop a taxonomy of data marketplace business models (Nickerson et al., 2013). The main objective of this study was:

Design a taxonomy of data marketplace business models, using the standard taxonomy development method by Nickerson et al. (2013), that can function as a tool to classify business model characteristics of data marketplaces. The taxonomy shall be developed based on both existing scientific concepts and characteristics founds in empirical cases. The term data marketplace shall be broadly interpreted during the taxonomy development process and the business models of data marketplaces shall be viewed from a multi-stakeholder perspective. Newly set up data marketplaces that have not yet been considered in existing studies shall be considered to take into account novel alterations of data marketplace business model characteristics.

In order to fulfil this research objective, the notion ‘data marketplace business model’ was first explained on the basis of a number of key concepts: business models, digital platforms, electronic marketplaces and data marketplaces. The first research question was:

SQ 1: How can a data marketplace business model be defined?

From a multi-stakeholder perspective, a business model can be viewed as the mapping of how a network of organizations collaborates together as an enterprise to create and capture value (Bouwman et al., 2008). It was argued that the business models of data marketplaces can be viewed from a multi-stakeholder perspective, since these marketplaces are part of a

platform ecosystem that generally consists of a data marketplace provider, data providers, data consumers and third-party service providers (Spiekermann, 2019).

In the light of digital platform theory, a data marketplace can function as a digital intermediary that matches multiple data buyers, data sellers and external service providers (Carnelley et al., 2016). However, it is important to note that not all data marketplaces function like a platform that matches multiple data buyers with multiple data sellers, nor do all marketplaces invite third-party service providers to offer their services on the marketplace. Instead, some data marketplaces may take on the form of a data 'store', where data buyers go to buy data goods and services. In this simple form of a data marketplace, the offerings on the marketplace are developed and provided by the marketplace provider itself, and no external service providers are involved. In a more advanced of data marketplace, the marketplace may function like a multi-sided digital platform, where the marketplace provider takes on the role of a neutral intermediary that matches multiple buyers, sellers and external service providers. In this form, the data marketplace operates as a multi-sided digital platform that allows value creation in the form of transactions and innovation.

Furthermore, the economic notions of markets, marketplaces and electronic markets and marketplaces were introduced, and the differences between them were discussed. Electronic marketplaces provide the digital infrastructure on which marketplace participants interact and where the negotiation phase of the trading process is carried out electronically (Stahl et al., 2016). Electronic marketplaces differ from general marketplaces, because they make use of information technologies to match buyers and sellers (Bakos, 1998). This may result in increased personalization, cost-effective customization of product offerings, decreased search costs for buyers, lower communication costs for sellers, and new ways of price discovery (Bakos, 1998).

In this study, the term data marketplace was broadly interpreted as *a marketplace that provides a digital infrastructure to enable the commercial trade of data as a valuable good*. While the two existing taxonomies studies mainly focus on the classification of business models of multilateral data marketplaces on which anybody can upload and maintain datasets (Fruhwirth et al., 2020; Spiekermann, 2019), in the definition of a data marketplace that is employed in this study, no assumption is made about the number of users on each side of the data marketplace. Therefore, this study went beyond the state of the art by not only considering multilateral data marketplaces, but also other types of data marketplaces, such as bilateral, dispersal and harvest data marketplaces (Koutroumpis et al., 2017). Furthermore, this definition rules out data marketplaces that provide free or open data, because these marketplaces do not adopt for-profit business models with a commercial purpose (Carnelley et al., 2016).

Combining the knowledge from the key concepts, a data marketplace business model was defined as: *The mapping of how a data marketplace enterprise aims to create and capture value by providing a marketplace and additional value adding services for the commercial trade of data between data providers and data buyers.*

Having established a profound understanding of the object of interest, existing solutions in the state of the art research were reviewed. The second sub questions of this study was:

SQ 2: What theories and artifacts are currently available in the state of the start research that classify or are relevant to the classification of data marketplace business models?

A literature review was conducted to investigate the state of the art scientific theories and artifacts on data marketplace business models. The literature review resulted in the discovery of a number of studies that presented dimensions of data marketplaces and data providers (Stahl et al., 2017; 2014a; 2014b; Schomm et al., 2013), a classification framework for electronic marketplaces (Stahl et al., 2016), marketplace designs for data marketplaces (Koutroumpis et al., 2020; 2017) and pricing approaches for data marketplaces (Fricker & Maksimov, 2017; Muschalle et al., 2012). Furthermore, two existing taxonomies of data marketplaces were identified (Fruhwirth et al., 2020; Spiekermann, 2019), as well as a number of business model taxonomies that were found to be closely related to data marketplaces, in the fields of digital marketplaces (Täuscher & Laudien, 2018; Täuscher, 2016), digital business (Bock & Wiener, 2017) and data-driven business models (Hartmann et al., 2014).

Based on the existing concepts from the data marketplace literature and relevant dimensions from the identified business model taxonomies, a preliminary business model taxonomy was constituted. At this point, the preliminary taxonomy comprised of 4 meta-dimensions, 20 dimensions and 77 characteristics of data marketplace business models.

After defining the object of interest and reviewing relevant theories and existing artifacts in the research domain of the object of interest, the taxonomy development process took off. The taxonomy development process started with the establishment of the meta-characteristics of the taxonomy and the definition of requirements for the ending conditions of the iterative design process (Nickerson et al., 2013). Therefore, the third sub question was:

SQ 3: What are the meta-characteristics of data marketplace business models and what ending conditions terminate the taxonomy development?

The meta-characteristics set the basic conditions which the taxonomy is built upon (Nickerson et al., 2013). The STOF ontology by Bouwman et al. (2008) was selected as the leading perspective on data marketplace business models in this study. The STOF ontology takes service as the unit of analysis and employs a multi-stakeholder perspective on business models (Bouwman et al., 2008). This approach suits data marketplace business models, because data marketplaces are active in an ecosystem of multiple stakeholders, among which the data marketplace provider, data buyers, data sellers and third-party service providers (Spiekermann, 2019; Chakrabarti et al., 2018; Thomas & Leiponen, 2016; Muschalle et al., 2012). Furthermore, the STOF ontology is appropriate for data marketplace business models because the main aim of a data marketplace company is to provide a marketplace as a service for the commercial trade of data between data buyers and data sellers (Koutroumpis et al., 2020). Additional services such as data processing and visualization services may be offered on top of the data, to enhance the value for marketplace participants (Spiekermann, 2019). Therefore, the four dimensions of the STOF model were selected as relevant meta-characteristics for the characterization of the business model of data marketplaces: the service domain, technology domain, organization domain and finance domain (Bouwman et al., 2008).

The eight objective ending conditions and five subjective ending conditions that are suggested by the authors of the selected taxonomy development method were employed to terminate the iterative taxonomy development process (Nickerson et al., 2013).

SQ 4: What business model dimensions and characteristics can be derived from the business models of existing data marketplaces and how can these be added to or revise the preliminary taxonomy?

Based on desk research for empirical cases, a database of 178 data marketplaces was compiled. Most of the cases were derived from the website datarade.ai, a profit-based platform that aims to match data buyers with data providers by providing a repository of over 1800 data providers and more than 200 data platforms. It was found that the majority of the data marketplaces in the database were active in the audience data industry domain (N=112). Therefore, the disproportionate stratified sampling method was employed to ensure that the sample included a variety of data marketplaces based on the type of data traded on the platform. A set of N=40 cases was sampled from the database using the disproportionate stratified sampling method (Daniel, 2011).

Subsequently, a within case analysis was conducted on the selected cases in the sample by searching for information about the business models of the companies. The dimensions and characteristics in the preliminary taxonomy were applied to the empirical cases in two conceptual-to-empirical design iterations (Nickerson et al., 2013). This resulted in a number of revisions of the taxonomy: 9 characteristics were merged or split, 5 dimensions were removed and 7 characteristics were deleted. Followed by that, two empirical-to-conceptual iterations were conducted (Nickerson et al., 2013). This resulted in the addition of 2 new dimensions and 9 characteristics. During the iterative design process, the ending conditions were checking after every iteration (Nickerson et al., 2013). After four design iterations, the taxonomy was finished and it comprised of 4 meta-dimensions, 17 dimensions and 59 characteristics. The final taxonomy is presented in Table 13 in this Chapter.

The identified meta-characteristics and corresponding business model dimensions of data marketplaces are: *value proposition*, *enterprise data marketplace*, *data processing and/or analytics tools*, *marketplace participants*, *industry domain*, *geographic scope* and *time frame* in the *Service domain*; *platform architecture*, *data access* and *data source* in the *Technology domain*; *matching mechanism* and *platform sponsor* in the *Organization domain*; and *revenue model*, *pricing model*, *price discovery*, *smart contract* and *payment currency* in the *Finance domain.*

SQ 5: How can the business model taxonomy be used to classify business models of different types of data marketplaces?

Finally, the use of the taxonomy was demonstrated on the basis of the empirical illustration of the classification of three empirical cases of data marketplaces, namely Wibson, QueXopa and Advaneo. The cases were classified based on publicly available information found on the companies' websites, white papers, news articles and academic studies. Based on the discovered information, most of the business model characteristics of the selected companies could be classified with the use of the taxonomy. However, not all information sources provided sufficient information about the respective business model characteristics of the

companies. Therefore, not all characteristics could be classified using the taxonomy. Thus, an important take-away from the demonstration is that sufficient information needs to available about the business model characteristics of the selected data marketplaces, in order for the taxonomy to be a useful tool. In the demonstration in this study, publicly available information was consulted to discover the characteristics of the selected cases. However, the data marketplace provider may contacted to retrieve additional information.

In this study, data marketplaces were broad interpreted as the digital infrastructure on which the commercial trade of data takes place, and where the data is traded as a valuable good. Whereas existing taxonomies studies mainly focus on the classiciation of multilateral data marketplaces (Fruhwirth et al., 2020; Spiekermann, 2019), the demonstration of the use of the taxonomy illustrated that the taxonomy is not only suitable to classify multilateral data marketplaces, but it also well suited to classify various of other types of data marketplaces, such as bilateral, harvest and dispersal data marketplaces (Koutroumpis et al., 2017).

8.2. Answering the main research question

In this research, a design science research approach was adopted and a standard taxonomy development method was employed to develop a taxonomy of data marketplace business models. The research aimed to go beyond the state of the art by taking on a broad interpretation of a data marketplace and adopting a multi-stakeholder perspective on the business models of data marketplaces.

The main research question that summarized the academic challenge and that structures the research was:

How can the business model characteristics of different types of data marketplaces be classified into a taxonomy from a multi-stakeholder perspective?

In this study, data marketplaces were broadly interpreted as digital infrastructures that enable the commercial trade of data as a valuable good. A data marketplace business model was defined as *the mapping of how a data marketplace enterprise aims to create and capture value by providing a marketplace and additional value adding services for the commercial trade of data between data providers and data buyers.* Following an iterative taxonomy development process (Nickerson et al. 2013), a taxonomy of data marketplace business models was developed. The final taxonomy is presented in Table 13, and comprises of 4 meta-dimensions, 17 business model dimensions and 59 business model characteristics.

Table 13: Taxonomy of data marketplace business models

	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			

8.3. Managerial and societal relevance

The developed taxonomy is relevant to managers and society, as it provides knowledge, information and transparency about the business models of data marketplaces. The taxonomy that was developed in this study can be used by managers and other decision-makers who are exploring the options of setting up a data marketplace or that are considering to join an existing data marketplace, to provide guidance in making business model design choices. Appendix III provides an overview of the questions that practitioners can ask to assess the business model of a data marketplace, and the definitions of the various business model characteristics in the taxonomy. An improved understanding about data marketplace business models may result in an increase of data marketplaces, which may make data more accessible and exploitable to a wide range of stakeholders, including individuals, businesses and authorities. Furthermore, a number of existing data marketplaces that enable the transactions of audience, health and personal data was exposed during this research, that may be directly gather data from individuals or communities in society. Therefore, this study may promote the dialogue about the existence of companies that gather personal data of consumers with the goal of exchanging it for commercial purposes. Moreover, the results of this study may raise awareness about the emergence of harvesting data marketplaces, that enable consumers to monetize their personal and health data.

8.4. Relevance to CoSEM programme

CoSEM engineers focus on designing in socio-technical systems. Data marketplaces may comprise of a platform ecosystem that is socially complex, because it may consist of a multitude of stakeholders such as data buyers, sellers and external service providers, that interact and are interconnected with each other. Furthermore, this study comprises of a technological component by addressing how novel technologies such as such as distributed ledgers, smart contracts and cryptocurrencies function and how they are applied in existing data marketplaces. Furthermore, CoSEM engineers design interventions in real world-decision making processes. In this study, a design science research approach was employed to design an artifact in the form of a taxonomy (Hevner, 2007). The taxonomy structures the knowledge about data marketplace business models and can can support academic researchers and decision-makers in the industry to provide guidance in the process of classification, design and reconfiguration of data marketplace business models.

9. Reflection

In this chapter, a reflection is given on the research process and results of the study. First, the challenges that were faced during the taxonomy development process are discussed. Then, a comparison of the created taxonomy and the two existing taxonomies by Fruhwirth et al. (2020) and Spiekermann (2019) is presented, and an argumentation is given how this study goes beyond the state of the art research. After that, the scientific contributions of this research are argued for and the limitations of the research process that possibly impacted the results are addressed. Lastly, recommendations for further research are presented.

9.1. Challenges faced during the taxonomy development process

During the taxonomy development process, a number of hindrances were faced that made the research challenging. In this section, the challenges of the research process and how they were dealt with are discussed.

First of all, when conducting desk research for empirical cases of data marketplaces, it was found that in practice the terms data marketplace, exchange and data platform are often interchanged. For instance, the companies DAWEX and Snowflake host a data marketplace and offer an enterprise data marketplace as an additional service, naming this service a 'Data Exchange'. However, for some companies the enterprise data marketplace is the main offering, in which proprietary data provided by the data marketplace provider is integrated in the enterprise marketplace service for customers to enrich their data and enhance their data projects. For these atypical examples of data marketplaces, a variety of terms are used in practice. Examples of enterprise data marketplaces that adopted divergent terms are the OpenPrise Data Orchestration Platform, Data Republic's Senate Platform and Informatica's B2B Exchange.

The same challenge was faced with regard to how the terms data, information and insights are used in practice. For example, companies such as CARTO, HERE and Knoema offer data processing and analytics tooling on top of their data, that customers can use to directly process their proprietary data into comprehensive tables, graphs and other visualizations. Often, this type of companies also offer data sets to enrich the customers' proprietary data and improve their insights. While data is the underlying good in the data exchange process, the exchanged data is transformed into insights or information, regardless of whether the data is provided by the customer or whether access to external data is provided by the data marketplace. This resonates the statement by Koutroumpis et al. (2020) that found that data marketplaces often exchange access to data and data-related services rather than explicitly selling data goods.

To deal with the variety of terminologies used in practice, the term "data marketplace" was broadly interpreted in this research, as *the digital infrastructure on which the commercial trading of data as a valuable good takes place*. This broad interpretation allowed for the inclusion of atypical forms of data marketplaces, as long as data was commercially traded as a valuable good via a digital infrastructure.

9.2. Comparison of the developed taxonomy and existing taxonomies

In this section, the developed taxonomy is compared with the two existing taxonomies by Fruhwirth et al. (2020) and Spiekermann (2019). An argumentation is given how this study goes beyond the state of the art. Furthermore, a reflection is given on the broad interpretation of a data marketplace and the multi-stakeholder perspective on data marketplaces that were adopted in this study. An illustration of the taxonomy dimensions and comparison with the two existing taxonomies is provided in Appendix IV.

In the rigor cycle of this design science research (Hevner, 2007), existing theories and artifacts about the object of interest were examined and concepts were derived to form a preliminary taxonomy (see Chapter 3). Therefore, on the one hand this study builds on the knowledge and understanding of the object of interest by Fruhwirth et al. (2020) and Spiekermann (2019), but on the other hand this study refines the knowledge about data marketplace business models by including novel business model dimensions and characteristics.

Three dimensions that are included in the newly developed taxonomy are also present or similar to dimensions in the two existing taxonomies: the *industry domain*, *platform architecture* and *pricing model*. Furthermore, a number of dimensions that were exclusive for each of the two taxonomies were incorporated in the newly developed taxonomy. The business model dimensions *value proposition* and *revenue model* were derived from Spiekermann (2019), and the dimensions *marketplace participants*, *time frame*, *data access*, *data source*, *price discovery*, *smart contract* and *payment currency* in the taxonomy of Fruhwirth et al. (2020) were also employed in the taxonomy in this study.

This study goes beyond the state of the art by the refinements of the characteristics in dimensions that were derived from the existing taxonomies (Fruhwirth et al., 2020; Spiekermann, 2019). For instance, the *industry domain* dimension that is included in the newly developed taxonomy is similar to the characteristics *domain* by Fruhwirth et al. (2020) and *integration* by Spiekerman (2019). Whereas Spiekermann (2019) makes a distinction between *domain specific* data marketplaces and *domain unspecific* data marketplaces, this study refines the domain dimension by making a distinction between six types of industry domains that data marketplaces are active in. Compared to the domain dimension by Fruhwirth et al. (2020), the *industry domain* dimension was refined by introducing the *audience data* industry domain and refining the *finance*, *address*, *personal* and *sensor* domains to *financial & alternative data*, *health & personal data* and *sensor & mobility data*. These new characteristics were discovered during the analysis of the sample of 40 data marketplaces in this study.

