Emina Prlja, BSc

Discovering Business Models of Data Marketplaces

MASTER'S THESIS

to achieve the university degree of

Diplom-Ingenieurin

Master's degree programme: Software Engineering and Management

submitted to

Graz University of Technology

Supervisor

Univ.-Prof. Dr. Stefan Vorbach

Institute for General Management and Organization

Adviser: Univ.-Ass. DDipl.-Ing. Michael Rachinger
Adviser: Dipl.-Ing. Michael Fruhwirth MLBT, Know-Center GmbH

Graz, September 2019

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master‘s thesis.

_____	_____
Date	Signature

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Meaning
API	Application Programming Interface
XML	eXtensible Markup Language
CSV	Comma Separated Values
XLS	eXceL Spreadsheet
JSON	JavaScript Object Notation
RDF	Resource Description Framework
PDF	Portable Document Format
DOC	DOCument (file format/ extension)
JPEG	Joint Photographic Experts Group (image file format/extension)
GDPR	General Data Protection Regulation
B2B	Business-to-Business
B2C	Business-to-Customer
C2B	Customer-to-Business
C2C	Customer-to-Customer

ABSTRACT

The modern economy heavily relies on data as a resource for advancement and growth. A huge amount of data is produced continuously, and only a fragment of the amount is handled properly and efficiently. Data marketplaces are increasingly gaining attention. They provide possibilities to exchange, trade and access different kinds of datasets across organizations, between interested data providers and data buyers. Data marketplaces need stable and efficient infrastructure for their operations, and a suitable business model in order to provide and gain value. Due to the rapid development of the field, and its recent high increase in popularity, the research on business models of data marketplaces is fragmented. This thesis aims to address the issue by identifying dimensions and characteristics of data marketplaces, which outline the characteristics of their business models. Following a rigorous process for taxonomy building, a business model taxonomy for data marketplaces is proposed. Using the evidence from a final sample of twenty available data marketplaces, the frequency of characteristics of data marketplaces is analyzed. In addition, four data marketplace business model archetypes are identified. The findings reveal the impact of the structure of data marketplaces as well as the relevance of infrastructure, regulations and security issues handling for identified business model archetypes. Therefore, this study contributes to the growing body of literature on digital business strategies.

1 INTRODUCTION

This chapter will provide an introduction to the topic being researched throughout this thesis, as well as the research question of this study. Furthermore, the conceptual structure of the thesis will be presented, followed by a description of how the chapters of this thesis are organized.

1.1 INITIAL SITUATION AND RESEARCH QUESTION

The amount of available data has increased exponentially in recent years. Combined with new possibilities in the field of data analysis, machine learning and storage technologies, data can be increasingly leveraged for economic benefits. Thus, data utilization and analytics have a high potential for transforming a firm's business models (Agarwal *et al.*, 2019; Günther *et al.*, 2017; Loebbecke and Picot, 2015; Woerner and Wixom, 2015). While the monetization of data, in the form of datasets, is an essential focus of organizations and even industries today, there is a lack of market mechanisms to exchange and price such datasets and match buyers to sellers (Agarwal *et al.*, 2019). Data marketplaces are one approach to overcome the challenges of data monetization (Agarwal *et al.*, 2019; Ozyilmaz *et al.*, 2018). Data marketplaces offer to companies the opportunities to use external data for the improvement of their business, but also to increase revenue by reselling internal data collected through their business (Carnelley *et al.*, 2016.). Data is triggering improvements of products and services, that consequently produce new data which can be a product itself (Carnelley *et al.*, 2016; Spiekermann *et al.*, 2018). In order to be sustainable and efficient, data marketplaces need an appropriate business model. However, in the emerging data exchange environment, the definition of appropriate business models and their continuous evaluation is a rather challenging task (Chakrabarti *et al.*, 2018). Consequently, companies currently underutilize most of the collected data (Deichmann *et al.*, 2016).

Overall, there has been little research conducted on the emerging field of data marketplaces in general (Thomas and Leiponen, 2016) and business models of data marketplaces in particular. In order to address this gap, this thesis serves the purpose of answering the following research question:

"What are the characteristic elements of data market business models?"

A well-accepted and a commonly used research approach in the area of business models are taxonomies, that classify objects of interest in the domain of interest and help to understand the complexity of the domain and its existing or possible concepts (Nickerson *et al.*, 2013). Research on business model taxonomies of digital business models is of interest for current Information Systems research (Beinke *et al.*, 2018; Bock and Wiener, 2017; Remane *et al.*, 2016). Therefore, a taxonomy of data marketplaces was developed in order to support answering the research question of this thesis.

For the development of the taxonomy, a commonly accepted method proposed by Nickerson *et al.* (2013) was used. A literature review was conducted in order to identify the relevant concepts and characteristics of data marketplaces. Moreover, twenty representative cases of data marketplaces were used to empirically revise the taxonomy. Thus, throughout this research sixteen key dimensions of data marketplaces were identified. The dimensions help to distinguish and explain business models of data marketplaces. Each dimension consists of a distinct set of characteristics. The resulting taxonomy was used to further identify four distinct business model archetypes for the representative sample of data marketplaces.

1.2 CONTRIBUTION

The taxonomy developed for this research contributes to the existing body of knowledge by establishing a common understanding of business models of data marketplaces. It should serve as one step forward to narrowing down the gap that exists in the research of the emerging topic on data trading ecosystem.

The emerging nature of data marketplaces indicates expected changes already in the near future. Therefore, the taxonomy developed throughout this research can be reused as a tool for monitoring the change of data marketplaces' business models. Furthermore, the taxonomy can be extended with new dimensions and characteristics as needed prior to the repetition of the analysis. Furthermore, the taxonomy can be adapted to fit similar researches as well. Additionally, the taxonomy outlines the possibilities of current data marketplaces' business models. Therefore, a data marketplace vendor whose business model needs modifications, or is still in the development phase, can use the taxonomy as a source of inspiration for new ideas. Similarly, new players who are eager to join the market can use the taxonomy as a guideline for business model development. The developed taxonomy is a basis for the future development of business models and innovations within the data market.

1.3 THE CONCEPTUAL STRUCTURE OF THE THESIS

The process of answering the research question of this thesis is conceptualized by separating it into two main structural parts – a theoretical part and a practical part. The visual overview of the concept that the thesis follows is presented in Figure 1 and described afterward.

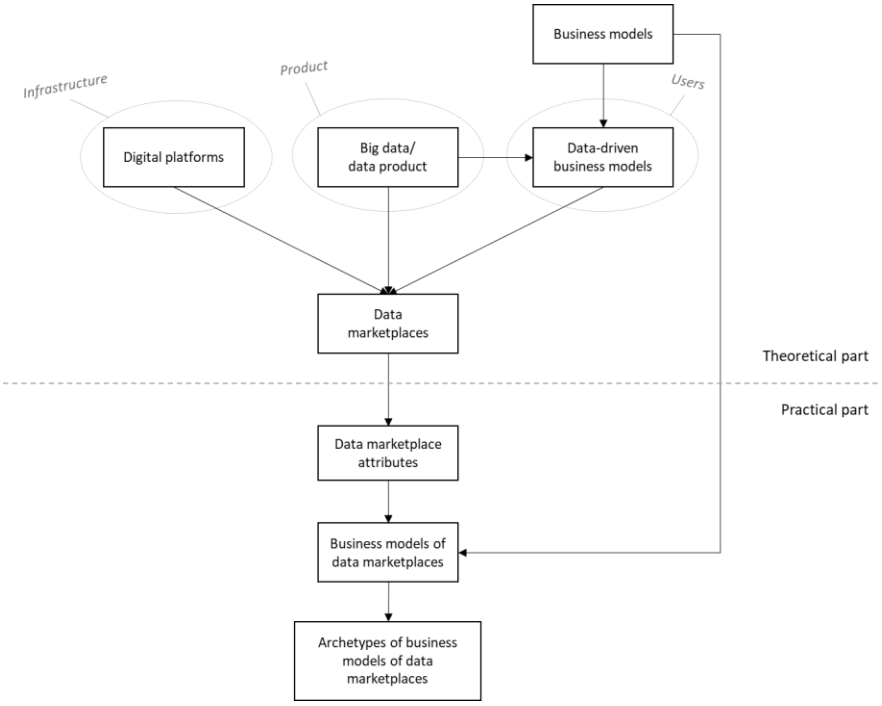

Figure 1 – Conceptual structure of the thesis (own illustration)