A number of dimensions were added to the taxonomy that were not yet considered in the existing taxonomy studies (Fruhwirth et al., 2020; Spiekermann, 2019). In the Service domain, the binary dimensions *Enterprise data marketplace* and *data processing and/or analytics tools* were included to indicate the possible additional services that may be offered on top of the data offering on the data marketplace. Furthermore, the *Geographic scope* dimension was included to enable the differentiation between globally operating data marketplaces, regional focussed data marketplaces and marketplaces that are locally active.

In this study, the term data marketplace was broadly interpreted as *the digital infrastructure on which the commercial trading of data as a valuable good takes place*. Whereas the two existing taxonomies mainly focus on the classification of multilateral data marketplaces (Fruhwirth et al., 2020; Spiekermann, 2019), the objective of this study was to develop a taxonomy that enables the classification of other types of data marketplaces as well. Therefore, the *matching mechanism* and *platform sponsor* dimensions were added to the Organization domain. The *matching mechanism* dimension allows for the classification of different types of data marketplaces based on the number of marketplace participants active on each side of the market (Koutroumpis et al., 2017). Furthermore, the *platform sponsor* dimension indicates the positioning of the sponsor of the marketplace. This dimension is quite similar to the *market positioning* dimension by Spiekermann (2019). However it was found that the two separate dimension in the Organization domain of the taxonomy better capture the market positioning of the data marketplace.

The business models of data marketplaces were viewed from a multi-stakeholder perspective in this study. The meta-characteristics of this study were based on the STOF ontology by Bouwman et al. (2008), which provided the lens through which the various business model characteristics of data marketplaces are classified. In an advanced form, a data marketplace may function as a digital intermediary platform that allows third-party service providers to offer and improve their services via the platform (Carnelley et al., 2016). The taxonomy dimensions *Enterprise data marketplace* and *Data processing and/or analytics tools* capture services and functions of the data marketplace that may possibly be provided by external service providers or the data marketplace itself. Not all data marketplaces in the sample provided additional services such as an enterprise data marketplace and data analytics tools on top their data (see Appendix II). Moreover, it was not always clear from the publicly available information whether the services were developed by the data marketplace provider itself, or whether they were provided by external service providers. Therefore, future research can focus on capturing and improving the multi-stakeholder aspect of data marketplace business models.

A number of dimensions that were included in the two existing taxonomies (Fruhwirth et al., 2020; Spiekermann, 2019) were not included in the newly developed taxonomy in this study. The dimensions *data quality guarantee*, *privacy*, and *additional purchase support* by Fruhwirth et al. (2020) were not included because they were not included in the preliminary taxonomy (Chapter 3) to be distinctive business model characteristics for data marketplaces. The dimensions *pre-purchase testability*, *review system* and *data output* were considered during the taxonomy development process (section 5.3.1), however it was found that the data marketplaces in the sample either did not provide sufficient information for these dimensions or the data marketplaces were advancing to one type of the characteristics, and therefore these dimensions were not found to be suitable for the classification of different types of data marketplaces. The dimension *transformation* and *market positioning* by Spiekermann (2019) are partly covered by the dimensions *Data processing and/or analytics tools* and *platform sponsor* in the newly developed taxonomy. Lastly, the dimension *market access* by Spiekermann (2019) was not considered during the taxonomy development process, because the research by Spiekermann (2019) indicated that most of the existing data marketplaces had adopted a *hybrid* form of *market access*, and the data marketplaces that adopted an *open* form of market access were closed. Furthermore, during the empirical-to-conceptual design phase of this study, it was found to be hard to determine the market access or openness of the data marketplaces. Therefore, this dimension is suggested for further research.

Lastly, Fruhwirth et al. (2020) includes a *no info* characteristic in their taxonomy, to indicate that no sufficient information was found for certain cases of data marketplaces they analysed. In this study, the choice was made to not include a *no info* characteristic because it did not seem an appropriate business model characteristic. Rather, if an empirical case did not provide sufficient information about the respective business model characteristics, the cells in the classification table were blank and the case will not be classified on the respective business model dimension (see Appendix II).

9.3. Scientific contributions

The developed taxonomy of data marketplace business model contributes to science in a number of ways. In this section, the scientific contributions of the research are presented and discussed.

First, the results of the study contribute to understanding the notion of data marketplace business models by developing a taxonomy that describes the most important dimensions and characteristics of data marketplace business models. This way, the research contributes to the scarce knowledge about data marketplaces and their respective business models (Thomas & Leiponen, 2016). This study goes beyond the state of the art by adopting a multi-stakeholder perspective on data marketplace business models, by emphasizing the roles of players in the data marketplace platform ecosystem. For academic researchers, the taxonomy can function as a knowledge map that displays the contemporary knowledge from both scientific research and practical applications. It may expose certain areas that require further research. The developed taxonomy provides academic researchers with an overview of the characteristics of data marketplace business models. The taxonomy may expose certain areas of research where new business model alterations are emerging, that were derived from empirical cases and that may have not been thoroughly research in science yet. For instance, the taxonomy exposes the emergence of novel technological applications by marketplace providers in the industry that enhance trust among data marketplace participants, such as enterprise data marketplaces, decentralized platform architectures, smart contracts and cryptocurrencies as a payment method.

A second contribution made by this study is related to the interpretation of a data marketplace. Compared to existing data marketplace business model taxonomies, this research takes on a broad perspective and interpretation of data marketplaces. Existing data marketplace business model taxonomies focus on studying one variant of data marketplaces, multilateral data marketplaces, where the marketplace functions as neutral intermediary that connects data buyers and data sellers (Fruhwirth et al., 2020; Spiekermann, 2019). In this study, data marketplaces are more broadly interpreted as *the digital infrastructure on which the commercial trading of data as a valuable good takes place*. In this interpretation, no assumptions are made about the number of users on each side of the market, or the role of the data marketplace provider. Therefore, marketplaces that comprise of only a single data seller or a single data buyer on each side of the market were also considered to be a data marketplace in this study. Furthermore, it is assumed that the provider of the data marketplace may also take on the role of data buyer or data seller. Additionally, this interpretation excludes data marketplaces hosted by governmental agencies and NGOs that provide free or open data, because data marketplaces are interpreted to enable commercial transactions, rather

than providing data for free. This definition also implies that it is assumed that data is traded on a digital infrastructure, rather than a physical infrastructure, where participants interact and exchange data in commercial transactions. By providing a sound definition of a data marketplace, this study aims to contribute to the knowledge and understanding of data marketplaces.

9.4. Limitations

The research process and results of this study are subject to a number of limitations. In this section, the limitations are addressed.

Firstly, the interpretation of qualitative information about the business models of empirical cases is prone to the subjectivity of the researcher. In this study, desk research was conducted to gather information about the business models of existing data marketplaces by consulting company websites, online news articles and other sources, and within case analysis was conducted on the gathered information. As this research was conducted by a single researcher, there is a chance that valuable information may have been missed or that the gathered information was misinterpreted. The interpretation of the found information is subject to researchers' knowledge about the object of interest. Therefore, other researchers may find more information or interpret the information differently, which may result in the finding of different business model dimensions characteristics. However, as multiple design iterations were conducted in this research to develop the taxonomy, desk research for relevant information was performed multiple times and the discovered information was analysed in multiple iterations.

Secondly, the database of existing data marketplaces was constituted by consulting data marketplaces included in the repository of datarade.ai and existing scientific studies, and was complemented with desk research for empirical cases in Google. This search process resulted in the discovery of 178 data marketplaces. It may be that some existing data marketplaces have been missed during the desk research, or that new data marketplaces have been set up during the writing of this thesis. Therefore, future research on data marketplace business models may discover additional and new empirical cases that will improve the understanding of the object of interest.

Thirdly, not all data marketplace companies provided sufficient information about all of their business model characteristics. Therefore, not all empirical cases could be classified on the conceptually derived dimensions (see Appendix II). The business model dimensions *main revenue partner* and *key costs* (Täuscher & Laudien, 2018; Täuscher, 2016) may be relevant for the classification of data marketplace business models, but they were omitted during the taxonomy development process because no sufficient information was found to make well-founded statements about these dimensions. In this research, there has not been any direct contact with providers of data marketplaces to verify the information about their respective business models. Further research on the object of interest may therefore focus on case studies of specific data marketplaces, that involve interviews, surveys and other qualitative research methods to gain in-depth insight in business model dimensions and characteristics that have not been addressed during this research.

9.5. Recommendations for further research

A number of possible opportunities and directions for further research arise from the results and limitations of this research. In this section, recommendations for further research are outlined.

Firstly, future researchers can utilize the designed taxonomy to derive business model patterns and archetypes in existing data marketplaces. Business model patterns are business model characteristics that are commonly used in practical cases (Remane et al., 2017; Abdelkafi et al., 2013). Subsequently, business model archetypes are configurations of business model characteristics that are common among existing companies (Fruhwirth et al., 2019; Hodapp et al., 2019; Weking et al., 2019). The statistical derivation of business model patterns and archetypes may be conducted existing data marketplaces in a specific industry domain, such as the audience data industry, or by comparing business model patterns across different industries. Using the developed taxonomy for the statistical analysis of patterns and archetypes may uncover the frequency of certain business model characteristics and business model configurations, which can provide valuable information about the relative importance of certain business model dimensions and characteristics.

Secondly, the different alterations of data marketplaces that occur in practice may be studied, taking the variety of terms used by data marketplace providers in practice in consideration. During the desk research, it was found that the terms data marketplace, exchange and data platform are often interchanged. The same was true for the use of the terms data, information and insights. To allow the inclusion of atypical examples of data marketplaces, the term data marketplace was broadly interpreted in this study. However, future research may focus on providing a concise definition of a data marketplace, and take into consideration the various terms used in practice. This way, a more clear differentiation can be made between different types of data marketplaces.

Thirdly, in-depth case studies may be conducted on specific data marketplaces or in certain industry domains, by conducting interviews, surveys and other qualitative empirical analyses. The in-depth analyses of existing data marketplace companies may provide deeper insight about certain business model dimensions and characteristics that were not highlighted in this research, such as the key costs and main revenue partners of data marketplaces. From a strategy perspective on business models, such in-depth case studies may uncover the considerations that are made by data marketplace providers with regard to choosing between certain business model characteristics, and provide insight about which business model characteristics provide certain data marketplace companies with a competitive advantage over other companies. Furthermore, the in-depth case studies of data marketplaces may deepen the knowledge about the multi-stakeholder perspective on data marketplaces, such as the market access or openness of the platforms to external service providers. During the desk research, it was found to be hard to determine the openness of the data marketplace based on the publicly available information. Through interviews with relevant experts in the field of data marketplaces, the taxonomy may also be validated to see whether the derived business model dimensions and characteristics are relevant, whether the taxonomy should be revised, or if new concepts can be added to the taxonomy on the basis of expert knowledge.

References

Abdelkafi, N., Makhotin, S., & Posselt, T. (2013). Business model innovations for electric mobility—what can be learned from existing business model patterns?. *International Journal of Innovation Management*, *17*(01), 1340003.

Al-Debei, M. M., El-Haddadeh, R., & Avison, D. (2008). Defining the business model in the new world of digital business. *AMCIS 2008 proceedings*, 300.

Al-Debei, M. M., & Avison, D. (2010). Developing a unified framework of the business model concept. *European Journal of Information Systems*, *19*(3), 359-376.

Amit, R., & Zott, C. (2001). Value creation in e-business. *Strategic management journal*, *22*(6-7), 493-520.

Applegate, L. M. (2001). Emerging E-Business Models: Lessons from the Field, Harvard Business School.

Arrow, K. J. (1972). Economic welfare and the allocation of resources for invention. In *Readings in industrial economics* (pp. 219-236). Palgrave, London.

Azkan, C., Iggena, L., Gür, I., Möller, F., & Otto, B. (2020). A Taxonomy for Data-Driven Services in Manufacturing Industries. In *PACIS* (p. 184).

Bailey, K. D. (1994), Typologies and Taxonomies: *An Introduction to Classification Techniques*. Sage Publications Inc Los Angeles.

Bakos, Y. (1998). The emerging role of electronic marketplaces on the Internet. *Communications of the ACM*, *41*(8), 35-42.

Baldwin, C. Y., & Woodard, C. J. (2009). The architecture of platforms: A unified view. *Platforms, markets and innovation*, *32*.

Bock, M., & Wiener, M. (2017). Towards a Taxonomy of Digital Business Models-Conceptual Dimensions and Empirical Illustrations. In *ICIS*.

Boudreau, K. J., & Hagiu, A. (2009). Platform rules: Multi-sided platforms as regulators. *Platforms, markets and innovation*, *1*, 163-191.

Bounie, D., Marcocchia, G., & Quinn, M. (2018). Deliverable D6.7 PUSH AND EFFECTUAL APPROACH BUSINESS MODEL Value Chain, Value Network, Economic model and Marketplace P&L simulation. Retrieved 9 July 2020, from https://www.automat-project.eu/sites/default/files/automat/public/content-files/articles/AutoMat-D6.7%20Push%20and%20effectual%20business%20model_final.pdf

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

Bouwman, H., Nikou, S., & de Reuver, M. (2019). Digitalization, business models, and SMEs: How do business model innovation practices improve performance of digitalizing SMEs?. *Telecommunications Policy*, *43*(9), 101828.

Cantrell, L. J., & Linder, J. (2000). Changing business models: Surveying the landscape. *Accenture Institute for Strategic Change*.

Carnelley, P., Schwenk, H., Cattaneo, G., Micheletti, G., & Osimo, D. (2016). Europe's Data Marketplaces-Current Status and Future Perspectives. *European Data Market SMART 2013/0063 D*, *39*.

Casadesus-Masanell, R., & Ricart, J. E. (2010). From strategy to business models and onto tactics. *Long range planning*, *43*(2-3), 195-215.

Chakrabarti, A., Quix, C., Geisler, S., Pullmann, J., Khromov, A., & Jarke, M. (2018). Goal-Oriented Modelling of Relations and Dependencies in Data Marketplaces. In *iSTAR@ CAiSE*.

Chesborough, H., & Rosenbloom, R. S. (2002). The role of the business model in capturing value from innovation: evidence from Xerox Corporation's technology spin-off companies. *Industrial and corporate change*, *11*(3), 529-555.

Daniel, J. (2011). *Sampling essentials: Practical guidelines for making sampling choices*. Sage Publications.

Dodds, L. (2012). Kasabi – Shutting Down Kasabi. Retrieved March May 12th, from https://kasabi.wordpress.com/2012/07/09/shutting-down-kasabi/

DOMO. (2019). Domo Resource - Data Never Sleeps 7.0. Retrieved March 12th, 2020, from https://www.domo.com/learn/data-never-sleeps-7

Eisenmann, T., Parker, G., & Van Alstyne, M. W. (2006). Strategies for two-sided markets. *Harvard business review*, *84*(10), 92.

Eisenmann, T. R., Parker, G., & Van Alstyne, M. (2009). *Opening Platforms: How, When and Why?*. Edward Elgar Publishing.

European Commission. (2017). Communication on Building a European Data Economy. Retrieved March 12, 2020, from https://ec.europa.eu/digital-single-market/en/news/communication-building-european-data-economy

European Commission. (2019). *Information and Communication Technologies - Horizon 2020.* Retrieved on September 29th 2019, from https://ec.europa.eu/programmes/horizon2020/en/h2020-section/information-and-communication-technologies

The European data market study update. (2020). Retrieved 9 July 2020, from https://ec.europa.eu/digital-single-market/en/news/european-data-market-study-update

Faber, E., Ballon, P., Bouwman, H., Haaker, T., Rietkerk, O., & Steen, M. (2003). Designing business models for mobile ICT services. In *Workshop on concepts, metrics & visualization, at the 16th Bled electronic commerce conference etransformation, Bled, Slovenia*.

Fernandez, D., Futoransky, A., Ajzenman, G., Travizano, M., & Sarraute, C. (2020). Wibson protocol for secure data exchange and batch payments. *arXiv preprint arXiv:2001.08832*.

Fricker, S. A., & Maksimov, Y. V. (2017). Pricing of data products in data marketplaces. In *International Conference of Software Business* (pp. 49-66). Springer, Cham.

Fruhwirth, M., Rachinger, M., & Prlja, E. (2020, January). Discovering Business Models of Data Marketplaces. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*.

Gassmann, O., Frankenberger, K., & Csik, M. (2014). *The business model navigator: 55 models that will revolutionise your business*. Pearson UK.

Giesen, E., Berman, S. J., Bell, R., & Blitz, A. (2007). Three ways to successfully innovate your business model. *Strategy & leadership*.

Gordijn, J., & Akkermans, J. M. (2003). Value-based requirements engineering: exploring innovative e-commerce ideas. *Requirements engineering*, *8*(2), 114-134.

Hartmann, P. M., Zaki, M., Feldmann, N., & Neely, A. (2014). Big data for big business? A taxonomy of data-driven business models used by start-up firms. *A taxonomy of data-driven business models used by start-up firms*.

Henderson, R. M., & Clark, K. B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative science quarterly*, 9-30.