The first major part of the concept in Figure 1 is the theoretical framework. It consists of the background information gathered from the existing literature on business models, data-driven business models, platform and electronic marketplace business models as well as data marketplaces. The investigated literature was found in the following databases: Science Direct, IEEE Xplore, Google Scholar, Scopus, and ResearchGate. Identifying the exact definition of data marketplaces and understanding their characteristics was approached by taking a look at the broader picture, which involves three major concepts. The concepts subsequently bundle into the framework of data marketplaces ecosystem. The three concepts that are considered are:

- *The infrastructure* that the marketplace is built on – the underlying mechanism for a data marketplace is an electronic marketplace (also referred to as "digital platform")
- *The product* that is being traded or used within the data marketplace ecosystem - big data/data product
- *Users* who are the biggest demanders of useful and efficient data marketplaces – businesses that implement data-driven business models.

The interdependence of the involved concepts is presented in Figure 1, and it is not completely straightforward. Big data/data product is not only a concept that is involved in data marketplaces, but it is a key asset of data-driven business models as well. Moreover, understanding the definition of business models is necessary before defining data-driven business model, as well as for discovering characteristics of data marketplaces' business models, the step that belongs to the practical part of the thesis.

The second major part of the concept of this thesis is the practical part (see Figure 1). The input for the practical part was the fundamental understanding of all involved concepts and characteristics of data marketplaces, which was established in the theoretical part. The practical part involved the identification of characteristics of data marketplaces, that are relevant for describing their business models. Moreover, it involved assigning objects from the representative set of data marketplaces to these characteristics. This step was done in order to learn about business models of data marketplaces from practical examples. It supported the identification of different business models archetypes that characterize data marketplaces which are currently available on the market.

1.4 ORGANIZATION OF CHAPTERS

A visual overview of the organization of chapters throughout this thesis is provided in Figure 2. Further details about the consistency of each chapter are described in the following.

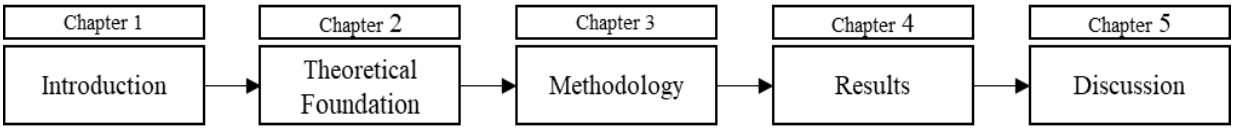

Figure 2 – Outline of the thesis chapters (own illustration)

Chapter 1 provides the introduction to the topic and research question of this thesis. Further, the conceptual structure of the thesis document is presented, followed by the description of the organization of chapters.

The theoretical part, containing literature overview, is presented in chapter 2. In subchapter 2.1 the concept of big data and data, in general, is explained. The subchapter gives an overview of the characteristics of data, describes the possibilities of creating value from data, and includes the challenges and opportunities with data when it is seen as a product. In subchapter 2.2, the concept of electronic platforms and marketplaces is explained, which is the introductory overview that discusses different types of platforms, and possibilities of innovations within them, as well as competition among the platforms. The introduction to electronic platforms is followed by the theoretical overview of the specific type of platforms – data platforms. The subchapter explains the main components and stakeholders of a data marketplace, possible and most common evolvement of new data marketplaces from setup phase to the mature phase, as well as different possible structures of data marketplaces. It continues with the existing definitions of data marketplaces, proposed by different researches on the topic, and bundles the definitions into the one which will be followed throughout this work. The subchapter further lists possible data marketplaces participants, the typical challenges current data marketplaces face, and the typical pricing mechanisms data marketplaces providers implement. Furthermore, data marketplaces are discussed from the economic perspective, and statistical facts and figures are presented (including those measured in 2016 and 2017, as well as those predicted and expected for 2020 and 2025). Last but not least, the characteristics of two typical infrastructures of data marketplaces are described. Finally, the theoretical part of subchapter 2.3 gives an overview of the business model as a concept in general. That overview combined with the understanding of possibilities of creating value from data (see subchapter 2.1), and typical data marketplaces stakeholders (see subchapter 2.2), gives a foundation for the understanding of data-driven business models concept. The concept is a special case of business models and the stakeholder with the biggest demand for data marketplaces.

Chapter 3 describes the research methodology, which explains in details all the steps taken in order to answer the research question of the thesis. It describes the steps of taxonomy development, including the methods of collecting and analyzing empirical data.

The results of the research are presented in chapter 4. It consists of individual elements of the developed taxonomy, a frequency analysis results, four archetypes of data marketplaces as well as four illustrative cases for each of the archetypes.

Finally, the discussion of the empirical results is presented in chapter 5. It involves interpretation and evaluation of the results and information about how they relate with the knowledge gathered from the existing literature reviewed in the theoretical part of the thesis. Furthermore, the discussion includes the limitations of this thesis, as well as the possible directions for future research about the topic.

2 THEORETICAL FOUNDATION

This chapter introduces the theoretical foundation needed for understanding the data market ecosystem, and the concepts that the ecosystem involves. It starts with the description of big data and the possibility for extraction of additional value from it. Furthermore, the theoretical overview of data marketplaces is provided, including the explanation of their underlying infrastructure and the challenges they face. Last but not least, the chapter tackles the concept of data-driven business models, including the introductory overview of business models in general.

2.1 BIG DATA AND DATA VALUE

The concept of big data is explained in this subchapter. The main characteristics of big data are described, as well as how it can be used as an asset for creating additional value out of it. Furthermore, the characteristics of data as a product or a service are tackled, including the metadata model of the data product.

By 2020 every second, every person on earth will produce 1.7 MB of data. Every minute, 159,362,760 emails are being sent, 4,333,560 YouTube videos are being viewed, 1,111 Amazon packages are being delivered, 12,054 gallons of wine are consumed, 49,380 videos are uploaded to Instagram (Domo, Inc., 2018). Every two years, the amount of data doubles, and it is estimated that 44 zettabytes (44 trillion gigabytes) of data will be created and copied every year by 2020 (Turner *et al.*, 2014). As a comparison, the amount of data in 2010 was estimated to 1,200 exabytes (1.17 zettabytes) (Gantz and Reinsel, 2013). Nowadays, data is considered to be the new oil (Haupt, 2016).

Due to the growth of data that is created, stored and manipulated, it is often labeled with the term "big data", but the term is not uniformly defined and depends on the perspective of individuals. Big data is often described as "a huge amount of data", but the volume is not its only characteristic. Big data is diverse, continuously produced, complex, difficult or impossible to manipulate by traditional processes or tools, and potentially powerful – if properly utilized. However, with the growth of big data, its utilization is becoming more and more challenging. Big data is the data that can be captured, communicated, aggregated, stored, analyzed, and visualized, which is exactly what makes it challenging to handle. Despite the challenges, it became the ultimate basis for growth and advancement of the modern economy. (Philip Chen and Zhang, 2014; Vega-gorgojo *et al.*, 2016; Schüritz *et al.*, 2017; Liang *et al.*, 2018; Manyika *et al.*, 2011; Zikopoulos *et al.*, 2012)

In general, there are three main characteristics of big data that make it a valuable resource: volume, variety, and velocity (see Figure 3). This is a concept known as 3Vs, which is often used to describe big data (McAfee and Brynjolfsson, 2012; Liang *et al.*, 2018). The three characteristics are explained in more details in Table 1.

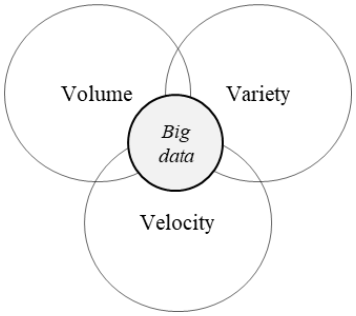

Figure 3 – 3Vs of big data (Liang et al., 2018, p. 15134)

Big data characteristic	Explanation
Volume	It describes the amount of data that exists, and that increases continuously. New data sources are being added continuously as well. The data can be saved in tables, records or files, and the amount of data is expressed in Bytes. The amount of data is way bigger than the amount that can be handled by a typical database software, and its precise amount globally can only be estimated, rather than precisely calculated.
Variety	This characteristic of big data describes that it is available from different kinds of sources, which makes data diverse in the context of content as well as format; it can be created either by human or by machines, and it can be structured, semi-structured or unstructured.
Velocity	Velocity is the characteristic that describes the speed of creation and accessing of big data. It is continuously generated and the huge set of available data is, therefore, being constantly altered, which forces companies to perform data analysis and make decisions in a rather agile manner. This constant data flow is the reason why traditional systems struggle to handle big data efficiently.

Table 1 – 3Vs of big data (own presentation based on Liang et al. (2018), McAfee and Brynjolfsson (2012) and Zikopoulos et al. (2012))

Data and big data will not be further distinguished from each other in this thesis. Both of the concepts will be referred to as "data", in order to reduce the complexity of the terminology, and since the big data is indeed data itself, only with special characteristics.

2.1.1 Creating value from data

Companies and organizations contribute to the creation and growth of data by collecting data about all individuals, products, processes, machines or sensors that they are in contact with. However, only collecting data principally does not create any value to a company, no matter how much of it they acquire. It is often just stored in a raw and unstructured or semi-structured format, while companies are struggling with filtering the right information from it. Furthermore, companies are often trying to decide if the whole collected data has any worth, when the collected data is not properly handled. However, if properly and efficiently utilized, it has the potential of increasing economic value and profit. For example, a retailer who uses the full potential of available data could increase its operating margin by at least 60% (Manyika *et al.*, 2011). Also, except the possible improvement of business growth, utilization of data can contribute to the improved development of health systems, science, education and public administration (Cattabei *et al.*, 2019). (Manyika *et al.*, 2011; Zikopoulos *et al.*, 2012)

Utilization of data, with the goal of creating valuable information from it, means using it as raw material and performing a set of operations on it. That set of operations is executed with specialized software and hardware, that can be seen as a plant and equipment for the production of information from data (see Figure 4). Consequently, the produced information is seen as a product, or an asset for creation of additional value. (Moody and Walsh, 1999)

Figure 4 illustrates the process of production of information that potentially has value, from a raw and unstructured data. Due to a high complexity and high amounts of data, the process is always supported by an Information system. (Moody and Walsh, 1999)

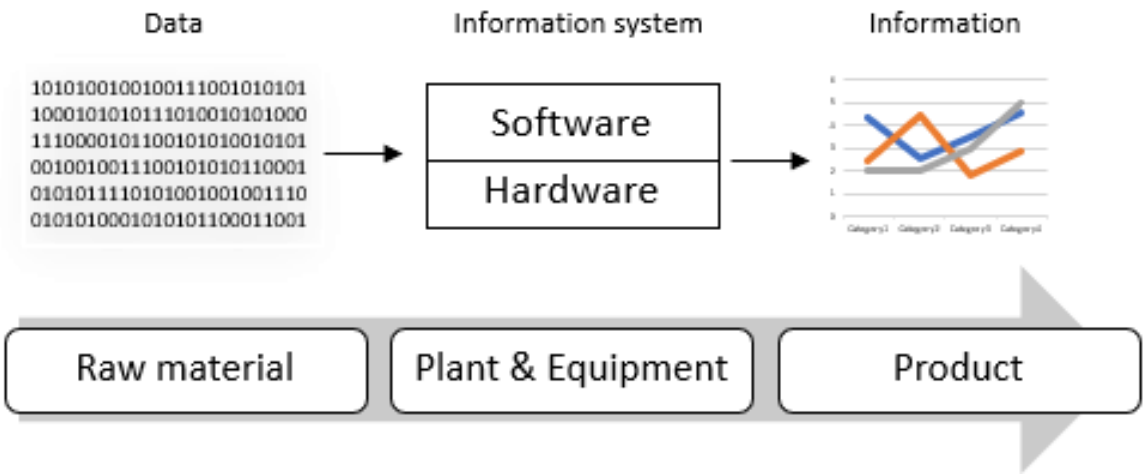

Figure 4 – Information production process (Moody and Walsh, 1999)

Big market players such as Google, Facebook, Amazon or Alibaba, daily collect huge amounts of data from their services. They use the collected data to create even more value and improve their business, by using their advanced knowledge and experience of how to extract valuable and useful information from it. However, collecting the needed data is not possible for all companies, since this process requires long-term investments and efforts. Additionally, it is not always possible to collect the data that could improve the business, by the business itself. In that case, it is challenging for companies to obtain the needed data. Therefore, one way of increasing benefits possible with data is to enable its sharing and trading, that can provide benefits to both those who collect it and make it available to others and those who obtain it. (Liang *et al.*, 2018)

According to Liang *et al.*, 2018, organizations can create commercial value from data considering two aspects. The first one is the possibility to use data for improvement and optimization of current business processes, services, and practices. The second one is the possibility to utilize data in the process of development of new business models, products and practices. Because of data, managers can measure and understand their businesses, and utilize that knowledge into improved decision making and performance. That makes data either a powerful tool or a bunch of unmanaged resources, depending on the way it is handled. Other than business and management, utilization of data has influenced research and science, by bringing new perspectives, but also challenging currently available computing power, which makes involved parties working more towards the improvement of data potential utilization. (McAfee and Brynjolfsson, 2012; Philip Chen and Zhang, 2014)

El Arass and Souissi (2018) recommended a lifecycle that data should go through in order to utilize its true value in a smart way. It is called “Smart Data Life Cycle”, and it consists of fourteen phases, which either operate directly on data or support the complete data lifecycle processes. All phases are listed in the following Table 2.

Phase	Explanation
Planning	Identification of potentially valuable data within an organization, and deciding on data management processes.
Management	Making sure that all the phases are effective individually, as well as between each other.
Collection	This step includes gathering the raw data and arranging it into a predefined structure. That involves cleaning the data and assuring data quality. Data can be gathered from internal sources, external sources or both.
Integration	If data is gathered from different sources that do not have a coherent structure, it needs to be integrated into the previously well-defined structure.
Filtering	Excluding data parts, even if they have good quality, in case they do not add value to the business. This is done to save the resources of storage, calculation, and optimization.
Enrichment	Involves adding information to the collected data in order to increase its quality. This step has to be well-planned and defined in advance as well.
Analysis	This is the most important phase of the data lifecycle. It involves a special set of operations on data, with the goal of extracting meaningful and useful information and conclusions from it, that serve as an input for decision-making processes.
Access	It is the step that involves Data Consumer that is accessing the data through some software component or application.
Visualization	Data can be shown and inspected in various graphic representations, that can increase understanding of the analysis results and influence decision-making processes.
Storage	It is important to ensure that data is securely stored and flexibly accessible at any phase of the cycle. Storage should be reliable and available and should make traceability of data possible at any time.
Destruction	Once the data is successfully used, its storing often does not add value to the business anymore. However, this step involves great attention, so that really only the useless data is destroyed, and that it is performed in a smart and efficient way.
Archiving	Data can be stored in a long-term manner, which involves encryption techniques, long-distance storage, and data retrieval mechanism. This kind of storage is isolated from "fresh" data, in case the company wants to re-use it again. In that case, it will re-involve all the steps of the data lifecycle.
Security	This aspect incorporates prevention of unauthorized alteration of data since the data is often a crucial part of decision-making processes. Therefore, its quality and reliability must be assured. Furthermore, it includes privacy protection, by doing operations on data that will prevent leaking of personal information about individuals. (Khan *et al.*, 2014) These challenges will be discussed in more details in 2.2.3
Quality	Quality requirements have to be set in the early phases of the data lifecycle and consistently checked throughout all of the phases. This involves setting parameters for data quality and having a procedure of measuring data against them.