HERE. (2019). HERE Technologies welcomes Mitsubishi Corporation and NTT as new investors. Retrieved 9 July 2020, from https://www.here.com/company/press-releases/en/2019-20-12

Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian journal of information systems*, *19*(2), 4.

Hodapp, D., Remane, G., Hanelt, A., & Kolbe, L. M. (2019). Business Models for Internet of Things Platforms: Empirical Development of a Taxonomy and Archetypes.

Hürtgen, H., & Mohr, N. (2018). Achieving business impact with data. Retrieved March 12th, 2020, from https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/achieving-business-impact-with-data

Iivari, J. (2007). A paradigmatic analysis of Information Systems as a design science. *Scandinavian Journal of Information Systems.*

Johnson, B. (2012). Kasabi shuts down, says data marketplace ‘too slow’. Retrieved from https://gigaom.com/2012/07/09/kasabi-shuts-down-says-data-marketplace-too-slow/

Katz, M. L., & Shapiro, C. (1985). Network externalities, competition, and compatibility. *The American economic review*, *75*(3), 424-440.

Kosara, R. (2010). The rise and fall of swivel. com. Retrieved May 12th, from https://eagereyes.org/criticism/the-rise-and-fall-of-swivel

Koutroumpis, P., Leiponen, A., & Thomas, L. D. (2017). *The (unfulfilled) potential of data marketplaces* (No. 53). ETLA Working Papers.

Koutroumpis, P., Leiponen, A., & Thomas, L. D. (2020). Markets for data. *Industrial and Corporate Change*, *29*(3), 645-660.

Krcmar, H., Böhm, M., Friesike, S., & Schildhauer, T. (2011). Innovation, Society and Business: Internet-based Business Models and their Implications.

Lambert, S. (2015). The importance of classification to business model research. *Journal of Business Models*, *3*(1).

Lange, J., Stahl, F., & Vossen, G. (2018). Datenmarktplätze in verschiedenen Forschungsdisziplinen: Eine Übersicht. *Informatik-Spektrum*, *41*(3), 170-180.

Lawrenz, S., Sharma, P., & Rausch, A. (2019, March). Blockchain technology as an approach for data marketplaces. In *Proceedings of the 2019 International Conference on Blockchain Technology* (pp. 55-59).

McKelvey, B. (1982). *Organizational systematics--taxonomy, evolution, classification*. Univ of California Press.

Meisel, L., Spiekermann, M., Otto, B., Rehof, J. (2019). Datenmarktplätze - Plattformen für Datenaustausch und Datenmonetarisierung in der Data Economy. *Fraunhofer-Institut für Software- und Systemtechnik ISST*

Muschalle, A., Stahl, F., Löser, A., & Vossen, G. (2012). Pricing approaches for data markets. In *International workshop on business intelligence for the real-time enterprise* (pp. 129-144). Springer, Berlin, Heidelberg.

Nickerson, R. C., Varshney, U., & Muntermann, J. (2013). A method for taxonomy development and its application in information systems. *European Journal of Information Systems, 22*(3), 336-359.

Opher, A., Chou, A., Onda, A., & Sounderrajan, K. (2016). The rise of the data economy: driving value through internet of things data monetization. *IBM Corporation: Somers, NY, USA*.

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

Permira. (2015). Informatica Announces Agreement to be Acquired by The Permira Funds and Canada Pension Plan Investment Board | Permira. Retrieved 9 July 2020, from https://www.permira.com/news-views/news/informatica-announces-agreement-to-be-acquired-by-the-permira-funds-and-canada-pension-plan-investment-board/

Prlja, E. (2019). *Discovering Business Models of Data Marketplaces* (Unpublished master's thesis). Graz University of Technology, Graz, Austria.

Ramel, D. (2016). "Microsoft Closing Azure DataMarket." *Application Development Trends MAG.* Retrieved May 12th, from https://adtmag.com/articles/2016/11/18/azure-datamarket-shutdown.aspx

Remane, G., Nickerson, R., Hanelt, A., Tesch, J. F., & Kolbe, L. M. (2016). A taxonomy of carsharing business models.

Remane, G., Hanelt, A., Tesch, J. F., & Kolbe, L. M. (2017). The business model pattern database—a tool for systematic business model innovation. *International Journal of Innovation Management*, *21*(01), 1750004.

Richardson, J. (2008). The business model: An integrative framework for strategy execution. *Strategic Change*.

Rochet, J. C., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the european economic association*, *1*(4), 990-1029.

Schomm, F., Stahl, F., & Vossen, G. (2013). Marketplaces for data: an initial survey. *ACM SIGMOD Record*, *42*(1), 15-26.

Spiekermann, M., Tebernum, D., Wenzel, S., & Otto, B. (2018). A metadata model for data goods. *MKWI 2018-Multikonferenz Wirtschaftsinformatik, 2018–March*, 326-337.

Spiekermann, M. (2019). Data Marketplaces: Trends and Monetisation of Data Goods. *Intereconomics*, *54*(4), 208-216.

Stahl, F., Schomm, F., & Vossen, G. (2014a). Data Marketplaces: An Emerging Species. In *DB&IS* (pp. 145-158).

Stahl, F., Schomm, F., & Vossen, G. (2014b). *The data marketplace survey revisited* (No. 18). Working Papers, ERCIS-European Research Center for Information Systems.

Stahl, F. (2016). *High-Quality web information provisioning and quality-based data pricing* (Doctoral dissertation, Universitäts-und Landesbibliothek Münster).

Stahl, F., Schomm, F., Vossen, G., & Vomfell, L. (2016). A classification framework for data marketplaces. *Vietnam Journal of Computer Science*, *3*(3), 137-143.

Stahl, F., Schomm, F., Vomfell, L., & Vossen, G. (2017). Marketplaces for Digital Data: Quo Vadis?. *Computer and Information Science*, *10*(4).

Tapscott, D., Ticoll, D., & Lowy, A. (2000). Digital capital: Harnessing the power of business webs. *Ubiquity*, *2000*(May), 3-es.

Tashakkori, A., & Teddlie, C. (2003). *Handbook of mixed methods in social & behavioral research*. Thousand Oaks, Calif: SAGE Publications.

Täuscher, K. (2016). Business Models in the Digital Economy: An Empirical Classification of Digital Marketplaces. *Fraunhofer Center For International Management and Knowledge Economy.*

Täuscher, K., Hilbig, R., & Abdelkafi, N. (2017). Geschäftsmodellelemente mehrseitiger Plattformen. In *Digitale Transformation von Geschäftsmodellen* (pp. 179-211). Springer Gabler, Wiesbaden.

Täuscher, K., & Laudien, S. M. (2018). Understanding platform business models: A mixed methods study of marketplaces. *European Management Journal*, *36*(3), 319-329.

Teece, D. J. (2010). Business models, business strategy and innovation. *Long range planning*, *43*(2-3), 172-194.

Thomas, L. D., & Leiponen, A. (2016). Big data commercialization. *IEEE Engineering Management Review, 44*(2), 74-90.

Timmers, P. (1998). Business models for electronic markets. *Electronic markets*, *8*(2), 3-8.

Tiwana, A., Konsynski, B., & Bush, A. A. (2010). Research commentary—Platform evolution: Coevolution of platform architecture, governance, and environmental dynamics. *Information systems research*, *21*(4), 675-687.

Tiwana, A. (2014). *Platform ecosystems: Aligning architecture, governance, and strategy.* Burlington: Morgan Kaufmann.

Van Alstyne, M. W., Parker, G. G., & Choudary, S. P. (2016). Pipelines, platforms, and the new rules of strategy. *Harvard business review*, *94*(4), 54-62.

Vessey, I., Ramesh, V., & Glass, R. L. (2005). A unified classification system for research in the computing disciplines. *Information and Software Technology*, *47*(4), 245-255.

Von Engelhardt, S., Wangler, L., & Wischmann, S. (2017). Eigenschaften und Erfolgsfaktoren digitaler Plattformen–Eine Studie im Rahmen der Begleitforschung zum Technologieprogramm AUTONOMIK für Industrie 4.0 des Bundesministeriums für Wirtschaft und Energie.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly*, xiii-xxiii.

Weill, P., & Vitale, M. (2001). *Place to space: Migrating to eBusiness Models*. Harvard Business Press.

Weill, P., Malone, T. W., & Apel, T. G. (2011). The business models investors prefer. *MIT Sloan Management Review*, *52*(4), 17.

Weking, J., Mandalenakis, M., Hein, A., Hermes, S., Böhm, M., & Krcmar, H. (2019). The impact of blockchain technology on business models–a taxonomy and archetypal patterns. *Electronic Markets*, 1-21.

Yoo, Y., Boland Jr, R. J., Lyytinen, K., & Majchrzak, A. (2012). Organizing for innovation in the digitized world. *Organization science*, *23*(5), 1398-1408.

Zott, C., & Amit, R. (2008). The fit between product market strategy and business model: implications for firm performance. *Strategic management journal*, *29*(1), 1-26.

Appendix

Appendix I: Overview of sample of empirical cases

Table 14: Overview of sample of empirical cases

Data marketplace	Data type	Industry domain	Website
Datahub	Any data	Any data	https://datahub.io/
Data Intelligence Hub	Any data	Any data	https://dih.telekom.net/en/
DAWEX	Any data	Any data	https://www.dawex.com/en/
Streamr	Sensor data	Sensor	https://streamr.network/marketplace
Databroker	Any data	Any data	https://databroker.global/
BattleFin Ensemble	Financial & Alternative data	Financial & Alternative data	https://www.battlefin.com/ensemble-new
Intrinio	Financial data	Financial & Alternative data	https://intrinio.com/
QuantConnect	Financial data	Financial & Alternative data	https://www.quantconnect.com/
Knoema	Any data	Any data	https://knoema.com/
Open:Factset Marketplace	Alternative data	Financial & Alternative data	https://open.factset.com/en-us
Amazon DSP	Audience	Audience data	https://advertising.amazon.com/products/amazon-dsp
Salesforce Audience Studio	Audience	Audience data	https://www.salesforce.com/products/marketing-cloud/data-management/
BidTheatre	Audience	Audience data	https://www.bidtheatre.com/
RollWorks	Audience	Audience data	https://www.rollworks.com/
SimilarWeb	Audience	Audience data	https://www.similarweb.com/
Adsquare	Audience	Audience data	https://www.adsquare.com/
OpenPrise Data Orchestration Platform	Audience	Audience data	https://www.openprisetech.com/
Marketscan	Audience	Audience data	https://www.marketscan.co.uk/
Snowflake Data Marketplace	Any data	Any data	https://www.snowflake.com/data-marketplace/
Opendatasoft	Any data	Any data	https://www.opendatasoft.com/
Informatica B2B Exchange	Any data	Any data	https://www.informatica.com/nl/products/data-integration/b2b-data-exchange.html
Data Republic	Audience data	Audience data	https://www.datarepublic.com/
Otonomo	Connected car data, automotive data	Sensor & Mobility data	https://otonomo.io/platform/
Caruso	Connected car data, automotive data	Sensor & Mobility data	https://www.caruso-dataplace.com/
IOTA	Sensor data	Sensor & Mobility data	https://www.iota.org/
ThinkDataWorks	Any data	Any data	https://www.thinkdataworks.com/products/marketplace
Datax	Personal data	Health & Persondal data	https://datax.io/
HERE Platform	Location data	Geo data	https://www.here.com/platform/
CARTO	Location data	Geo data	https://carto.com/
Factual	Location data	Geo data	https://www.factual.com/
Red Lion Data	Location data	Geo data	https://www.redliondata.com/
Veracity	Environmental data, industrial data	Geo data	https://store.veracity.com/
HealthVerity	Health data	Health & Personal data	https://www.burstiq.com/
CoverUS	Health data	Health & Personal data	https://coverus.health/
Medicalchain	Health data	Health & Personal data	https://medicalchain.com/en/
Datum Data Marketplace	Personal data	Health & Personal data	https://datum.org/
BIGToken	Personal data	Health & Personal data	https://bigtoken.com/
BurstIQ	Health & Personal		https://www.burstiq.com/
Mobility Data Marketplace	traffic data, petrol price data, parking data	Sensor & mobility data	https://www.mdm-portal.de/en/
oneTRANSPORT	traffic data, petrol price data, parking data	Sensor & Mobility data	https://service.onetransport.io/

Appendix II: Overview of the classification of cases in the sample

Table 15: Overview of classification of cases in the Service domain

		Service domain						
#	Data marketplace	Value proposition	Enterprise data marketplace	Data processing and analytics tools	Marketplace participants	Industry domain	Geographic scope	Time frame
1	Datahub	Easy data access and/or tooling	No	No	Any	Any data	Global	Up-to-date
2	Data Intelligence Hub	Easy data access and/or tooling	Yes	Yes	Any	Any data		Multiple
3	DAWEX	Secure data sharing	Yes	Yes	Any	Any data	Global	Multiple
4	Streamr	Easy data access and/or tooling	No	Yes	Any	Sensor & Mobility data	Global	(Near) Real-time
5	Databroker	Easy data access and/or tooling	Yes		B2B	Any data	Global	Multiple
6	BattleFin Ensemble	Easy data access and/or tooling	No	Yes	B2B	Financial & Alternative data	Global	Up-to-date
7	Intrinio	Easy data access and/or tooling	No	Yes	Any	Financial & Alternative data		Multiple
8	QuantConnect	Easy data access and/or tooling	No	Yes	Any	Financial & Alternative data		Multiple
9	Knoema	Easy data access and/or tooling	Yes	Yes	Any	Any data	Global	Up-to-date
10	Open:Factset Marketplace	Easy data access and/or tooling	No	Yes	B2B	Financial & Alternative data	Global	Multiple
11	Amazon DSP	High quality and unique data	No	Yes	B2B	Audience data	Global	
12	Salesforce Audience Studio	High quality and unique data	Yes	Yes	B2B	Audience data		
13	BidTheatre	All services in single platform	No	Yes	B2B	Audience data	Regional	
14	RollWorks	All services in single platform	No	Yes	B2B	Audience data		
15	SimilarWeb	High quality and unique data	No	Yes	B2B	Audience data	Global	
16	Adsquare	Easy data access and/or tooling	Yes	Yes	B2B	Audience data	Global	
17	OpenPrise Data Orchestration Platform	All services in single platform	No	Yes	B2B	Audience data		
18	Marketscan	High quality and unique data	No	Yes	B2B	Audience data	Local	Up-to-date
19	Snowflake Data Marketplace	Secure data sharing	Yes	Yes	B2B	Any data	Global	Multiple
20	Opendatasoft	Easy data access and/or tooling	Yes	Yes	B2B	Any data	Global	Multiple
21	Informatica B2B Exchange	Easy data access and/or tooling	Yes	Yes	B2B	Any data	Global	
22	Data Republic	Easy data access and/or tooling	Yes	Yes	B2B	Audience data	Regional	
23	Otonomo	Secure data sharing	No	Yes	B2B	Sensor & Mobility data	Global	(Near) Real-time
24	Caruso	Easy data access and/or tooling	No	Yes	B2B	Sensor & Mobility data		Multiple
25	IOTA	Secure data sharing	No	No	B2B	Sensor & Mobility data	Global	(Near) Real-time
26	ThinkDataWorks	Easy data access and/or tooling	Yes	Yes	Any	Any data	Global	Up-to-date
27	Datax	Easy data access and/or tooling	No	Yes	C2B	Health & Personal data		Static
28	HERE Platform	Easy data access and/or tooling	Yes	Yes	Any	Geo data	Global	Multiple
29	CARTO	All services in single platform	Yes	Yes	Any	Geo data	Global	Up-to-date
30	Factual	All services in single platform	No	Yes	B2B	Geo data	Global	Multiple
31	Red Lion Data	High quality and unique data	No	No	Any	Geo data	Regional	Up-to-date
32	Veracity	Easy data access and/or tooling	No	Yes	B2B	Any data	Global	
33	HealthVerity	Easy data access and/or tooling	Yes	Yes	B2B	Health & Personal data	Local	
34	CoverUS	Secure data sharing	No	No	C2B	Health & Personal data		
35	Medicalchain	Secure data sharing	No	No	C2B	Health & Personal data		
36	Datum Data Marketplace	Secure data sharing	No	No	C2B	Health & Personal data		
37	BIGToken	Secure data sharing	No	No	C2B	Health & Personal data		
38	BurstIQ	Secure data sharing	Yes	Yes	Any	Health & Personal dat	Global	Multiple
39	Mobility Data Marketplace	High quality and unique data	No	No	Any	Sensor & Mobility data	Local	Multiple
40	oneTRANSPORT	Easy data access and/or tooling	No	No	B2B	Sensor & Mobility data	Local	Multiple