Table 2 - Data lifecycle (own presentation based on El Arass and Souissi (2018))

2.1.2 Data as a product or service

A data product is a package of data, also referred to as a dataset, with a price attributed to it (Fricker and Maksimov, 2017). There are different types of data vendors who are offering data as product or data services. The categorization of those vendors considering their products/services are listed in the following (Schomm *et al.*, 2013):

- *Web Crawler*: a specific service that is crawling one or more websites. It is an automated process with a purpose to collect some specific type of information.
- *Customizable Crawler:* a crawler which allows users to specify properties for crawling of specific websites.
- *Search Engine*: services where user can specify keywords as input to data collections and get relevant search result as an output.
- *Raw Data Vendor*: data is not processed and structured, but rather raw, in the form of a table or list.
- *Complex Data Vendor*: data that has gone through a sort of structuring or analysis, and does not represent raw data as before.
- *Matching Data Vendor*: instead of rebuying the complete dataset that potentially changed throughout the time, the customer can use matching data services, that correct his/her internal dataset with the up-to-date one, in order to keep it valid.
- *Data Enrichment Services*: adding value to the existing data by doing one of the following:
    - *Tagging*: adding attributes to the input data in the form of tags
    - *Sentiment*: discovering a feeling or opinion of people about something through a special way of data interpretation, and adding it to the offered data
    - *Analysis*: discovering facts, descriptive and predictive information and adding it to the offered data
- *Data Marketplace*: data is being sold and bought through the platform infrastructure that supports those transactions.

When any physical product is ready to be traded, it is labeled with its specific attributes and properties. Those attributes and properties can be analyzed and measured, and they are necessary for the further handling of the product within the market (for price determination, qualitative evaluation or the definition of user authorizations). However, when it comes to data as a product, for many companies it is still a challenge to handle and manage its specific attributes. (Spiekermann *et al.*, 2018)

Data source properties can be classified into three primary concepts (Dekkers, 2013):

- *Repository:* storage and maintenance facility for descriptions of Assets and Asset Distributions.
- *Asset:* the abstraction of real content and its attributes.
- *Asset Distribution:* a material form of Asset, which can take the format of downloadable computer file, an API or a paper document.

In addition to the listed three properties, Spiekermann *et al.* (2018) argue that if data is going to be sold and considered as a product, it should be treated as any other product as well. Therefore, they extended the classification from Dekkers (2013) in order to describe data sources, and proposed the following metadata properties for data goods:

- *Curator*: the person responsible for the data source, which does not have to be a publisher
- *Subscriber*: user of data source; can be application, process or individual
- *Accessibility:* how accessible it is on the scale from private to public
- *Owner:* does not have to be the publisher as well, but someone who have ownership rights over the data sources
- *Rating:* a value estimation of data product coming from users
- *Alternatives:* other possibilities for data gathering
- *Origin:* where is it collected from
- *Dataset count:* number of included datasets in the data source
- *Update cycle:* frequency of updating the data source
- *Encryption:* the method used to make data secure
- *Authorization type*: user management system, if exists
- *Compression type:* how the data is compressed
- *Access type:* what technical methods have to be performed to retrieve the data
- *Sample*: a representative subset of data for exhibition
- *Volume:* can be static or dynamic
- *Billing model:* includes pricing model, discounts and way of payment
- *License document:* Service Level Agreements (SLA), rights and restrictions
- *Right statements:* who owns the data, who can access it and what is allowed to do with it

Data that is being traded as a product can come from different sources. It can be pulled directly from public online resources, self-generated by its vendors, collected from user inputs or community participation, captured by government or different authorities. It can be static, meaning that it consists of non-changeable facts that are relevant for a long period of time, or dynamic, which means it has relevance for the short time after creation and it needs regular updates to remain valid. Naturally, the language of data depends on its source and purpose. However, most of the data products are available in English and German language. It can be offered for its users through API, traditional download, specialized client software or a Web interface. (Schomm *et al.*, 2013)

The research conducted by Stahl *et al.* (2014) shows that raw and unstructured data is being offered much less then enrichment services and processed data. Moreover, specialized software solutions and Web interfaces are being more often provided than a few years before the survey was done. Also, more and more

datasets are offered in languages other than English and German, and the demand for up-to-date data is increasing. (Stahl *et al.*, 2014)

Data as an asset, as well as information extracted from it, has a unique characteristic comparing to a normal assets that can be shared or traded as a product (e.g. physical product): no matter what amount of data is already shared between any number of people, businesses or organizations, it neither loses its amount nor value (see Figure 5). On the contrary, the value of information extracted from data can be increased once it is shared among more beneficiaries (see Figure 6), but it can also bring more power to those who have it. This is sometimes the reason not to share data openly. In other words, since knowledge is considered to bring power, not everybody is eager to share power with anyone else. (Moody and Walsh, 1999)

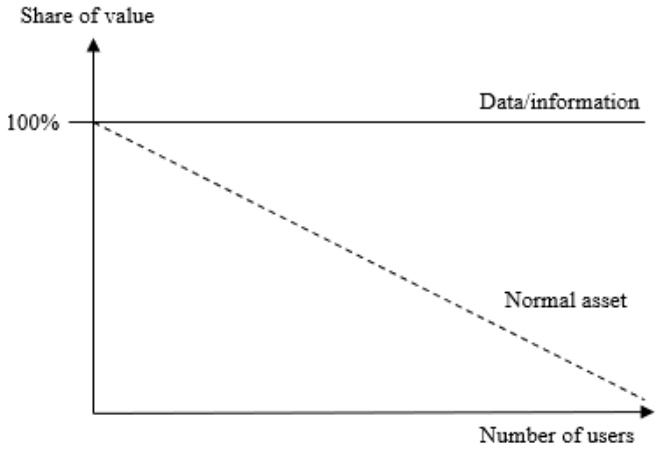

Figure 5 - Data shareability comparing with any other asset (Moody and Walsh, 1999)

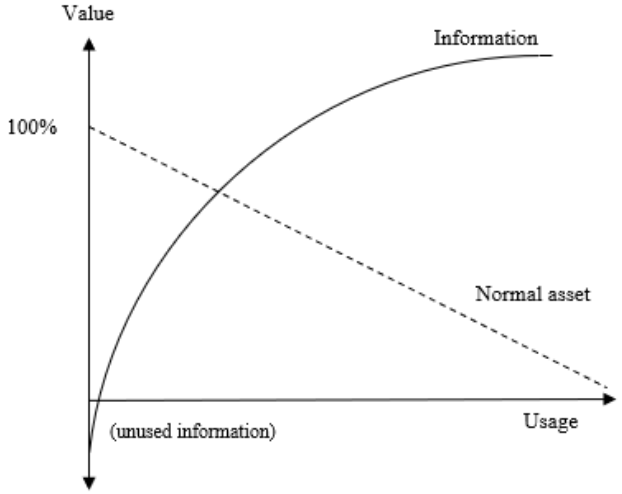

Figure 6 –Increase of information value with usage comparing with any other asset (Moody and Walsh, 1999)

2.2 DIGITAL PLATFORMS AND DATA MARKETPLACES

This subchapter provides the theoretical basis on data marketplaces, by introducing digital platforms concept first, followed by data platforms and marketplaces, as a specific type of the concept. Types of structures of data marketplaces are described. Furthermore, existing definitions of data marketplaces are reviewed and combined into one main definition that the thesis follows. Moreover, an overview of data marketplace participants is provided. The subchapter continues with typical challenges data marketplaces are facing, price setting mechanisms for datasets, and economic perspective on data marketplaces. The subchapter closes with the description of possible infrastructures of data marketplaces.