Table 16: Overview of classification of cases in the Technology domain

		Technology domain		
#	Data marketplace	Platform architecture	Data access	Data source
1	Datahub	Centralized	Multiple options	Customer provided data
2	Data Intelligence Hub	Decentralized	Specialized software	Multiple sources
3	DAWEX	Centralized	Multiple options	Customer provided data
4	Streamr	Decentralized	Multiple options	Customer provided data
5	Databroker	Decentralized	API	Customer provided data
6	BattleFin Ensemble			Customer provided data
7	Intrinio	Centralized	Multiple options	Customer provided data
8	QuantConnect		API	
9	Knoema		Multiple options	Multiple sources
10	Open:Factset Marketplace		Specialized software	Multiple sources
11	Amazon DSP			Multiple sources
12	Salesforce Audience Studio		Specialized software	Customer provided data
13	BidTheatre		API	Customer provided data
14	RollWorks		Specialized software	Customer provided data
15	SimilarWeb		Multiple options	Multiple sources
16	Adsquare		Multiple options	Customer provided data
17	OpenPrise Data Orchestration Platform	Centralized	Specialized software	Multiple sources
18	Marketscan			Acquired data
19	Snowflake Data Marketplace	Decentralized	Multiple options	Customer provided data
20	Opendatasoft	Centralized	Specialized software	Customer provided data
21	Informatica B2B Exchange	Centralized	Specialized software	
22	Data Republic	Decentralized	Specialized software	Customer provided data
23	Otonomo	Decentralized	API	Customer provided data
24	Caruso	Centralized	API	Customer provided data
25	IOTA	Decentralized	API	Customer provided data
26	ThinkDataWorks		Multiple options	Multiple sources
27	Datax			Customer provided data
28	HERE Platform	Decentralized	API	Multiple sources
29	CARTO		Multiple options	Multiple sources
30	Factual		API	Self-generated
31	Red Lion Data		Download	Self-generated
32	Veracity	Centralized	API	Customer provided data
33	HealthVerity			Multiple sources
34	CoverUS	Centralized		Customer provided data
35	Medicalchain	Decentralized	Multiple options	Customer provided data
36	Datum Data Marketplace	Decentralized		Customer provided data
37	BIGToken	Decentralized		Customer provided data
38	BurstIQ	Decentralized	API	Multiple sources
39	Mobility Data Marketplace	Centralized		Customer provided data
40	oneTRANSPORT	Centralized	API	Self-generated

Table 17: Overview of classification of cases in the Organization domain

		Organization domain	
#	Data marketplace	Platform sponsor	Matching mechanism
1	Datahub	Independent	Many-to-many
2	Data Intelligence Hub	Independent	Many-to-many
3	DAWEX	Independent	Many-to-many
4	Streamr	Independent	Many-to-many
5	Databroker	Independent	Many-to-many
6	BattleFin Ensemble	Independent	Many-to-many
7	Intrinio	Independent	Many-to-many
8	QuantConnect	Independent	Many-to-many
9	Knoema	Independent	Many-to-many
10	Open:Factset Marketplace	Independent	Many-to-many
11	Amazon DSP	Private	One-to-many
12	Salesforce Audience Studio	Independent	Many-to-many
13	BidTheatre	Independent	Many-to-many
14	RollWorks	Private	One-to-many
15	SimilarWeb	Private	One-to-many
16	Adsquare	Independent	Many-to-many
17	OpenPrise Data Orchestration Platform	Private	One-to-many
18	Marketscan	Private	One-to-one
19	Snowflake Data Marketplace	Independent	Many-to-many
20	Opendatasoft	Independent	Many-to-many
21	Informatica B2B Exchange	Private	One-to-many
22	Data Republic	Independent	Many-to-many
23	Otonomo	Independent	Many-to-many
24	Caruso	Independent	Many-to-many
25	IOTA	Independent	Many-to-many
26	ThinkDataWorks	Private	One-to-many
27	Datax		Many-to-one
28	HERE Platform	Consortium	Many-to-many
29	CARTO	Private	One-to-many
30	Factual	Private	One-to-many
31	Red Lion Data	Private	One-to-many
32	Veracity	Independent	Many-to-many
33	HealthVerity	Independent	Many-to-many
34	CoverUS	Private	Many-to-one
35	Medicalchain	Independent	Many-to-many
36	Datum Data Marketplace	Private	One-to-one
37	BIGToken	Private	Many-to-one
38	BurstIQ	Independent	Many-to-many
39	Mobility Data Marketplace	Independent	Many-to-many
40	oneTRANSPORT	Independent	Many-to-many

Table 18: Overview of classification of cases in the Finance domain

		Finance domain				
#	Data marketplace	Revenue model	Pricing model	Price discovery	Smart contract	Payment currency
1	Datahub	Subscriptions	Freemium	Set by marketplace provider	No	Fiat
2	Data Intelligence Hub	Commissions	Multiple		No	Fiat
3	DAWEX	Subscriptions	Freemium	Set by external sellers	Yes	Fiat
4	Streamr			Set by external sellers	Yes	Crypto
5	Databroker	Commissions	Pay-per-use	Set by external sellers	No	Fiat
6	BattleFin Ensemble	Subscriptions	Flat fee tariff	Set by external sellers	No	Fiat
7	Intrinio	Subscriptions	Freemium	Set by marketplace provider	No	Fiat
8	QuantConnect	Subscriptions	Freemium	Set by external sellers	No	Fiat
9	Knoema	Subscriptions	Freemium		No	Fiat
10	Open:Factset Marketplace	Subscriptions	Pay-per-use	Set by external sellers	No	Fiat
11	Amazon DSP				No	Fiat
12	Salesforce Audience Studio	Subscriptions	Multiple	Set by external sellers	No	Fiat
13	BidTheatre	Subscriptions		Set by external sellers	No	Fiat
14	RollWorks	Subscriptions	Pay-per-use	Set by marketplace provider	No	Fiat
15	SimilarWeb	Subscriptions	Freemium	Set by marketplace provider	No	Fiat
16	Adsquare	Commissions	Multiple		No	Fiat
17	OpenPrise Data Orchestration Platform	Subscriptions			No	Fiat
18	Marketscan	Subscriptions	Multiple		No	Fiat
19	Snowflake Data Marketplace	Usage fees	Pay-per-use	Set by external sellers	No	Fiat
20	Opendatasoft	Usage fees	Freemium		No	Fiat
21	Informatica B2B Exchange	Subscriptions			No	Fiat
22	Data Republic	Subscriptions	Multiple	Set by external sellers		Fiat
23	Otonomo	Commissions	Multiple		No	Fiat
24	Caruso	Commissions	Multiple	Set by external sellers	No	Fiat
25	IOTA	Commissions		Set by external sellers		Crypto
26	ThinkDataWorks	Subscriptions	Multiple	Set by marketplace provider	No	Fiat
27	Datax	Commissions	Pay-per-use	Set by marketplace provider	No	Fiat
28	HERE Platform	Subscriptions	Freemium	Set by external sellers	No	Fiat
29	CARTO	Subscriptions	Freemium		No	Fiat
30	Factual				No	Fiat
31	Red Lion Data	Asset sales	Package based pricing	Set by marketplace provider	No	Fiat
32	Veracity	Commissions	Multiple	Set by external sellers	No	Fiat
33	HealthVerity					Fiat
34	CoverUS	Asset sales				
35	Medicalchain				Yes	Crypto
36	Datum Data Marketplace	Usage fees	Pay-per-use	Negotiation	Yes	Crypto
37	BIGToken			Set by buyers		Fiat
38	BurstIQ	Subscriptions			Yes	Fiat
39	Mobility Data Marketplace		Multiple	Set by external sellers	No	Fiat
40	oneTRANSPORT	Subscriptions	Freemium	Set by external sellers	No	Fiat

Appendix III: Definitions of the Taxonomy of Data Marketplace Business Models

Table 19: Overview of meta-characteristics, dimensions and corresponding questions to ask

Meta-characteristic	Dimension	Question to ask
Service domain	Value proposition	What is the value proposition of the data marketplace?
	Enterprise data marketplace	Does the data marketplace offer an enterprise data marketplace as a service for the private exchange of data among organizations?
	Data processing and/or analytics tools	Does the data marketplace offer data processing and/or analytics tools?
	Marketplace participants	What type of marketplace participants engage in transactions on the platform?
	Industry domain	What industry domain is the data marketplace active in?
	Geographic scope	What is the geographic scope of the data marketplace?
	Time frame	Is the data offered via the platform static data, up-to-date, (near) real-time data or multiple time frames?
Technology domain	Platform architecture	What platform architecture does the data marketplace adopt to store and access data?
	Data access	What types of data access does the data marketplace offer to users?
	Data source	Where does the data originate from and who is the author?
Organization domain	Matching mechanism	How many buyers and sellers on each side of the market are matched by the data marketplace?
	Platform sponsor	What type of actor is the sponsor and holder of the intellectual property rights of the marketplace?
Finance domain	Revenue model	What is the revenue model of the data marketplace?
	Pricing model	What is the pricing model of the data marketplace?
	Price discovery	What price discovery mechanism is employed by the data marketplace?
	Smart contract	Does the data marketplace offer smart contracts enabled by blockchain technology?
	Payment currency	What payment currency does the data marketplace accept?

Table 20: Definitions of business model characteristics in the Service domain

Service domain	Value proposition	Easy data access and/or tooling	The main value proposition of the data marketplace is easy data access and/or tooling
		Secure data sharing	The main value proposition of the data marketplace is secure data sharing among marketplace participants
		High quality and unique data	The main value proposition of the data marketplace is high quality data that is unique
		All services in a single platform	The main value proposition of the data marketplace is that all services are offered on a single platform
	Enterprise data marketplace	Yes	The data marketplace offers an enterprise data marketplace as a service
		No	The data marketplace does not offer an enterprise data marketplace as a service
	Data processing and/or analytics tools	Yes	The data marketplace offers data processing and/or analytics tools on the marketplace on top of the data offering
		No	The data marketplace does not offer data processing and/or analytics tools on the marketplace on top of the data offering
	Marketplace participants	B2B	The marketplace participants engage in business-to-business transactions
		C2B	The marketplace participants engage in consumer-to-business transactions
		Any	Any type of marketplace participant is welcome to participate in transactions on the data marketplace
	Industry domain	Any data	Any type of data is offered on the data marketplace
		Geo data	Geo data is offered on the data marketplace (e.g. agriculture, environmental, industry location and satellite data)
		Financial & Alternative data	Financial and/or alternative data is offered on the data marketplace (e.g. finance data, credit card transactions, website usage, product reviews and price trackers)
		Health & Personal data	Health and/or personal data is offered on the data marketplace (e.g. patient names, birth dates, medical treatments, health conditions, name, sex, age, home address, income)
		Audience data	Audience data is offered on the data marketplace (e.g. cookie data, clicking behaviour, timestamps individual geographic location)
		Sensor & Mobility data	Sensor and/or mobility data is offered on the data marketplace (e.g. smart city data, traffic data, parking data and automotive data)
	Geographic scope	Global	The data marketplace operates and is available to users on a global scale
		Regional	The data marketplace operates and is available to users in a certain region
		Local	The data marketplace operates and is available to users in a certain country
	Time frame	Static	The data offered on the data marketplace is static
		Up-to-date	The data offered on the data marketplace is regularly updated
		(Near)-real time	The data offered on the data marketplace is (near) real time
		Multiple	The data from multiple different time frames is offered the data marketplace

Table 21: Definitions of business model characteristics in the Technology domain

Technology domain	Platform architecture	Centralized	The data marketplace has adopted a centralized platform architecture for data storage and access
		Decentralized	The data marketplace has adopted a decentralized platform architecture for data storage and access
	Data access	API	Access to the data is offered via API
		Download	Access to the data is offered via a download option
		Specialized software	Access to the data is offered via specialized software
		Multiple options	Multiple options are offered to access the data
	Data source	Self-generated	The data marketplace generates the data offered on the data marketplace themselves
		Customer provided data	The data offered on the data marketplace is provided by customers of the marketplace
		Acquired data	The data offered on the data marketplace is acquired by the data marketplace company
		Multiple sources	The data offered on the data marketplace originates from multiple different sources

Table 22: Definitions of business model characteristics in the Organization domain

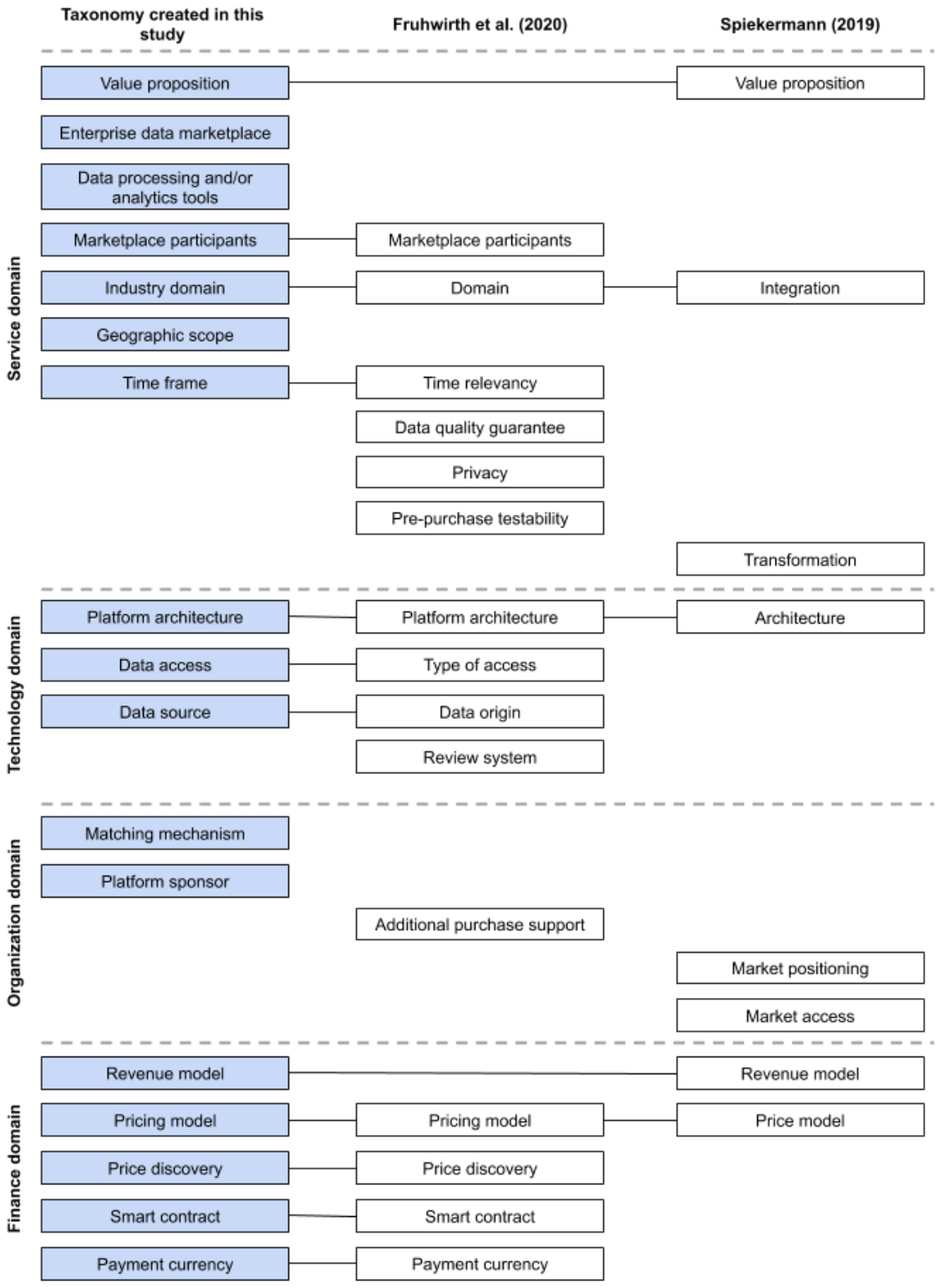
Organization domain	Matching mechanism	One-to-one	The data marketplace matches single sellers with single buyers on the marketplace
		One-to-many	The data marketplace matches a single seller with multiple buyers
		Many-to-one	The data marketplace matches a single buyer with multiple sellers
		Many-to-many	The data marketplace matches multiple sellers with multiple buyers
	Platform sponsor	Private	The data marketplace is sponsored by a private individual or group, that holds the property rights of the marketplace
		Consortium	The data marketplace is sponsored by a consortium of buyers or sellers, that hold the property rights of the marketplace
		Independent	The data marketplace is sponsored by a party that is independent of data buyers and data sellers, that positions themselves as neutral, and that holds the property rights of the marketplace

Table 23: Definitions of business model characteristics in the Finance domain

Finance domain	Revenue model	Commissions	Commissions or transactions fees are the main source of revenue for the data marketplace
		Subscriptions	Subscriptions are the main source of revenue for the data marketplace
		Usage fees	Fees paid for the usage of data products and services on the marketplace are the main source of revenue for the data marketplace
		Asset sales	The sales of data assets is the main source of revenue for the data marketplace
	Pricing model	Freemium	The data marketplace provides basic functions for free, but asks marketplace participants to pay a certain fee for premium functions on the marketplace
		Pay-per-use	Customers pay a price that is proportional to the amount of units they consume
		Flat fee tariff	Marketplace participants are offered full access to the data marketplace services for a recurring fee
		Package based pricing	Data goods and service offered on the data marketplace are bundled in packages of a certain size, for which the price may vary based on the size of the package
		Multiple	Multiple pricing models are adopted for the data product and services offered on the marketplace
	Price discovery	Set by buyers	The prices for the data offered on the marketplace are set by data buyers
		Negotiation	The prices for the data offered on the data marketplace are determined by negotiation between data buyers and data sellers
		Set by marketplace provider	The prices for the data offered on the data marketplace are set by the marketplace provider
		Set by external sellers	The prices for the data offered on the data marketplace are set by external sellers
	Smart contract	Yes	The data marketplace offers smart contracts as an option to handle transactions
		No	The data marketplace does not offer smart contracts to handle transactions
	Payment currency	Fiat money	The data marketplace utilizes fiat money as a means of payment currency for handling transactions on the marketplace
		Cryptocurrency	The data marketplace utilizes cryptocurrencies as a means of payment for handling transactions on the marketplace

Appendix IV: Comparison of the developed taxonomy and existing taxonomies

Figure 6: Comparison of the developed taxonomy and existing taxonomies



CREATING A TAXONOMY OF BUSINESS MODELS FOR DATA MARKETPLACES

Academic article

M.R. van de Ven, Delft University of Technology, Delft, The Netherlands
Word count: 11.101

Abstract

With the amount of available data growing and data posing as a strategic asset to firms, the data economy has started to evolve. Data marketplaces can fulfil a key role in realizing the data economy. The way a data marketplace operates and conducts business can be mapped and managed using a business model. As data marketplaces are a new area of research, not much research has been conducted on this type of digital platforms yet, nor on the business models of data marketplaces. Existing taxonomies of data marketplace business models mainly focus on the classification of multilateral data marketplaces and are developed from a single firm perspective on business models. This study aims to go beyond the state of the art by developing a taxonomy of data marketplaces business models from a multi-stakeholder perspective on business models. The term data marketplace is broadly interpreted in this research to also allow the inclusion of atypical forms of data marketplaces. A design science approach is employed and a standard taxonomy development method by Nickerson et al. (2013) is followed to develop the taxonomy. The final taxonomy comprises of 4 meta-dimensions, 17 business model dimensions and 59 business model characteristics. The results of this study contribute to the literature by improving the understanding of the notion of data marketplace business models and by providing a framework that can be utilized for the classification of data marketplace business models and for the analysis of business model patterns and business model archetypes.

Keywords: Business Model, Data Marketplace, Taxonomy, Dimensions, Characteristics, Literature review, Desk research

1. Introduction

In recent years, the amount of available and generated data has staggered. DOMO (2019) estimated that by 2020 we will have generated 40 times more data bytes than the number of stars in the observable universe. As data can be processed into comprehensible information, the vast amounts of data have become an important resource for innovation and economic growth for businesses and society in general. Organizations have long been using data as input for decision-making and process optimization. However, the deployment of connected devices on the Internet of Things and advances in data analytics have made it easier for organizations to collect and analyze data (Hürtgen & Mohr, 2018). As all the newly collected and processed data can be monetized and traded, it can pose as a strategic asset to organizations (Opher et al., 2016).

As a result of this development, data economies have started to evolve (European Commission, 2017). A data economy is a digital ecosystem in which a network of vendors gather, organize and exchange data (European Commission, 2017). The vendors trade their proprietary data with firms or individuals, often charging a fee for the exchanged products and services. Data marketplaces can fulfil a key role in realizing the data economy. As an organizations may not always possess the required data to carry out or improve their processes and services, they may wish to purchase these data from other organizations. A

data marketplace can address this issue by providing a digital platform through which individuals and organizations can exchange data (Stahl et al., 2016; Schomm et al., 2013).

Despite the potential benefits of data marketplaces, in practise very little data is shared or traded via platforms (Koutroumpis et al., 2020). Many data marketplaces that have been set up have failed or are shut down (Ramel, 2016, Johnson, 2012; Dodds, 2012; Kosara, 2010). The way a data marketplace operates and conducts business can be mapped and managed using a business model. A business model is a description of how a network of organizations creates and captures value (Bouwman et al., 2008). Only a small number of data marketplaces currently exist, and most of the existing marketplaces adopt non-profit business models (Koutroumpis et al., 2017). Surveys among data marketplace providers show that existing data marketplaces are adopting limited business models, focussed on secure revenue streams and hierarchical organizational settings (Stahl et al., 2017).

In general, little research has been conducted on data marketplaces (Thomas & Leiponen, 2016) and data marketplace business models in particular (Fruhwirth et al., 2020; Spiekermann, 2019). Two taxonomies of data marketplace business models are currently available in the literature (Fruhwirth et al., 2020; Spiekermann, 2019). A taxonomy is a classification scheme of an object of interest, from which wider generalizations can be made (Lambert, 2015). The existing taxonomies are lacking in some areas, that this study aims to improve. Firstly, the two studies mainly focus on multilateral data marketplaces, in which the data marketplace functions as neutral intermediary that matches multiple data buyers with multiple data sellers (Koutroumpis et al., 2017). However, in practice data is rarely traded via multilateral data marketplaces, and instead data trading often happens via bilaterally negotiated contracts (Koutroumpis et al., 2017). Secondly, the two existing taxonomies are structured using business model ontologies that view the business model from a single firm perspective: Fruhwirth et al. (2020), a paper based on the Master's thesis of Prlja (2019), utilizes a combination of Teece (2010) and Remane et al. (2017) to structure their taxonomy, and Spiekermann (2019) makes use of a combination of Gassman et al. (2014) and Al-Debei et al. (2008) respectively. However, studies show that data marketplaces take part in a network of stakeholders, among which data analysts, application vendors, algorithm developers, data providers, consultants, licensing entities, platform providers (Spiekermann, 2019; Chakrabarti et al., 2018; Thomas & Leiponen, 2016; Muschalle et al., 2012). Thirdly, the data economy is emerging and more data marketplaces are being set up, and therefore new business model alterations may have been produced in practice that were not considered during the development of the existing taxonomies.

Therefore, this study aims to go beyond the state of the art by developing a taxonomy from a multi-stakeholder perspective on business models (Bouwman et al., 2008). The term data marketplace is broadly interpreted in this research, to also allow for the inclusion of atypical types of data marketplaces other than multilateral data marketplaces. New business model alterations that have been produced in practice will be considered. The main research question that summarizes the academic challenge is:

How can the business model characteristics of different types of data marketplaces be classified into a taxonomy from a multi-stakeholder perspective?

To build a taxonomy of data marketplace business models, the taxonomy development method by Nickerson et al. (2013) is used. This method combines techniques from typology development (conceptual) and taxonomy development (empirical) (Nickerson et al., 2013). The

method by Nickerson et al. (2013) has previously been used in business model taxonomy development studies, in the fields of among others Blockchain-based business models (Weking et al., 2019), IoT platform business models (Hodapp et al., 2019), and carsharing business models (Remane et al., 2016). To structure the research, a design science approach is employed, that comprises of three complementary cycles of research activities: the rigor cycle, relevance cycle and design cycle (Hevner, 2007). The taxonomy development method was designed in such a way that it is analogous to the design science approach, and therefore includes a rigor, relevance and design cycle (Nickerson et al., 2013).

The remainder of this paper is structured as follows: section 2 outlines the theoretical background of this research by introducing the concepts markets, marketplaces and electronic marketplaces, the notion of a data marketplace, and the object of interest of this study, data marketplace business models. In section 3, the taxonomy development process is described. Subsequently, section 4 presents the final taxonomy on the basis of the identified business model dimensions and characteristics. Section 5 provides a demonstration of the use of the taxonomy. Lastly, the he scientific contribution, societal and managerial implications and limitations of the research are discussed in section 6, and possible directions for future research are addressed.

2. Theoretical background

2.1. Business Models

In the business model literature, three main research streams can be distinguished: Information Systems, Strategy, and Innovation and Technology Management (Bouwman et al., 2019). An important difference between the various business model definitions and ontologies in the tree research streams is that some authors view the business model from the perspective of a single company (Teece, 2010; Casadesus-Masanell & Ricart, 2010; Chesborough & Rosenboom, 2002), while other authors view the business model from a multi-stakeholder perspective, where a network of companies collaborates together (Bouwman et al., 2008; Gordijn & Akkermans, 2001; Amit & Zott, 2001; Timmers, 1998).

In this study, the business model definition by Bouwman et al. (2008) is employed as the leading perspective on data marketplace business models. Bouwman et al. (2008) define the business model as *"a blueprint for a service to be delivered, describing the service definition and the intended value for the target group, the sources of revenue, and providing an architecture for the service delivery, including a description of the resources required, and the organizational and financial arrangements between the involved business actors, including a description of their roles and the division of costs and revenues over the business actors"* (p. 3). In this definition, the perspective on business models is extended from a single company to the level of an enterprise: a network of companies that collaborate together to offer consumers a joint value proposition. The service that is offered to the customer and delivered by the network of stakeholders is central in this definition.

2.2. Digital Platforms

The data marketplace provider offers a digital infrastructure that enables marketplace participants to buy or sell data goods (Stahl et al., 2016). In an advanced form of a data marketplace, the platform functions as a digital intermediary that provides value to data buyers, data sellers, and third-party service providers (Carnelley et al., 2016). In this form, the data

marketplace operates as a multi-sided digital platform. Digital platforms have two key characteristics that data marketplaces may hold.

Firstly, digital platforms function as two-sided or multi-sided intermediaries. In economic theory, a platform functions as a mediator between different user groups (Rochet & Tirole, 2003). Economic value is created on the digital platform through interactions between pairs of end users, such as buyers and sellers, in the form of transactions. Platforms that mediate between multiple groups of users are also called multi-sided platforms (Boudreau & Hagiu, 2009). By connecting different user groups, digital platforms create network externalities (also referred to as network effects) that represent the increase in usefulness of a good for a user that arises when the usage of the good by other users increases (Katz & Shapiro, 1985).

Secondly, a digital platform functions as a breeding ground for innovation. From a technical perspective, a digital platform functions as a code base that can be extended with third-party modules, such as software and applications (Tiwana et al., 2010). The platform has interfaces through which third-party complementors can interact and exchange data and information with the platform. The extendibility of the digital platform is enabled by its technical architecture: digital platforms comprise of a modular and stable set of core components with a variable periphery of complementary components (Baldwin & Woodward, 2009). The two sets of components interact and are governed via interfaces at the boundaries of the platform. This combination of stability and variety in the technical architecture allows modular innovation: while the interfaces of the platform remain stable, the core components and complementary components may change over time (Henderson & Clark, 1990).

It is important to note that not all data marketplaces function as multi-sided intermediaries or breeding grounds for innovation. While matchmaking between one or more buyers and sellers is an important requirement for electronic marketplaces and data marketplaces, multi-sided matchmaking between a multitude of parties (buyers, sellers and third-party service providers) is not.

2.3. Electronic Marketplaces

In practise, the terms market and marketplace are often confused. From an economics perspective, markets are viewed as an abstract place where actors (often: buyers and sellers) meet to exchange goods and services at certain price and quantity that they agree on (Stahl et al., 2016). Conversely, a marketplace is an explicit place in terms of time and location, where the exchange of goods or services among actors takes place (Stahl et al., 2016). In other words, a marketplace provides the physical or virtual infrastructure where products are traded, which enables the abstract concept of a market. For a market to be classified as an electronic market, at least the negotiation between the buyer and the seller of a product offering needs to be carried out electronically (Stahl et al., 2016). In the same way, an electronic marketplace provides a digital infrastructure on which market participants interact online (Stahl et al., 2016).

Electronic marketplaces have a major impact on the functions of a market, compared to general markets. Electronic marketplaces make use of information technologies to improve the matching of buyers and sellers (Bakos, 1998). Advantages that the use of information technologies in electronic marketplaces may offer are: increased personalization, cost-effective customization of product offerings, decreased search costs for buyers, lower communication costs for sellers, and new ways of price discovery (Bakos, 1998). Furthermore, electronic marketplace decrease facilitation costs by enabling online information sharing to decrease logistics costs and by offering online payment methods to lower transaction costs (Bakos, 1998).

2.4. Data Marketplaces

In early research, data marketplaces are viewed as a platforms on which any individual or organization is invited to buy, sell, upload and maintain datasets, and where data access and usage is regulated via a variety of licensing models (Schomm et al., 2013). In more recent research, such platforms are denoted as multilateral data marketplaces, where the marketplace provider functions as a neutral intermediary that matches multiple data buyers with multiple data sellers (Koutroumpis et al., 2017). It is important to note that in practise, data marketplaces often exchange access to data and data-related services rather than explicitly selling data goods (Koutroumpis et al., 2020).

Data marketplaces are more than just repositories of data sets or providers of cloud services; they function as market makers, enabling exchange of data between data providers and data consumers (Carnelley et al., 2016). In contrast with most other platforms, where data is utilized to improve services or manage customer relationships, on data marketplaces data is actually the product itself (Spiekermann et al., 2018). This way, data marketplaces enable new data-driven business models for data buyers, data sellers, and third-party service providers (Hartmann et al., 2014).

The data marketplace provider offers a digital infrastructure that enables marketplace participants to buy or sell data goods (Stah et al., 2016). In an advanced form of a data marketplace, the platform functions as a digital intermediary that provides value to data buyers, data sellers, and third-party service providers (Carnelley et al., 2016). In this form, the data marketplace operates as a multi-sided digital platform that allows value creation in the form of transactions and innovation.

In this study, the term data marketplace is broadly interpreted on the basis of two assumptions. Firstly, a data marketplace is interpreted as a marketplace that enables the commercial trading of data as a valuable good (Koutroumpis et al., 2020). Secondly, as data is a digital good, it is assumed that the trading of data takes place on an electronic marketplace, a digital infrastructure that is provided by the a marketplace provider (Stahl et al., 2016). Thus, in this study a data marketplace is defined as *the digital infrastructure on which the commercial trading of data as a valuable good takes place.*

A data marketplace may decide to function as a multi-sided digital platform by allowing third-party service providers to interact with data buyers and data sellers on the marketplace. In principle, all platforms comprise of an ecosystem with the same fundamental structure, that comprises of four different players: platform owners, providers, producers, and consumers (Van Alstyne et al., 2016). In the case of data marketplaces, these translate to the data marketplace owner, data providers, third-party service providers and data buyers (Spiekermann, 2019).

2.5. Data Marketplace Business Models

The STOF ontology by Bouwman et al. (2008) was selected as the leading perspective on data marketplace business models in this study. In the STOF ontology, a business model is viewed as the mapping of how a network of organizations aims to create and capture value (Bouwman et al., 2008). The STOF approach takes service as a unit of analysis and employs a multi-stakeholder perspective on business models. This approach suits data marketplaces, as a network of business actors are involved in and around data marketplaces, among which data buyers, data sellers and external service providers (Spiekermann, 2019; Chakrabarti et al.,

2018; Thomas & Leiponen, 2016; Muschalle et al., 2012). Moreover, the approach is well-suited for data marketplace business models because the main aim of data marketplace companies is to provide a marketplace as a service for the commercial trade of data between data buyers and data sellers (Koutroumpis et al., 2020). Additional value adding services such as data processing and visualization services may be offered on top of the data (Spiekermann, 2019).

Based on the theories on electronic marketplaces, a data marketplace was defined as *the digital infrastructure on which the commercial trading of data as a valuable good takes place*. The four main roles of players in the data marketplace ecosystem were defined as: the data marketplace owner, data providers, third-party service providers and data buyers. Combining this definition with the STOF approach to business models, a data marketplace business model is defined as: *The mapping of how a data marketplace enterprise aims to create and capture value by providing a marketplace and additional value adding services for the commercial trade of data between data providers and data buyers.* In this definition, the data marketplace enterprise may comprise of the single data marketplace provider or a combination of a data marketplace provider and external service providers.

3. Methodology

In this research, a design science approach is employed to design a taxonomy of data marketplace business models. This approach ensures that the taxonomy is based on practical relevance and scientific rigor (Hevner, 2007). The taxonomy of data marketplace business models is designed and evaluated by following a standard taxonomy development method (Nickerson et al., 2013). In this section, the research method that was followed is described.

3.1. Literature review

To ensure scientific rigor, a literature review was conducted to discover existing scientific theories and artifacts about the object of interest (Webster & Watson, 2002). The scientific database Google Scholar was consulted to find relevant academic sources, using the search string “Data marketplaces” AND (“Business models” OR “Digital platform” OR “Digital marketplace” OR “Data trading” OR “Data economy”). This string resulted in a total of 359 articles.

The articles were scanned based on their title, abstract and relevance, which resulted in a preliminary selection of 17 articles. After making this pre-selection of articles, the full text of the articles was read. Special attention was paid to whether the studies discussed dimensions and characteristics of data marketplaces and data marketplace business models. The full reading of the text resulted in the exclusion of 7 articles, that did not explicitly discuss dimensions or characteristics. The articles that were excluded from the list of relevant articles were not fully omitted, but they were used to provide background information in Chapters 1 and 2 of this research and to complement the information from the relevant sources.Based on the literature mentioned in the selected articles, 4 additional articles that presented topic-relevant business model taxonomies were added to the list.

The literature review resulted in a final set of 14 articles, which are presented in Table 1. For every article, an indication is given about the type of research. The characteristics of data marketplace business models that result from the literature review serve as input for input for the taxonomy design process.