2.2.1 Introduction to digital platforms and marketplaces

Digital platforms can be seen from two different perspectives: *economic theory* and *engineering design*. The former one sees platforms as a kind of business that serves customers and suppliers who interact and exchange between each other through pricing strategies. This type of digital platforms can also be referred to as "two-sided market", "multi-sided market", but the very commonly used term is "electronic marketplace" as well. The latter perspective sees platforms as technological architectures that have a purpose to facilitate innovation. This design supports and stimulates firms in achieving economies of scope in innovation. This is rather a perspective with some limitations – innovation happens on modules, within stable system architectures, and facilitated by stable interfaces. (Gawer, 2014)

The type of digital platforms that this thesis focuses on is the one that observes platforms from the economic theory perspective since this type of platforms is the underlying infrastructure of data marketplaces.

Electronic marketplaces are a rapidly emerging category of digital platforms which have an open business model, that enables and supports transactions between independent participants. The participants of an electronic marketplace co-create value. This phenomenon is called the "network effect", an effect that happens when a good or a service has more value for its user the more users adopt it (Shapiro and Varian, 2010). The network effect is a self-reinforcing mechanism which can lead to a "winner-take-all" outcome – first players who reach the market leadership and recruit the majority of customers, reach high economies of scale and set too high boundaries for other players (Eisenmann *et al.*, 2006; Carnelley *et al.*, 2016.). For example, potential customers of a video game console will be interested in buying one only if there are enough video games available for it. On the contrary, game developers will be interested in developing games for the console only if there are enough customers who are already using the console. It means that both buyers and sellers, as well as other participants, have clear direct benefits when being matched with each other in order to exchange goods or services. The electronic marketplace itself is a facilitator of those exchanges. In other words, electronic marketplaces are the type of digital platforms, seen from the perspective of economic theory. Moreover, the participants of an electronic marketplace do not have to act only on one side, as a buyer or seller. They can participate in both, the supply and the demand side, and therefore they do not necessarily represent two different groups of users. However, electronic marketplaces do not create or trade goods or services by themselves, but they are rather an institutional and regulatory frame for transactions. (Koutroumpis *et al.*, 2017; Täuscher and Laudien, 2017; Osterwalder *et al.*, 2010)

There are three levels of organizational settings within which digital platforms can be observed: *firms, supply-chains, and industry ecosystems*. Therefore, three types of platforms can be distinguished, as presented in Table 3 (Gawer, 2014):

Type of platform	Description
Internal platform	The platform and its accessible innovative capabilities are observed on the level of the firm, including its constituent subunits. Its interfaces are open internally but closed externally. Coordination mechanisms are achieved through an authoritative managerial hierarchy. Example of this type is consumer electronics platforms.
Supply-chain platform	Observed on the supply-chain level of analysis and includes assembler and suppliers. Its interfaces are selectively open across the supply chain. Accessible innovative capabilities depend on the supply chain's capabilities. Similarly, its coordination mechanisms depend on contractual relations between supply-chain members. An example is automotive or aerospace manufacturing.
Industry platform	Observed on the industry ecosystems analysis level, which includes platform leader and complementors. It has open interfaces and a potentially unlimited pool of external capabilities. Coordination mechanisms are achieved through the governance of the ecosystem, and in the case of multi-sided markets – through pricing. An example is a social networking platform.

Table 3 – Typology of digital platforms (own presentation based on Gawer (2014))

Different types of platforms, that are presented in Table 3, provide different organizational continuum when it comes to innovation and competition. An organization can evolve and move from an internal platform to a more open platform type. Platform interfaces that are more open, attract more agents into its ecosystem, and the platform leader is potentially able to achieve more attractive innovations. But, as the platform becomes more open, for example by providing APIs to other smaller and independent parties which want to incorporate it into their own solutions, it opens for complementary innovation to the platform, but most certainly to a competitive innovation as well. This is visually presented in Figure 7. (Gawer, 2014)

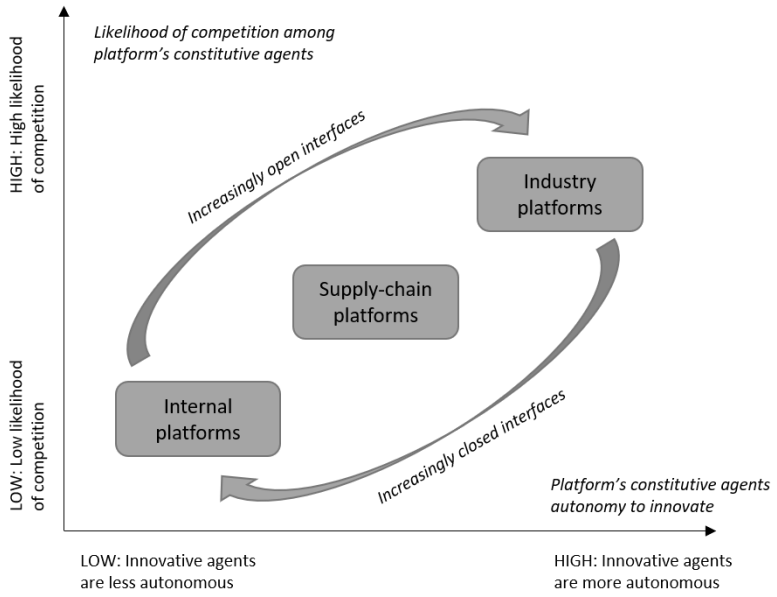

Figure 7 – Platform innovation and competition (Gawer, 2014)

In general, marketplaces can facilitate the trading of any type of products and services. They can take any size of the business, either when it comes to the number of employees or market participants. For the sake of some representative information, some general statistical details will be presented in the following. According to the research done by Täuscher and Laudien (2017), 90 out of 100 digital marketplaces are businesses that have up to 50 employees, and their core businesses are almost equally distributed over the trading of physical products (28), offline services (28) and digital services (30). Most of the businesses (75%) provide value through increased efficiency or cost savings, and the majority of them (60%) implement C2C model, where customer can take both sides of trading – supplier or buyer. Furthermore, revenue is mostly gained through commission fees (72%), followed by subscription fees (22%). (Täuscher and Laudien, 2017)

Two common terms are often being confused in everyday conversations without taking their differences into account: *market* and *marketplace*. Markets are concrete places where buyers and sellers interact by determining the price and the quantity of a good or service. On the other hand, marketplace provides the infrastructure for trading; it is the explicit time and location when and where market participants prepare and execute their transactions (Stahl *et al.*, 2016; Schmid and Lindemann, 1998). However, following the definition from Schmid and Lindemann (1998), electronic marketplaces are a special case of traditional marketplaces, since they are completely independent of time and space, due to their digital nature. Therefore, it is enough to define digital marketplaces as a system that matches buyers and sellers to trade goods under predefined terms of exchange (Koutroumpis *et al.*, 2017).