Table 1: Overview of classifications and taxonomies relevant to data marketplace business models

Author(s) (Year)	Title	Type	Citations (14.04.2020)
Schomm et al. (2013)	Marketplaces for data: an initial survey	Dimensions of data providers and data marketplaces	73
Stahl et al. (2014a)	Data Marketplaces: An Emerging Species.		14
Stahl et al. (2014b)	The data marketplace survey revisited		16
Stahl et al. (2017)	Marketplaces for digital data: Quo vadis?		12
Stahl et al. (2016)	A classification framework for data marketplaces	Classification of electronic marketplaces	30
Koutroumpis et al. (2017)	The (unfulfilled) potential of data marketplaces	Market designs for data marketplaces	19
Muschalle et al. (2012)	Pricing approaches for data markets	Pricing models for data marketplaces	74
Fricker & Maksimov (2017)	Pricing of data products in data marketplaces		8
Spiekermann (2019)	Data marketplaces: Trends and monetisation of data goods	Taxonomy of data marketplace business models	9
Fruhwirth et al. (2020)	Discovering Business Models of Data Marketplaces		1
Bock & Wiener (2017)	Towards a Taxonomy of Digital Business Models-Conceptual Dimensions and Empirical Illustrations	Taxonomy of digital business models	22
Täuscher (2016)	Business Models in the Digital Economy: An Empirical Classification of Digital Marketplaces	Taxonomy of digital marketplace business models	6
Täuscher & Laudien (2018)	Understanding platform business models: A mixed methods study of marketplaces	Taxonomy of marketplace business models	153
Hartmann et al. (2014)	Big data for big business? A taxonomy of data-driven business models used by start-up firms	Taxonomy of data-driven business models	131

3.2. Constituting a database of empirical cases

To account for the practical relevance of the to be designed artifact, desk research was conducted to constitute a database of empirical cases of data marketplaces (Hevner, 2007). A number of different sources that link to data marketplace websites were considered. 65 websites of data marketplaces that were mentioned and analysed in existing studies of data marketplaces were included in the database (Koutroumpis et al., 2020, 2017; Prlja, 2019; Spiekermann, 2019, Stahl et al., 2016; Carnelley et al., 2016). The data discovery platform datarade.ai was consulted, a website that provides an overview of 1800+ data providers, 200+ data platforms and 200+ data categories. In the database of data platforms on the website, the categories ‘audience data marketplace’, ‘data marketplaces & exchanges’, ‘personal data marketplaces’, ‘IoT data marketplaces’, ‘alternative data marketplaces & platforms’, ‘financial data marketplaces’, ‘second party data marketplaces’ and ‘B2B data marketplaces’ were consulted to find data marketplaces to add to the database of this research. In total, the search in the categories of datarade.ai resulted in the discovery of 187 data marketplaces. To complement the database with data marketplaces that were not considered in the existing studies or part of the datarade.ai database, the search engine Google was utilized to conduct a desk research. The keywords “data marketplace”, “data market” and “data trading platform”

were applied during the search. This resulted in an additional 15 data marketplaces that were added to the database.

To ensure that the sample of empirical cases contained relevant data marketplaces, a number of criteria were applied to the companies that resulted from the desk research: Data marketplaces that turned out to be closed after inspection of the website were excluded from the database. The websites were inspected and analysed to make sure that the companies fit the definition of a data marketplace that is employed in this study. The companies that did not fit the definition were excluded. Data marketplaces that did not have an English version of their website or of which the English version seemed outdated compared to the webpage in the native language were excluded from the database. Data marketplaces that only provided open data, such as governmental organizations and NGOs were excluded from the database. Data marketplaces that were still in the construction phase were excluded. The application of these five criteria to the cases resulting from desk research lead to the exclusion of 89 cases. The final database consisted of 178 cases of data marketplaces for further consideration.

To explore the variety between cases in the database, the 178 cases in the database were segmented and labelled based on the different types of data that is traded on the data marketplaces. The database of datarade.ai provided an indication of the type of data marketplace based on the labelling per category of data marketplaces (‘audience data marketplace’, ‘data marketplaces & exchanges’, ‘personal data marketplaces’, ‘IoT data marketplaces’, ‘alternative data marketplaces & platforms’, ‘financial data marketplaces’, ‘second party data marketplaces’ and ‘B2B data marketplaces’). Therefore, 138 cases that originated from datarade.ai database could be labelled. If applicable, the remaining 40 cases were labelled based on the classification of data marketplaces in existing scientific classification studies (Fruhwirth et al., 2020; Spiekermann, 2019), and through the inspection of the companies’ website.

3.3. Sample selection

To analyse the business models of existing data marketplaces, a representative sample was taken from the database of cases. In the taxonomy development process, a within case analysis is conducted to analyse the business model characteristics of the empirical cases. The results of the within case analysis serve as input for the iterative taxonomy development process.

The empiricist philosophy of classification prescribes to build a taxonomy based on the consideration of many characteristics (Lambert, 2015). Therefore, the cases of data marketplaces in the database was first segmented into groups based on the similarity of their characteristics, to ensure that the sample size included data marketplaces with varying characteristics. To explore the variety between cases in the database, the 178 cases in the database were segmented and labelled based on the different types of data that is traded on the data marketplaces. The database of datarade.ai provided an indication of the type of data marketplace based on the labelling per category of data marketplaces. Therefore, 138 cases that originated from datarade.ai database could be labelled. If applicable, the remaining 40 cases were labelled based on the classification of data marketplaces in existing scientific classification studies (Fruhwirth et al., 2020; Spiekermann, 2019), and through the inspection of the companies’ website.

From the segmentation of data marketplaces by type of data traded on the platform, it became clear that some data marketplace types in the database were overrepresented compared to others. This was especially the case for audience data marketplaces, that made up over 60% of the cases (N=112). Therefore, instead of random sampling, the disproportionate stratified

sampling method was applied to compensate for the overrepresentation of some types of data marketplaces in the database (Daniel, 2011).

A sample of N=40 cases was sampled from the database, following the eight steps of disproportionate stratified sampling by Daniel (2011). The final sample of 40 data marketplaces consisted of 10 data marketplaces on which *any type* of data is traded (25% of the sample), 4 *financial and alternative data* marketplaces (10%), 9 *audience data* marketplaces (22,5%), 6 *sensor and mobility* data marketplaces (15%), 4 *geo data* marketplaces (10%) and 7 *health and personal* data marketplaces (17,5%). By lowering the proportion of audience data marketplaces to 22,5%, the larger size of this segment was still taken into account compared to other categories, but it would not dominate or be overrepresented in the sample.

3.4. Taxonomy development process

The first step of the taxonomy development process is to define the meta-characteristics of the taxonomy (Nickerson et al., 2013). The meta-characteristic functions as an overarching characteristic that provides the foundation for choosing characteristics of the object of interest (Nickerson et al., 2013). Each determined characteristic follows logically from the meta-characteristic. When the meta-characteristics and ending conditions are set, the design phase can start. In the design phase of the taxonomy development method, researchers can choose between two possible design approaches: the conceptual-to-empirical approach and the empirical-to-conceptual approach (Nickerson et al., 2013). In this study, both approaches are employed to build the taxonomy. The taxonomy design is evaluated by checking the framework with the pre-defined ending conditions. When the results of the evaluation of ending conditions are not satisfactory, researchers may choose to go back to the drawing table and pick one of two approaches to revise and improve the taxonomy. This way, multiple iterations may be conducted. When the ending conditions are met, the taxonomy development process ends.

In this study, the four business model domains of the STOF ontology provide the meta-characteristics of the business model taxonomy (Bouwman et al., 2008; Faber et al., 2003):

- **Service domain:** a description of the value that the network of actors aims to deliver to the customer, specifically the service offering.
- **Technology domain:** a description of the technical architecture that the value network utilizes to deliver the proposed service offering as explicated in the service domain.
- **Organization domain:** a description of the organization of actors in the network, the value network, explicating the roles they take on and the value activities they perform to deliver the service and create value for the customer.
- **Finance domain:** a description of how the value network aims to make money from the service offering and how the costs and revenue streams are split among the different actors in the network.

Besides providing the four overarching dimensions of the taxonomy, the STOF ontology provides logic by which the various business model characteristics of data marketplaces are classified. Then, requirements are set that define the ending conditions of the taxonomy development process (Nickerson et al., 2013). The eight objective ending conditions and five subjective ending conditions that are suggested by the authors of the selected taxonomy development method were employed to terminate the iterative taxonomy development process (Nickerson et al., 2013).

When the meta-characteristics and ending conditions were set, the design phase took off. First, a conceptual framework was developed in the form of a preliminary taxonomy, based on the concepts from the data marketplace literature and relevant dimensions from the identified

business model taxonomies. Then, the iterative design process started. In this study, the design phase started with a conceptual-to-empirical approach. In these design iterations, the preliminary business model taxonomy was applied to the sample of empirical cases to see if the conceptual dimensions and characteristics corresponded with the empirical dimensions and characteristics. To map the business model characteristics of the selected cases on the preliminary taxonomy, information on the business models of the cases was collected from publicly available sources. Main sources of information were company websites and blogs and news articles, to which the companies commonly link to from their website. Many companies often provided a whitepaper on their website with the vision and mission of the company, that sometimes provided information about the business model of the data marketplace. To gather information about the business model characteristics of the sampled data marketplaces, a within case analysis was conducted to gather relevant information in the form of text fragments, pictures, screenshots and other informational elements. The discovered information fragments were coded using the dimensions and characteristics of the preliminary taxonomy as a guideline. Table 2 provides an example of the coding logic.

Table 2: Coding examples for the value proposition dimension

Characteristic	Case	Quote
Easy data access and/or tooling	Open:Factset Marketplace	*"FactSet creates data and technology solutions for investment professionals around the world, providing instant access to financial data and analytics that investors use to make crucial decisions."*
	Knoema	*"Knoema is a cloud-based data technology platform that makes data accessible and delivers intelligent data tools to enable data access and discovery."*
Secure data sharing	DAWEX	*"With Dawex Global Data Marketplace providers can highlight the value of their data while retaining full control over the distribution and configuration of usage rights."*
	Snowflake	*"Unlike other data marketplaces, Snowflake Data Marketplace leverages Snowflake's Secure Data Sharing technology, which means no data transfer and no need to squeeze data through APIs or use cloud storage."*
High quality and unique data	Amazon DSP	*"Use exclusive Amazon audiences to reach your ideal audience on and off Amazon."*
	Datax	*"Quality business data for better sales leads - Any campaign is only as good as the data it's built on – so make sure yours is the best.*

The identified characteristics that resulted from the within case analysis were specified in a comprehensive table for each case. If the identified characteristics of a data marketplace were not yet specified in the preliminary taxonomy, the taxonomy is revised by adding the new characteristics to the existing dimensions. The two conceptual-to-empirical iterations were resulted in the refinement of the meta-characteristics with the following business model dimensions: the *value proposition* (Spiekermann, 2019), *marketplace participants* (Fruhwirth et al., 2020), *industry domain* (Fricker & Maksimov, Schomm et al., 2013), *geographic scope* (Täuscher & Laudien; 2018; Täuscher, 2016) and *time frame* (Schomm et al., 2013) in the *Service domain*; the *platform architecture* (Fruhwirth et al., 2020; Spiekermann, 2019; Koutroumpis et al., 2017), *data access* (Fruhwirth et al., 2020; Schomm et a., 2013) and *data source* (Hartmann et al., 2014) in the *Technology domain*; the *matching mechanism* (Koutroumpis et al, 2017) and *platform sponsor* (Stahl et al., 2017, 2016) in the *Organization domain*; and the *revenue model* (Täuscher & Laudien, 2018; Täuscher, 2016), *pricing model* (Fruhwirth et al., 2020; Spiekermann, 2019; Fricker & Maksimov, 2017; Schomm et al., 2013),

price discovery (Täuscher & Laudien, 2018; Täuscher, 2016), *smart contract* (Fruhwirth et al., 2020) and *payment currency* (Fruhwirth et al., 2020) in the *Finance domain*. After that, two empirical-to-conceptual iterations were conducted, which resulted in the addition of two binary dimensions to the taxonomy: *enterprise data marketplace* and *data processing and analytics tools*.

After every design iteration, the ending conditions were checked. The selected taxonomy development method by Nickerson et al. (2013) provided eight objective ending conditions and five subjective ending conditions. After two conceptual-to-empirical iterations and two empirical-to-conceptual iterations, both the objective and subjective ending conditions were met.

4. Final taxonomy

The final taxonomy consists of four meta-dimensions, 17 dimensions and 59 characteristics, and is presented in Table 3. In the following sections, the business model dimensions and characteristics are discussed per meta-dimension.

4.1. Service domain

The business model dimensions in the service domain are the starting point of the STOF approach to business model design (Bouwman et al., 2008). 'Value' is the central issue in the service domain: the data marketplace provider and other service providers intend to collaborate as an enterprise to deliver a certain value proposition to the customer, that in turn expects or perceives a certain value from the value offering.

The **value proposition** is a statement that indicates the proposed value that an enterprise intends to deliver to the customer (Bouwman et al., 2008). It often describes how customers can benefit from using the service and how the enterprise aims to set itself apart from the competition. Organizations usually present their value proposition on the first page of their website, to clearly communicate their intended value to the customer. Data marketplaces can be characterized by five value propositions: easy data access and/or tooling, secure data sharing, high quality and unique data and all services in a single platform.

Some data marketplaces offer an **enterprise data marketplace** as an additional service. An enterprise data marketplace, sometimes also referred to as 'data exchange', functions as a private data marketplace that enables organizations to share data within the company or with external partners, such as suppliers, customers and other players that are invited to the platform by the focal organization. Marketplace participants are able to present data sets in a shared environment for other participants to use. This way data sharing is shifted from a demand-based model, in which departments and partners have to make requests for data, to a supply-based model, wherein the datasets available for sharing are presented on the platform. For some companies, the enterprise data marketplace is their main offering, and they incorporate external data in the data marketplace for data enrichment and analytics activities within the marketplace environment. Prominent companies that offer an enterprise data marketplace are DAWEX, Snowflake and Data Republic.

The **data processing and analytics tools** characteristic refers to the tooling that is offered on top of the data, often in a workspace environment, where data and tooling buyers can perform analytics activities on their proprietary data or data bought from the platform. Some companies offer a large variety of tools on top of their data, such as Data Intelligence Hub (by T-Systems), while other companies do not offer tooling, such as Red Lion Data, focussing solely on the data offering on their marketplace.

The specification of users or customers is a key element in business models (Bouwman et al., 2008). Data marketplaces can choose to direct their platform to individual consumers or businesses on both the supply-side and the demand-side (Fruhwirth et al., 2020). Three types of variations of **marketplace participants** are distinguished: business-to-business (B2B), consumer-to-business (C2B) or any combination of business and consumers (Fruhwirth et al., 2020). B2B data marketplaces direct themselves specifically to organizations and businesses that are willing to become more data-driven or possess a large amount of data that they wish to monetize or commercialize. Many C2B data marketplaces act as harvesting data marketplaces, that gather the personal data of users in exchange for rewards. Lastly, some data marketplaces are open for any party, business or consumer, to register and exchange data on the marketplace (Schomm et al., 2013).

Based on the analysis of empirical cases, data marketplaces are providing their data goods and services in the following **industry domains**: any data, geo data, financial and alternative data, health and personal data and sensor and mobility data. A number of data marketplaces allow the exchange of any data on their marketplace. An example of a data marketplace that provides the exchange of any type of data is Databroker (rebranded from Databroker DAO in 2019), that extended its scope from IoT data to all types of data, with the goal to be the to-go-to marketplace for data. Geo data refers to data that has a link with a location on the Earth. Geo data is sometimes also referred to as geospatial or geographical data. This type of data is often stored and used in geographical information systems (GIS). The finance and alternative data industry domain refers to data marketplaces active in the financial industry. Finance data comprises of datasets that provide information about the financial state of a company, such as data about a companies' assets, liabilities and equity. Alternative datasets provide information about a company that is published by sources outside of the company. Alternative data may provide unique insights about investment opportunities. Both financial and alternative data are used by investment professionals such as hedge fund managers, venture capitalists, private equity funds and investment bankers to make investment or divestment decisions. Data marketplaces in the health and personal data industry domain often function as harvesting data marketplaces that provide rewards to customers for providing their health or personal data. Health data refers to e.g. patient names, birth dates, medical treatments and health conditions of individuals or the population. Examples of personal data are name, sex, age, home address and income. Audience data is combined data about a certain target group of customers, the 'audience'. Marketeers aim to gather as much data about their envisioned audience as possible, to target the audience with highly personalized and relevant offers. In many cases, audience data is gathered by a data provider company through the automatic or manual scanning of user behaviour on websites and mobile applications. Data marketplaces in the sensor & mobility data industry provide sensor data gathered by Internet-of-Things sensors, such as smart city data, traffic data, parking data and automotive data. The data offered in this industry domain is often (near) real-time, because the sensor data is directly sent to the data marketplace by transferring the streaming data via APIs.

The **geographic scope** describes the regions in which the data marketplace is operating and available to users. A distinction is made between global data marketplaces, regional data marketplaces and local data marketplaces (Täuscher & Laudien, 2018; Täuscher, 2016). Global marketplaces serve clients across two or more continents. Regional data marketplaces focus on multiple countries in a single continent or region. Lastly, local marketplaces solely focus on a single country.

The data traded on the data marketplace may have a certain temporal context in a **time frame**, that describes whether or not the data needs frequent updates to maintain the relevancy of the data (Schomm et al., 2013). A distinction is made between static datasets, up-to-date datasets,

(near) real time datasets, and data marketplaces that offer datasets with multiple time frame relevancies. An example of a static data are the labelled datasets sold by Datax, that crowdsources data by asking consumers to label images, recordings and dialogues. Up-to-date datasets are essentially static datasets, that are repeatedly updated by the marketplace provider or the external data sellers on the data marketplace. A number of data marketplaces offers real-time or near-real time data. This type of data is often generated by IoT sensors or online data trackers, such as website and stock market trackers.

4.2. Technology domain

The requirements specified in the service domain determine the identification and specification of the technical architecture in the technology domain of the business model (Bouwman et al., 2008).

Data marketplaces may adopt two types of **platform architectures**: centralized and decentralized (Koutroumpis et al., 2017). In the centralized approach, data providers offer their data products via a predefined location central on the platform, such as a cloud repository. This type of platform architecture provides better control over data access and enables data buyers to directly process the data. In decentralized platforms, the data products remain at the data provider and the data is traded using distributed ledger technologies (Koutroumpis et al., 2017). The decentralized approach enhances data provenance, but makes data processing and storage more challenging for the platform users (Koutroumpis et al., 2017). Decentralized data marketplace architectures are emerging. Two examples of companies that have adopted such an architecture are the Data Intelligence Hub and Snowflake Data Marketplace. The Data Intelligence Hub has implemented the security standards by the International Data Spaces Association (IDSA). The platform has adopted a decentralized approach, in which data is transferred directly from data sellers to data buyers through a secured line. This way, the data never passes through the Data Intelligence Hub platform itself. A similar approach is adopted by Snowflake Data Marketplace: by implementing Snowflake's Secure Data Sharing technology, the data offered on the marketplace does not move from data provider to data consumer, but remains at the data provider. No data is transported, pushed through APIs or stored in a cloud. Instead, data consumers have direct 'read-only' access to the datasets of the data provider. The data providers in turn have control over who can access their data.

Platform providers may provide **access to the data** in a number of different ways (Schomm et al, 2013): via APIs, direct download options, specialized software or via multiple of the aforementioned options. Data marketplaces that offer data access via APIs develop a predefined software protocol to establish an interface that enables access and interaction with the platform. In the download option of data access, the data is accessed via a download file and there is no need for developing a software component. Some data marketplaces develop specialized software to provide access to the data on the marketplace. A large number of data marketplaces in the sample offered multiple options to access the data, either via APIs, direct download options and specialized software.

The **data source** dimension describes the origin where the data was gathered or collected by the data marketplace platform (Hartmann et al., 2014). The following data sources are distinguished for data marketplaces: self-generated data, customer provided data, acquired data or data from multiple of the aforementioned sources. Data marketplaces may have generated data themselves, by for instance gathering data manually or automatically from the internet. Furthermore, the data marketplace may also invite customers to provide their proprietary datasets on the platform. Marketscan is an example of a data marketplace that acquires data from external data providers. The company integrates data from the feeds of five large UK data suppliers, and then verifies and aggregates the data in the central database to

ensure high data quality and coverage. Lastly, some data marketplaces retrieve data from multiple types of sources. Data Intelligence Hub, a data marketplace that aims to offer all sorts of data, retrieves data from open data portals and publishes them on the platform, but also invites commercial data providers to sell their data offerings on the marketplace.

4.3. Organization domain

The technologies that are used to deliver the service to the customer depend on the organization design of actors that take ownership and invest in these technologies (Bouwman et al., 2008). Central in the organization domain of the business model is the value network of actors that is needed to realize the service offering.

The **matching mechanism** of a data marketplace determines the number of parties on each side of the platform (Koutroumpis et al., 2017). With regard to data marketplaces, the following variations of matching exist: one-to-one matching, one-to-many, many-to-one, and many-to-many. One-to-one matching mechanisms can be characterized by negotiated terms of exchange (Koutroumpis et al., 2017). Data marketplaces that adopt a one-to-many matching mechanism mediate between a single seller and many buyers (Koutroumpis et al., 2017). This type of data marketplaces are also called dispersal data marketplaces. In many-to-one matching, many sellers are trading data with a single buyer at the same time, is used in harvest marketplace designs (Koutroumpis et al., 2017). Finally, Data marketplaces that adopt the many-to-many matching model, allow any user to upload and maintain datasets on the platform (Schomm et al., 2013).

The **platform sponsor** constitutes and holds the property rights of the platform components, rules and ecosystem (Eisenmann et al., 2009). The platform can be sponsored by a private individual or group, a consortium of buyers or sellers on the supply or demand side of the platform, or an individual or group that is independent of other market players (Stahl et al., 2017, 2016). An example of a data marketplace with a private sponsor is Informatica B2B Exchange. Informatica is a software development company, with its proprietary software platform as main resource. In 2015, the company was acquired for $5.3B by Permira, a European private equity firm (Permira, 2015). The HERE Marketplace is an example of a data marketplace that is sponsored by a consortium of data buyers. HERE is a provider of location data and platform provider, with a service offering comprising of a development workspace, data marketplace, and map creation and visualization tools. The technology company is invested in by some main shareholders in the automotive industry, such as Audi, BMW and Daimler, and other engineering and service suppliers such as Bosch, Continental, Intel and Pioneer. Last year, HERE welcomed Mitsubishi as a major new shareholder, that took 30% ownership of the technology company (HERE, 2019). An example of an independent platform sponsor is oneTRANSPORT. The company aims to function as a neutral infrastructure provider in the UK data market.

4.4. Finance domain

In the finance domain of the STOF model, the financial arrangements between the different participants in the value network are specified (Bouwman et al., 2008). The value activities and technological architecture in the organization and technology domain are costs sources from the supply side of the service that affect the financial domain. Viable business models contain a balance between financial risks and benefits for the stakeholders involved in the value network. The finance domain therefore provides a description about how the network of actors intends to capture value. The *revenue model* depicts whether financial revenue comes directly from the buyers, or whether there are also other sources of revenue for the value network. The final price for the data good or service is specified by the *pricing model* of the data marketplace.

The *price discovery* function of a data marketplace describes how and by who the prices of the goods and services on the data marketplace are set. To provide safe payment, data marketplace may offer *smart contracts* that are enabled by blockchain. This may enhance privacy and trust among marketplace participants. Lastly, cryptocurrencies are emerging as an alternative *payment currency* to fiat money, as a way for marketplace providers to securely handle payments between data buyers and data sellers.

Financial revenue may come directly from the buyer of the good or service, but there are also other main sources of revenue for an enterprise (Bouwman et al., 2008). Five **revenue models** for data marketplaces are distinguished: the commission model, subscription model, usage fee model and asset sales model. In the commission or transaction fee model, the data marketplace receives a certain fee for every transaction that takes place on the platform (Spiekermann, 2019; Täuscher & Laudien, 2018; Täuscher, 2016). In the subscription model, the data marketplace signs a contract with platform users to provide a specific service for a recurring fee (Täuscher & Laudien, 2018; Täuscher, 2016). In the service sales model, the data marketplace sells services that are not standardly offered to all users (Täuscher & Laudien, 2018; Täuscher, 2016). Data marketplaces may charge are fee for the usage of their platform or services. In the asset sales revenue model (Osterwalder & Pigneur, 2010), the main source of revenue comes from the sales of data goods. For instance, the revenue of the location data marketplace Red Lion Data depends on the sales of their proprietary packages of data lists.

The **pricing model** specifies how the final price for the data good or service is composed. From the empirical analysis, it was found that data marketplaces employ seven types of pricing models: freemium, pay-per-use, flat fee tariff, package based pricing, and a combination of multiple of the aforementioned pricing models. In the freemium model the data marketplace provides basic functions for free, but marketplace users will need to pay a fee to make use of the premium functions (Fruhwirth et al., 2020; Spiekermann, 2019; Täuscher & Laudien, 2018; Täuscher, 2016). In pay-per-use or usage based pricing, customers pay a price that is proportional to the amount of units consumed by the data marketplace user (Fruhwirth et al., 2020; Spiekermann, 2019). The flat fee tariff or flat rate pricing model provides marketplace participants full access to the marketplace for a recurring fee (Fruhwirth et al., 2020; Schomm et al., 2013). In the package based pricing model, data goods or services are bundled in certain packages, of which the price may decrease by a certain discount rate when the size of the package increases (Fruhwirth et al., 2020; Spiekermann, 2019; Schomm et al., 2013). On some data marketplaces, the pricing of the data products and services are based on multiple pricing models.

A **price discovery** function allows buyers and sellers on the marketplace to determine a transaction price which they both agree on (Bakos, 1998). Data marketplaces make use of price discovery mechanisms to determine the price of a dataset before it is transacted on the platform: prices set by data buyers, discovery by negotiation, prices set by the marketplace provider and prices set by external sellers. The data marketplace may decide to let data buyers set the prices for the datasets they wish to buy. The harvesting data marketplace BIGToken aims to gather user data and information with the goal to selling it to advertisers. In this case, the data marketplace functions as buyer. BIGToken asks users to participate in brand and product surveys in return for rewards in the form of points, that can be exchanged for PayPal money or gift cards. The rates at which the earned points are exchanged is specified by the data marketplace. In the negotiation model, data marketplaces may allow data buyers and sellers to negotiate about the price before coming to an agreement. For example, Datum Data Marketplace allows data buyers to send a data purchase request to users to buy a copy of their encrypted personal data. A purchase request comprises of details about the purchaser

and the proposed price set by the data buyer. Users can agree to the proposed purchase price or send a counter offer to the data buyer. This way, the negotiation process about the data price takes place. The data marketplace provider may also decide to take charge of setting prices for the data goods and services on the platform. Lastly, the data marketplace may allow external sellers to set the prices for their own data offering on the marketplace. In this case, the data providers are free to set their own product description and prices.

Data marketplaces may implement **smart contracts** to enhance transparency and to enforce trust among marketplace participants (Fruhwirth et al., 2020). A smart contract comprises of an contractual agreement that is coded into a script that is automatically executed when the terms in the contract are met. The use of smart contracts by data marketplaces is emerging as a way to introduce transparency and to automatically handle payments made on the marketplace (Lawrenz et al., 2019). DAWEX is an example of a company that has implemented a smart contract in the Ethereum blockchain for the exchange of data on its data marketplace.

The **payment currency** dimension explicates which currencies are accepted by the data marketplace for the payments that are made by data buyers on the platform (Fruhwirth et al., 2020). Data marketplaces may handle their payments via cryptocurrencies or fiat money. Data marketplace companies that use cryptocurrencies as a payment method are emerging. Examples of marketplaces that offer payment in cryptocurrency are IOTA and Streamr, that both have developed their own coin, the IOTA and DATA.

Table 3: Taxonomy of data marketplace business models

	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 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	

5. Demonstration

The use of the taxonomy is demonstrated by applying the derived business model dimensions and characteristics to three empirical examples of data marketplaces. Demonstration on the basis of empirical illustration is employed in a number of other taxonomy development studies (Azkan et al., 2020; Bock & Wiener, 2017). In this study, three mini case studies are conducted to showcase how to use the taxonomy for the classification of business models of data marketplace companies. The three selected cases are part of the database of data marketplaces that was established in this study, but they were not included in the taxonomy development process. Therefore, the demonstration proves that the taxonomy can be used for data marketplaces outside of the sample of cases. The demonstration of the taxonomy can help researchers and practitioners to understand how to use the taxonomy for the classification and design of data marketplace business models.

Three data marketplace companies are selected from the sample of empirical cases to demonstrate the usefulness of the taxonomy. The three cases are part of the database of data marketplace companies that was established during the taxonomy development process, but they were not part of the sample of 40 cases. As the two existing taxonomies of data marketplaces mainly focussed on the classification of multilateral data marketplaces (Fruhwirth et al., 2020; Spiekermann, 2019), this study aims to go beyond the state of the art by developing a taxonomy that is also suitable for classifying other types of data marketplaces, based on their matching models and marketplace design. Therefore, the three selected companies differ in terms of their marketplace design: a bilateral data marketplace, harvest data marketplace and multilateral data marketplace (Koutroumpis et al., 2017). The first selected company is Wibson, a bilateral marketplace (one-to-one matching) in the personal data industry. QueXopa, an alternative data marketplace focussed on the Latin American market that has that has adopted a dispersal marketplace design (one-to-many matching), is the second selected company for the empirical illustration of the taxonomy. The third selected company is Advaneo, a multilateral data marketplace (many-to-many matching). Information

on the business models of the selected data marketplaces was derived from desk research that was conducted to gather information on the empirical cases. Main sources of information were the company websites of the selected cases, white papers, terms and conditions and news articles about the data marketplace companies. For the company Wibson, a scientific whitepaper by Fernandez et al. (2020) was available. Table 4 presents a summary of the business model characteristics of the selected cases.

Wibson is an example of a bilateral data marketplace that enables personal data trading between individuals and organizations. The company offers a decentralized data marketplace that makes use of smart contracts to enable inviduals to securely and anonymously share data in a trusted environment (Fernandez et al., 2020). Wibson is active in the personal data industry domain. The company provides an infrastructure for individuals to share information with data buyers. The company has implemented smart contracts to arrange the secure selling of data between data buyers and data sellers (Fernandez et al., 2020). Hence, data is provided by customers on the data marketplace. The individuals are in control of their personal data, and are able to monetize their personal data by giving organizations data access in return for money. This way, Wibson adopts a one-to-one matching mechanism, where the company itself functions as neutral intermediary that provides the blockchain infrastructure for the data exchange. Prices on the marketplace are set by data buyers, and buyers are matched with individual data sellers that are willing to sell their personal data for the set price. Wibson has implemented smart contracts to handle payments, and transactions are paid in cryptocurrencies, that can be exchanged with Wibson in return for fiat money (Fernandez et al., 2020).

QueXopa presents itself as the single source for Latin American alternative data. The company aims to set itself apart from the competition by finding, sourcing and aggregating alternative data, to provide exclusive, high quality and accurate alternative data. Customers of QueXopa are finance professionals such as investors, hedge funds, market analysts, retailers and corporations. The company is active in the alternative data industry domain, and provides credit card transactions, mobile location data, insurance policies, real estate listings, mobile app metrics, price monitoring, email receipts and maritime and port data from Latin American sources. This type of data is mostly static, and provides a snapshot of the moment of measurement of the data. On the one hand QueXopa sources data from major Latin American data providers, and on the other hand the company generates data themselves by conducting equity research and data scraping from websites. QueXopa makes use of the one-to-many matching mechanism, as it aims to sell its proprietary high quality alternative data to a multitude of governmental and industrial finance professionals. The company offers alternative data via its website, but also provides on custom tailored on demand data on request. Customers can subscribe to the data offerings of the company, and pricing is dependent on the frequency, history, quantities and regions of the data. The prices of the data are set by QueXopa itself, and the company handles payment in fiat money.

Advaneo is a data marketplace that aims to provide easy data access and tooling. The company offers a data science workbench on top of their data offering, that consists of a Jupyter Notebook operated via the Advaneo cloud. Furthermore, Advaneo offers an enterprise data marketplace solution in the form of Closed User Groups. This function of the data marketplace allows marketplace participants to control who has access to the proprietary datasets of participants, and allows organizations to invite both internal as well as external users to participate in projects. The Advaneo marketplace is open for any individual or organization to join, and offers four different member ship models: free, premium, small business and enterprise. The company aims to foster cross-domain innovations, and therefore any type of data is traded on the platform. The website is available in 16 languages, which

shows that Advaneo is open for global reach. The marketplace offers open data as well as commercial data, and while the number of datasets and portal on the platform is increasing, Advaneo conducts regular updates and maintenance of the datasets offered on the platform. Advaneo has adopted a decentralized architecture design, in which the data traded on the marketplace is transferred directly from the data seller to the data buyer through a secured line, without touching the platform. This way, Advaneo aims to take on a neutral intermediary position in the data market. The company offers multiple forms of access to the data, such as acces via API and access via specialized software (IDS connector). The main source of revenue for Advaneo are subscriptions, that vary depending on the type of membership of the marketplace. The company also offers a freemium option, that allows users to test data, build basic visualizations, trial the IDS-connector and use the workbench for up to 15 GB data. Prices for the data offerings on the marketplace are set by external data sellers, paired with a data license agreement. The marketplace also includes open data that is offered for free. Advaneo offers the possibility to pay in fiat money by credit card. Other digital options such as Apple pay are currently being built.

Table 4 provides an overview of the business model characteristics that were identified for the selected cases with the use of the taxonomy. If the information sources did not provide sufficient information about the business model characteristics of a selected case, this is denoted by *no info* in the table.

Table 4: Illustration of the use of the taxonomy by application to three empirical cases

		Company		
		Wibson	QueXopa	Advaneo
Service domain	Value proposition	Secure data sharing	High quality and unique data	Easy data access and tooling
	Enterprise data marketplace	No	No	Yes
	Data processing and analytics tools	No	No	Yes
	Marketplace participants	C2B	B2B	Any
	Industry domain	Health & personal data	Alternative data	Any data
	Geographic scope	*No info*	Regional	Global
	Time frame	Static data	Static data	Up-to-date
Technology domain	Platform architecture	Decentralized	*No info*	Decentralized
	Data access	Specialized software	*No info*	Multiple
	Data source	Customer provided	Multiple	Multiple
Organization domain	Matching mechanism	One-to-one	One-to-many	Many-to-many
	Platform sponsor	Independent	Private	Independent
Finance domain	Revenue model	*No info*	Subscriptions	Subscriptions
	Pricing model	Pay-per-use	Pay-per-use	Freemium
	Price discovery	Set by buyers	Set by marketplace provider	Set by external sellers

	Smart contract	Yes	No	No
	Payment currency	Crypto	Fiat	Fiat

Most of the business model characteristics of the three selected data marketplaces could be classified with the use of the taxonomy (see Table 4). The cases were classified based on publicly available information on the websites of the companies, white papers, news articles and academic studies. If the information sources did not provide sufficient information about the business model characteristics of a selected case, this is denoted by *no info* in the table. A key take away from the demonstration is the taxonomy is useful to classify the business models of data marketplace when sufficient information is available about the respective business model characteristics of the companies.