2.2.2 Data marketplaces

Markets for data trading have gained popularity in recent years, due to the need for the right information at the right time in almost all business areas. The term "data marketplace" was probably used for the first time in 1998 (Schomm *et al.*, 2013), but nowadays interest in data markets to lever public as well as private data and analyze it in order to create novel consumer and enterprise value is clearly visible. If data is not traded or exchanged, it remains static, while data trading makes it flow dynamically, and establishing a win-win market (Liang *et al.*, 2018). A data marketplace provides access to usable datasets for analytics providers and data-driven services, and they are becoming necessary as the data market matures. With the more data marketplace players joining the market, companies have more opportunities to use external data for the improvement of their business (as will be discussed more in 2.3.2), as well as new revenue opportunities by reselling the data created internally through their business. (Stahl *et al.*, 2014; Carnelley *et al.*, 2016; Muschalle *et al.*, 2012)

A simple overview of the most important components and stakeholders of a data marketplace is provided in Figure 8:

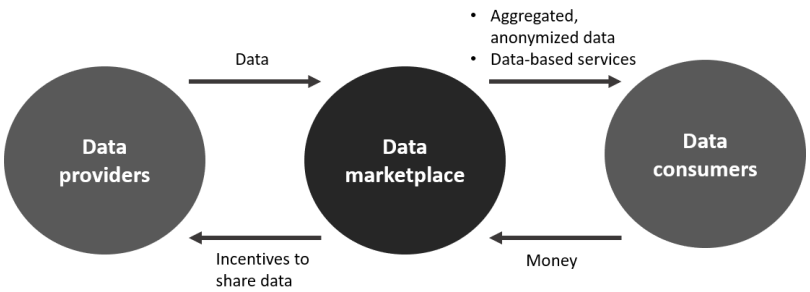

Figure 8 – Data marketplace (Deichmann et al., 2016)

Data providers are stakeholders that offer their data in the form of datasets on a marketplace. The datasets provided on the marketplace can be queried and explored. Data providers expect to profit from selling the provided datasets. They have knowledge about their datasets, but they also get to know their customers and the distribution of their demands. In other words, the marketplace, that they choose to sell their data on, has the authority to license their information by following predefined terms and conditions. On the other hand, data consumers are those participants who have an interest in buying needed datasets. The interest in buying the datasets means that they have "*Willingness to pay".* In return for their money, they get data and/or data-based services for further processing and analyzing. The middle layer is a data marketplace, which brings those providers and consumers together and makes trades possible, by providing them a central point of collaboration and trading. (Deichmann *et al.*, 2016; Kushal *et al.*, 2011)

Data consumers can take the role of data providers too, in case they have data that can be sold, and that they want to sell – the separation, between data providers and data consumers, is not strict. Additionally, as shown in Figure 8, data marketplaces can themselves use data they collect about trades that are active on the marketplace, learn about the demands and needs of data consumers and collaborate with data providers in order to stimulate them to provide their data on the marketplace. (Deichmann *et al.*, 2016)

Data sets that a data marketplace offers can be bought from data providers, and further resold to data consumers. This approach makes data marketplace a complete owner of the data sets. Another possible approach is to collect data from a variety of sources, aggregate it and resell it to interested parties. Furthermore, depending on the maturity of a platform provider, data marketplace evolves over the time, from selling raw data and acting only as a simple broker, through selling normalized and verified, or even aggregated data sets created from different data sets obtained from various providers, towards a marketplace that performs consistent checks, prevents trade of invalid data and assures data quality (see Figure 9). This is just a possible transition and not a rule. It is possible to target the quality assurance at the stage of setting up the business, although more challenging to succeed. (Deichmann *et al.*, 2016)

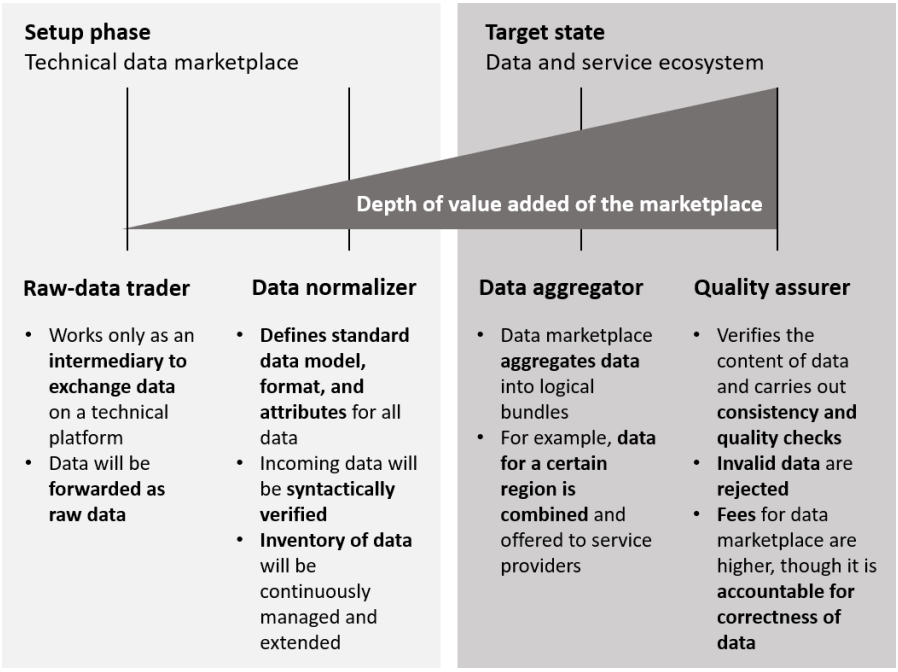

Figure 9 –Possible data marketplace evolvement considering maturity (Deichmann et al., 2016)

2.2.2.1 Structures of data platforms and marketplaces

Data marketplaces can be organized in different structures, which dictates how users will approach and use the market and how the pricing for data and associated services will be determined. As visually presented in Figure 10, electronic marketplaces providers can be positioned on a scale that varies from market orientation to hierarchy orientation. Market orientation means that the quantity and price are determined by market forces among competitive offerings. However, hierarchical orientation means that specific price and specific buyers are pre-determined and the market is limited to it. Furthermore, electronic marketplaces are also distinguished by ownership. They can be *private* (owned by a single company), *consortia-based* (owned by a small number of companies) or *independent* (run as a platform). (Stahl *et al.*, 2016)

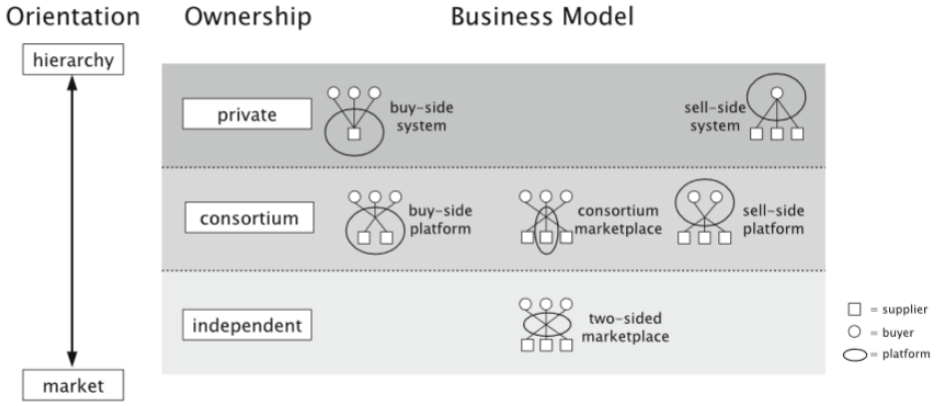

Figure 10 – Typology of electronic marketplaces (Stahl et al., 2016)

Based on this classification (see Figure 10), six business model types can be distinguished. Hierarchy oriented electronic marketplaces with private ownership, generally implement the *buy-side system* (one supplier offers goods to many buyers) or *sell-side system* (one buyer demands goods from many suppliers). This is also called "one-to-many" and "many-to-one" systems respectively. The former one has so-called *dispersal* marketplace design, where terms of exchange are standardized, rather than negotiated, with the purpose of reducing the time needed to individually decide about each relationship. In this case, data distribution is automated through APIs, without relationship monitoring, which makes it harder to strategically observe the behavior of buyers. On the other hand, many-to-one design, also called *harvest* marketplace design, has a characteristic of data harvesting, where a single service provider is collecting data from multiple users, who make it available under predefined terms of the exchange that each of them independently specifies. In other words, the provider of services offers those services for free, in exchange for the data (e.g. Google Search – users provide search terms in return for data about the search). (Stahl *et al.*, 2016; Koutroumpis *et al.*, 2017)

On the other side of orientation and ownership scale illustrated in Figure 10 is the *two-sided marketplace* business model, that follows *multilateral* marketplace design. It connects multiple suppliers with multiple customers independently from each other. Anybody can be a participant on any side of this market. Efficiently designing and running multilateral marketplaces can be challenging, because of the need for an environment supported with legal regulations, communication standards, data protocols, procedures for data import, storage, transformation, aggregation, analysis and delivery, adequate pricing and the

interaction of participants. Moreover, a multilateral marketplace needs clear licensing standards, security, information about data origin, and data validation. (Stahl *et al.*, 2016; Koutroumpis *et al.*, 2017)

And finally, in between, consortia-based ownership system distinguishes three more possible business models – *buy-side platform* (few-to-many principle), *consortium marketplace* (few-to-few principle) and *sell-side platform* (many-to-few principle). Those three are typically collaboration of several companies in the same industry, and most of the times the entry into this kind of platform, as a new player, is only theoretically possible. (Stahl *et al.*, 2016)

The only type of data marketplaces whose details about their business model are expected to be publicly accessible is multilateral data marketplaces. This is because they are market-oriented and they provide their business model design details openly in order to attract any interested parties to trade on their platform (Stahl *et al.*, 2016; Koutroumpis *et al.*, 2017). However, they are also the most complex and the highest emerging type of electronic marketplaces, that can succeed only with a well-structured business model (Stahl *et al.*, 2016). Therefore, this thesis will focus on business models of multilateral data marketplaces in further chapters.

2.2.2.2 Defining a data marketplace

Carnelley *et al.* (2016., p. 10) defined a data marketplace as "*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.*". They emphasize that a data marketplace has to be a platform and not a repository of datasets or cloud service providers. Therefore, they have to enable transactions between market participants, and they are presenting an online store for data, based on the multi-sided (multilateral) business model.

Stahl *et al.* (2016) propose inclusion and exclusion criteria for data marketplaces. Therefore, according to this definition, the primary business model of the electronic data marketplace needs to be providing data. Users of the marketplace infrastructure need to be able to upload, browse, download, buy, and sell machine-readable data (data that has a format that can be easily processed and directly understood by a computer, e.g. RDF, XML, JSON (Hendler and Pardo, 2012)). Moreover, data has to be hosted by the providers and its origin needs to be clear – if it comes from the community or the operator. On the contrary, if the data offered and traded is not in a machine-readable format, this case cannot be considered as a data marketplace. As an example, Wikipedia can look like a data marketplace because it allows users to upload, browse, download and use freely available data. This kind of data is shared for personal use among readers, but it does not consist of machine-readable data that can effectively be distributed and directly used for computer-based analysis. Furthermore, this definition excludes the services that only offer links to data locations, without hosting the data itself. Additionally, the definition excludes electronic frameworks that gather providers who do not offer transparency of data sources, because this leads to a lack of trust about data origin and type of offered data. Lastly, providing data for free by government agencies or non-government organizations is excluded as well. This means that the published data is only a side effect of their purpose in general, and they are not set on commoditizing data or finding an appropriate business model to financially utilize it.

Although definition from Carnelley *et al.* (2016.) and the one provided by Stahl *et al.* (2016) highly overlap, the major difference between the two is the emphasis of the former that data can be offered for free, or can be chargeable, depending on the decision of data provider. Consequently, they state that data

marketplace definition includes open data portals, that are usually run by the government and have no-profit business models, allowing data sets access to be open for public use.

Encouraged by the definition of electronic marketplaces provided by Schmid and Lindemann (1998), that involves keywords "*buying"* and "*selling"*, and the lack of profit-oriented nature of open data marketplaces (Zuiderwijk *et al.*, 2014), open data portals will be excluded from the definition of data marketplace for the purpose of this thesis.

2.2.2.3 Data marketplaces participants

To better understand the need and purpose of data marketplaces, it is necessary to know what kind of demands, interests and needs different data marketplaces participants have. Seven different groups of beneficiaries can be distinguished (Muschalle *et al.*, 2012; International Data Corporation (IDC) and the Lisbon Council, 2018):

Data participant	Description
Analysts	Members of this group are continuously discovering knowledge from various sources of data by searching it, querying it, combining it and summarizing results from it. The sources can be Web search engines, public datasets from the internet, internal datasets of the enterprise, or data collected from other services. Analysts are in constant demand for relevant datasets and data services, with the aim of creating useful and meaningful reports and insights from it. Typically, those are marketing experts, sales executives, product managers, brokers, marketing managers, and business analysists.
Application vendors	Application vendors work with data from data markets, prepare it and turn it into easier to use interfaces, so that other users, who do not have such technical knowledge, can take advantage of it. In other words, the requirements from analysts are formalized into a machine-readable representation. Examples are business analytics applications, customer relationship management applications or enterprise resource planning applications.
Developers of data associated algorithms	Both data analysts and application vendors have use of developers of data associated algorithms, who are performing data cleansing, computing, tracing and data mining algorithms on data, in order to prepare it for them. These algorithms are often domain, language or quality specific, and developers are sometimes implementing it and selling it as a black box so that others can buy them and try them out.
Data providers	These participants use data markets for storing, selling and advertising data, or offering data integration algorithms. They can be non-commercial providers like Bing or Google, Web archive or Web forum owners, or commercial like banks, government agencies, financial or geo sectors.
Consultants	Consultants support organizations in tasks like data source selection, integration, evaluation, and product development.

Licensing and certification entities	In order to help customers with buying data-related products and services, licensing and certification entities are putting labels on the data, applications, and algorithms.
Data market owner	Those are participants who own the marketplace and develop and provide an infrastructure for it. This leads to many technical, ethical, legal and economic challenges. It is necessary to establish a trusted brand and community, and provide users with the possibility to browse, store and exchange data and related algorithms through the platform in the minimized amount of time.

Table 4 – Data markets stakeholders (own presentation based on Muschalle et al. (2012) and International Data Corporation (IDC) and the Lisbon Council (2018))

2.2.3 Data marketplace challenges

In general, matching buyers and sellers for the purpose of data trading is not much different than any other traditional market. However, being a data marketplace company means being in an emerging, dynamic and challenging environment. In 2013, a survey was performed on (at that time) current data marketplaces, and a similar survey was repeated by the same authors in 2014. From 46 companies analyzed in 2013, 4 companies no longer existed or have changed their business. Also, 5 new companies appeared on the market. This example illustrates the changing dynamics of the market (Stahl *et al.*, 2014).