6. Discussion

In this study, a taxonomy of data marketplace business models was developed by employing a a design science approach (Hevner, 2007) and following a standard taxonomy development method (Nickerson et al., 2013). The final taxonomy comprises of 4 meta-dimensions, 17 business model dimensions and 59 business model characteristics.

6.1. Scientific contributions

The developed taxonomy of data marketplace business model contributes to science in a number of ways. First, the results of the study contribute to understanding the notion of data marketplace business models by developing a taxonomy that describes the most important dimensions and characteristics of data marketplace business models. This way, the research contributes to the scarce knowledge about data marketplaces and their respective business models (Thomas & Leiponen, 2016). This study goes beyond the state of the art by adopting a multi-stakeholder perspective on data marketplace business models, by emphasizing the roles of players in the data marketplace platform ecosystem. For academic researchers, the taxonomy can function as a knowledge map that displays the contemporary knowledge from both scientific research and practical applications. It may expose certain areas that require further research. The developed taxonomy provides academic researchers with an overview of the characteristics of data marketplace business models. The taxonomy may expose certain areas of research where new business model alterations are emerging, that were derived from empirical cases and that may have not been thoroughly research in science yet.

A second contribution made by this study is related to the interpretation of a data marketplace. Compared to existing data marketplace business model taxonomies, this research takes on a broad perspective and interpretation of data marketplaces. Existing data marketplace business model taxonomies focus on studying one variant of data marketplaces, multilateral data marketplaces, where the marketplace functions as neutral intermediary that connects data buyers and data sellers (Fruhwirth et al., 2020; Spiekermann, 2019). In this study, data marketplaces are more broadly interpreted as the digital infrastructure on which the commercial trading of data as a valuable good takes place. By providing a sound definition of a data marketplace, this study aims to contribute to the knowledge and understanding of data marketplaces.

6.2. Limitations

The research process and results of this study are subject to a number of limitations. Firstly, the interpretation of qualitative information about the business models of empirical cases is prone to the subjectivity of the researcher. In this study, desk research was conducted to gather

information about the business models of existing data marketplaces by consulting company websites, online news articles and other sources, and within case analysis was conducted on the gathered information. As this research was conducted by a single researcher, there is a chance that valuable information may have been missed or that the gathered information was misinterpreted. The interpretation of the found information is subject to researchers' knowledge about the object of interest. Therefore, other researchers may find more information or interpret the information differently, which may result in the finding of different business model dimensions characteristics. However, as multiple design iterations were conducted in this research to develop the taxonomy, desk research for relevant information was performed multiple times and the discovered information was analysed in multiple iterations.

Secondly, the database of existing data marketplaces was constituted by consulting data marketplaces included in the repository of database.ai and existing scientific studies, and was complemented with desk research for empirical cases in Google. This search process resulted in the discovery of 178 data marketplaces. It may be that some existing data marketplaces have been missed during the desk research, or that new data marketplaces have been set up during the writing of this thesis. Therefore, future research on data marketplace business models may discover additional and new empirical cases that will improve the understanding of the object of interest.

Thirdly, not all data marketplace companies provided sufficient information about all of their business model characteristics. Therefore, not all empirical cases could be classified on the conceptually derived dimensions. For instance, the business model dimensions *main revenue partner* and *key costs* (Täuscher & Laudien, 2018; Täuscher, 2016) may be relevant for the classification of data marketplace business models, but they were omitted during the taxonomy development process because no sufficient information was found to make well-founded statements about these dimensions. In this research, there has not been any direct contact with providers of data marketplaces to verify the information about their respective business models.

6.3 Recommendations for further research

A number of possible opportunities and directions for further research arise from the results and limitations of this research. Firstly, future researchers can utilize the designed taxonomy to derive business model patterns and archetypes in existing data marketplaces. Business model patterns are business model characteristics that are commonly used in practical cases (Remane et al., 2017; Abdelkafi et al., 2013). Subsequently, business model archetypes are configurations of business model characteristics that are common among existing companies (Fruhwirth et al., 2019; Hodapp et al., 2019; Weking et al., 2019). The statistical derivation of business model patterns and archetypes may be conducted existing data marketplaces in a specific industry domain, such as the audience data industry, or by comparing business model patterns across different industries. Using the developed taxonomy for the statistical analysis of patterns and archetypes may uncover the frequency of certain business model characteristics and business model configurations, which can provide valuable information about the relative importance of certain business model dimensions and characteristics.

Secondly, the different alterations of data marketplaces that occur in practice may be studied, taking the variety of terms used by data marketplace providers in practice in consideration. During the desk research, it was found that the terms data marketplace, exchange and data

platform are often interchanged. The same was true for the use of the terms data, information and insights. To allow the inclusion of atypical examples of data marketplaces, the term data marketplace was broadly interpreted in this study. However, future research may focus on providing a concise definition of a data marketplace, and take into consideration the various terms used in practice. This way, a more clear differentiation can be made between different types of data marketplaces.

Thirdly, in-depth case studies may be conducted on specific data marketplaces or in certain industry domains, by conducting interviews, surveys and other qualitative empirical analyses. The in-depth analyses of existing data marketplace companies may provide deeper insight about certain business model dimensions and characteristics that were not highlighted in this research, such as the key costs and main revenue partners of data marketplaces. From a strategy perspective on business models, such in-depth case studies may uncover the considerations that are made by data marketplace providers with regard to choosing between certain business model characteristics, and provide insight about which business model characteristics provide certain data marketplace companies with a competitive advantage over other companies. Through interviews with relevant experts in the field of data marketplaces, the taxonomy may also be validated to see whether the derived business model dimensions and characteristics are relevant, whether the taxonomy should be revised, or if new concepts can be added to the taxonomy on the basis of expert knowledge.

6.4. Managerial and societal relevance

The developed taxonomy is relevant to managers and society, as it provides knowledge, information and transparency about the business models of data marketplaces. The taxonomy that was developed in this study can be used by managers and other decision-makers who are exploring the options of setting up a data marketplace or that are considering to join an existing data marketplace, to provide guidance in making business model design choices. An improved understanding about data marketplace business models may result in an increase of data marketplaces, that will make data more accessible and exploitable to a wide range of stakeholders, including individuals, businesses and authorities. Furthermore, a number of existing data marketplaces that enable the transactions of audience, health and personal data was exposed during this research, that may be directly gather data from individuals or communities in society. Therefore, this study may promote the dialogue about the existence of companies that gather personal data of consumers with the goal of exchanging it for commercial purposes. Moreover, the results of this study may raise awareness about the emergence of harvesting data marketplaces, that enable consumers to monetize their personal and health data.

References

Abdelkafi, N., Makhotin, S., & Posselt, T. (2013). Business model innovations for electric mobility—what can be learned from existing business model patterns?. *International Journal of Innovation Management*, *17*(01), 1340003.

Al-Debei, M. M., El-Haddadeh, R., & Avison, D. (2008). Defining the business model in the new world of digital business. *AMCIS 2008 proceedings*, 300.

Amit, R., & Zott, C. (2001). Value creation in e-business. *Strategic management journal*, *22*(6-7), 493-520.

Azkan, C., Iggena, L., Gür, I., Möller, F., & Otto, B. (2020). A Taxonomy for Data-Driven Services in Manufacturing Industries. In *PACIS* (p. 184).

Bakos, Y. (1998). The emerging role of electronic marketplaces on the Internet. *Communications of the ACM*, *41*(8), 35-42.

Casadesus-Masanell, R., & Ricart, J. E. (2010). From strategy to business models and onto tactics. *Long range planning*, *43*(2-3), 195-215.

Chesborough, H., & Rosenbloom, R. S. (2002). The role of the business model in capturing value from innovation: evidence from Xerox Corporation's technology spin-off companies. *Industrial and corporate change*, *11*(3), 529-555.

Baldwin, C. Y., & Woodard, C. J. (2009). The architecture of platforms: A unified view. *Platforms, markets and innovation*, *32*.

Bock, M., & Wiener, M. (2017). Towards a Taxonomy of Digital Business Models-Conceptual Dimensions and Empirical Illustrations. In *ICIS*.

Boudreau, K. J., & Hagiu, A. (2009). Platform rules: Multi-sided platforms as regulators. *Platforms, markets and innovation*, *1*, 163-191.

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

Bouwman, H., Nikou, S., & de Reuver, M. (2019). Digitalization, business models, and SMEs: How do business model innovation practices improve performance of digitalizing SMEs?. *Telecommunications Policy*, *43*(9), 101828.

Carnelley, P., Schwenk, H., Cattaneo, G., Micheletti, G., & Osimo, D. (2016). Europe's Data Marketplaces-Current Status and Future Perspectives. *European Data Market SMART 2013/0063 D*, *39*.

Chakrabarti, A., Quix, C., Geisler, S., Pullmann, J., Khromov, A., & Jarke, M. (2018). Goal-Oriented Modelling of Relations and Dependencies in Data Marketplaces. In *iSTAR@ CAiSE*.

Daniel, J. (2011). *Sampling essentials: Practical guidelines for making sampling choices*. Sage Publications.

Dodds, L. (2012). Kasabi – Shutting Down Kasabi. Retrieved March May 12th, from https://kasabi.wordpress.com/2012/07/09/shutting-down-kasabi/

DOMO. (2019). Domo Resource - Data Never Sleeps 7.0. Retrieved March 12th, 2020, from https://www.domo.com/learn/data-never-sleeps-7

European Commission. (2017). Communication on Building a European Data Economy. Retrieved March 12, 2020, from https://ec.europa.eu/digital-single-market/en/news/communication-building-european-data-economy

Faber, E., Ballon, P., Bouwman, H., Haaker, T., Rietkerk, O., & Steen, M. (2003). Designing business models for mobile ICT services. In *Workshop on concepts, metrics & visualization, at the 16th Bled electronic commerce conference etransformation, Bled, Slovenia*.

Fernandez, D., Futoransky, A., Ajzenman, G., Travizano, M., & Sarraute, C. (2020). Wibson protocol for secure data exchange and batch payments. *arXiv preprint arXiv:2001.08832*.

Fricker, S. A., & Maksimov, Y. V. (2017). Pricing of data products in data marketplaces. In *International Conference of Software Business* (pp. 49-66). Springer, Cham.

Fruhwirth, M., Rachinger, M., & Prlja, E. (2020, January). Discovering Business Models of Data Marketplaces. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*.

Gassmann, O., Frankenberger, K., & Csik, M. (2014). *The business model navigator: 55 models that will revolutionise your business*. Pearson UK.

Gordijn, J., & Akkermans, J. M. (2003). Value-based requirements engineering: exploring innovative e-commerce ideas. *Requirements engineering*, *8*(2), 114-134.

Hartmann, P. M., Zaki, M., Feldmann, N., & Neely, A. (2014). Big data for big business? A taxonomy of data-driven business models used by start-up firms. *A taxonomy of data-driven business models used by start-up firms*.

Henderson, R. M., & Clark, K. B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative science quarterly*, 9-30.

HERE. (2019). HERE Technologies welcomes Mitsubishi Corporation and NTT as new investors. Retrieved 9 July 2020, from https://www.here.com/company/press-releases/en/2019-20-12

Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian journal of information systems*, *19*(2), 4.

Hodapp, D., Remane, G., Hanelt, A., & Kolbe, L. M. (2019). Business Models for Internet of Things Platforms: Empirical Development of a Taxonomy and Archetypes.

Hürtgen, H., & Mohr, N. (2018). Achieving business impact with data. Retrieved March 12th, 2020, from https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/achieving-business-impact-with-data

Johnson, B. (2012). Kasabi shuts down, says data marketplace 'too slow'. Retrieved from https://gigaom.com/2012/07/09/kasabi-shuts-down-says-data-marketplace-too-slow/

Katz, M. L., & Shapiro, C. (1985). Network externalities, competition, and compatibility. *The American economic review*, *75*(3), 424-440.

Kosara, R. (2010). The rise and fall of swivel. com. Retrieved May 12th, from https://eagereyes.org/criticism/the-rise-and-fall-of-swivel

Koutroumpis, P., Leiponen, A., & Thomas, L. D. (2017). *The (unfulfilled) potential of data marketplaces* (No. 53). ETLA Working Papers.

Koutroumpis, P., Leiponen, A., & Thomas, L. D. (2020). Markets for data. *Industrial and Corporate Change*, *29*(3), 645-660.

Lambert, S. (2015). The importance of classification to business model research. *Journal of Business Models*, *3*(1).

Muschalle, A., Stahl, F., Löser, A., & Vossen, G. (2012). Pricing approaches for data markets. In *International workshop on business intelligence for the real-time enterprise* (pp. 129-144). Springer, Berlin, Heidelberg.

Nickerson, R. C., Varshney, U., & Muntermann, J. (2013). A method for taxonomy development and its application in information systems. *European Journal of Information Systems, 22*(3), 336-359.

Opher, A., Chou, A., Onda, A., & Sounderrajan, K. (2016). The rise of the data economy: driving value through internet of things data monetization. IBM Corporation: Somers, NY, USA.

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

Permira. (2015). Informatica Announces Agreement to be Acquired by The Permira Funds and Canada Pension Plan Investment Board | Permira. Retrieved 9 July 2020, from https://www.permira.com/news-views/news/informatica-announces-agreement-to-be-acquired-by-the-permira-funds-and-canada-pension-plan-investment-board/

Prlja, E. (2019). *Discovering Business Models of Data Marketplaces* (Unpublished master's thesis). Graz University of Technology, Graz, Austria.

Ramel, D. (2016). "Microsoft Closing Azure DataMarket." *Application Development Trends MAG.* Retrieved May 12th, from https://adtmag.com/articles/2016/11/18/azure-datamarket-shutdown.aspx

Remane, G., Nickerson, R., Hanelt, A., Tesch, J. F., & Kolbe, L. M. (2016). A taxonomy of carsharing business models.

Remane, G., Hanelt, A., Tesch, J. F., & Kolbe, L. M. (2017). The business model pattern database—a tool for systematic business model innovation. *International Journal of Innovation Management*, *21*(01), 1750004.

Rochet, J. C., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the european economic association*, *1*(4), 990-1029.

Schomm, F., Stahl, F., & Vossen, G. (2013). Marketplaces for data: an initial survey. *ACM SIGMOD Record*, *42*(1), 15-26.

Spiekermann, M., Tebernum, D., Wenzel, S., & Otto, B. (2018). A metadata model for data goods. *MKWI 2018-Multikonferenz Wirtschaftsinformatik, 2018–March*, 326-337.

Spiekermann, M. (2019). Data Marketplaces: Trends and Monetisation of Data Goods. *Intereconomics*, *54*(4), 208-216.

Stahl, F., Schomm, F., & Vossen, G. (2014a). Data Marketplaces: An Emerging Species. In *DB&IS* (pp. 145-158).

Stahl, F., Schomm, F., & Vossen, G. (2014b). The data marketplace survey revisited (No. 18). Working Papers, ERCIS-European Research Center for Information Systems.

Stahl, F., Schomm, F., Vossen, G., & Vomfell, L. (2016). A classification framework for data marketplaces. *Vietnam Journal of Computer Science*, *3*(3), 137-143.

Stahl, F., Schomm, F., Vomfell, L., & Vossen, G. (2017). Marketplaces for Digital Data: Quo Vadis?. *Computer and Information Science*, *10*(4).

Täuscher, K. (2016). Business Models in the Digital Economy: An Empirical Classification of Digital Marketplaces. *Fraunhofer Center For International Management and Knowledge Economy.*

Täuscher, K., & Laudien, S. M. (2018). Understanding platform business models: A mixed methods study of marketplaces. *European Management Journal*, *36*(3), 319-329.

Teece, D. J. (2010). Business models, business strategy and innovation. *Long range planning*, *43*(2-3), 172-194.

Thomas, L. D., & Leiponen, A. (2016). Big data commercialization. *IEEE Engineering Management Review, 44*(2), 74-90.

Timmers, P. (1998). Business models for electronic markets. *Electronic markets*, *8*(2), 3-8.

Tiwana, A., Konsynski, B., & Bush, A. A. (2010). Research commentary—Platform evolution: Coevolution of platform architecture, governance, and environmental dynamics. *Information systems research*, *21*(4), 675-687.

Van Alstyne, M. W., Parker, G. G., & Choudary, S. P. (2016). Pipelines, platforms, and the new rules of strategy. *Harvard business review*, *94*(4), 54-62.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly*, xiii-xxiii.

Weking, J., Mandalenakis, M., Hein, A., Hermes, S., Böhm, M., & Krcmar, H. (2019). The impact of blockchain technology on business models–a taxonomy and archetypal patterns. *Electronic Markets*, 1-21.