Just like other digital platforms, data platforms are exposed to the challenges of the possible network effect, and since this kind of business is still shaping and emerging, it has a potential thread of dominance by only a few companies who reach market leadership and recruit the majority of potential users (Carnelley et al. 2016). However, data platform providers have additional challenges, that are not typical for platform and marketplace businesses in general. The main data marketplaces participants are data buyers and data sellers. As shown in Figure 11, the main features a data marketplace should provide for those participants are (Banerjee and Ruj, 2019; Koutroumpis *et al.*, 2017):

- *Fairness:* after the trade, both buyer and seller have to be satisfied – the former one with the quality of received goods considering the promised quality and the price previously agreed on, and the later one with the profit. Sometimes, this is solved by introducing service-level agreements that serve as a promise that the data marketplace will keep delivering high-quality data consistently (Deichmann *et al.*, 2016).

- *Transparency and Privacy:* except buyer (and only after a successful trade), no one else should have a privilege to see and use the dataset during its selling time. Also, only the buyer and seller should know which datasets have been traded, and the price, terms, and conditions of trade between the platform and the seller should be transparent to the buyer when buying it. This is one of the biggest factors that prevent data marketplaces from rapid development.

- *Safeness and Security*: data marketplaces need to implement an effective way of preventing participants with bad intentions to influence the marketplace since it affects the trust of trading participants towards the marketplace. It involves the sellers who are taking advantage of matching with interested buyers through the platform and then executing the transactions outside of it, or those buyers who are reselling the datasets illegally once they buy it on the platform, or any other way of misusing the platform.

- *Regulation (non-repugnancy): a* different type of data can have different amounts of sensitivity or legal boundaries for distribution, which can vary from country to country. This can produce many

challenges, and data marketplace has to implement social and legal policies for the affected because all the powerful algorithms and attractive benefits of data marketplaces can be useless if users sense the lack of legal support.

- *Thickness/liquidity:* suppliers and buyers should both have a chance to choose between a variety of trading partners with whom they are interacting. Otherwise, marketplace popularity growth could be negatively affected.
- *Efficiency:* in order to be usable and accepted, implementation of data platforms has to be comparable with the current technological trends. The platform popularity should not, in any case, influence the efficiency of making transactions. Once the market participants are willing to trade, they have to be able to do it in a painless way.

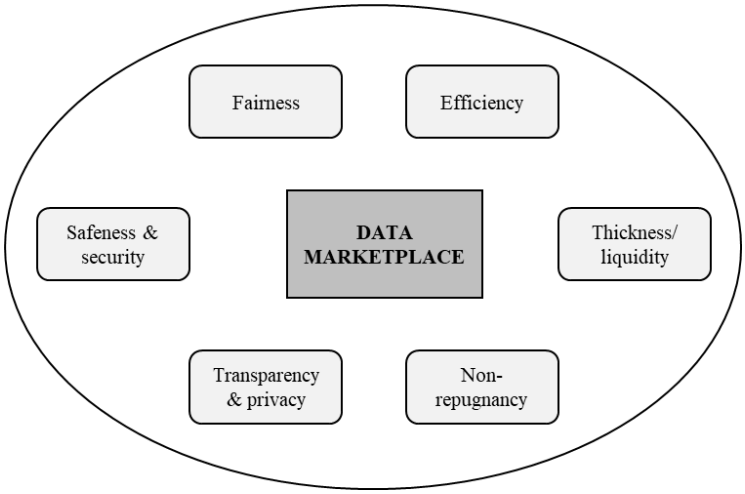

Figure 11 – Data marketplace challenges (own illustration)

Data marketplace should control which users can access it, by setting a limitation for thickness (liquidity) of the marketplace and controlling the access by identifying its participants. Usually, it is done by controlled user accounts creation. Thus, it increases the trust of buyers about the quality of traded data and its origin. Consequently, the number of participants should be under control just enough so that the marketplace remains needed liquidity, but also to prevent bad intentions to interfere with the user experience of serious clients. However, even with this regime, it is difficult to control how datasets will be used (or misused) once they are bought, since malicious sellers, mediators or buyers are likely to appear or collude with each other. Therefore, another important requirement is setting the legal rules for using data resources, on the platform, and outside the platform, as well as consequences for not following them. To be able to conduct those rules, stable monitoring by core users or third parties are implemented. The purpose of the monitoring is to stop suspicious activities, once they are noticed, as well as to decrease and prevent them in the future. (Koutroumpis *et al.*, 2017; Banerjee and Ruj, 2019)

Another challenge that data marketplaces are facing, unlike typical marketplaces, is that real data quality and value of information can be determined only by seeing data itself; metadata provided by seller or platform is not enough for conclusions. But, if data is uncovered before purchase, the paying part does not make sense, since the buyer already got data for free. This is generally called *Arrow's information paradox (Piazza and Pedicini, 2018)*, and data marketplaces suffer from it. Naturally, marketplace providers are not eager to show data in advance, which creates additional difficulties. Impossibility to check the value of good before buying it, influences the customer's willingness to pay, and often the price is the only criteria for making decisions about the purchase. (Stahl *et al.*, 2017; Kushal *et al.*, 2011)

2.2.3.1 Pricing mechanisms

In order to profit from providing an infrastructure for data trading, data marketplaces introduce transaction-related fees and commissions into their business model (Deichmann *et al.*, 2016). On the other hand, a pricing mechanism between buyers and sellers that participate in the trading activities must be predefined as a part of the data marketplace business model as well. Defining prices for products requires a set of rules, which describe product and variables that contribute to its value. This is called the pricing model, and it has to be applied to a data product before it is being traded. Pricing models can have different characteristics, therefore, there are six main categories which are common within data marketplaces (Muschalle *et al.*, 2012; Fricker and Maksimov, 2017; Stahl *et al.*, 2017):

Pricing model	Description
Free	This does not mean that all the data available on the platform is provided free of charge, but rather selected datasets. This is a strategic move to increase the popularity of the marketplace within consumers, with the aim of attracting more suppliers consequently.
Usage-based prices[1]	For this model, data has to have its unit (e.g. API calls) and price per unit. Customers pay proportionally to the number of units being bought. This a simple solution, but since data trading has some specific characteristics, it can be a tricky solution instead. Dataset is produced and prepared once before it is ready to be sold. Whether it is sold only once, a few times or many times, the seller does not have any new costs for producing it. Therefore, marginal costs are converging to zero, and prices become groundless, which leads to dissatisfaction of customers.
Package pricing	A selected amount of data or API calls is offered for a fixed price. This can lead to arbitrage, depending on the package size.
Flat fee tariff[2]	The simplest pricing model category, where the fee is charged regularly (on a monthly or annual basis), as agreed in the contract, and not dependent on usage. This model is often used for software licenses, and it has minimal transaction costs, but it lacks flexibility.
Two-part tariff	This model combines usage-based pricing and flat fee tariff pricing. The fee is paid regularly, but an additional fee is paid also according to usage.
Freemium	Users are free to access and use basic features of the platform, but they can also join premium services which have more value and benefits. The premium account needs to be paid a predefined price. This payment has to take some form of those listed above, therefore Freemium is almost always used in combination with the Usage-based or Flat fee model.

Table 5 – Pricing models for data marketplaces (own presentation based on Muschalle et al. (2012) and Stahl et al. (2017))

[1] Also known as *Pay-per-Use* model (Stahl *et al.* (2017))
[2] Also known as *Flat rate* model (Stahl *et al.* (2017))

A survey conducted by Stahl *et al.* (2017) among different data marketplace providers, showed that more than a half providers (54.2%) use only one pricing model, while 25% offer Freemium. When it comes to Freemium model, providers prefer to use it in combination with Flat fee model (61.1%), while many of them also provide to customers options to choose between Flat fee or Usage-based model in combination with Freemium model (27.8%). Only the small rest of providers offer the Freemium model in combination with Usage-based model. From providers that offer a pricing model other than Freemium, 59% offer Flat fee, while 41% offer Usage-based option. Flat fee model provides a higher possibility of profit than Usage-based model (Muschalle *et al.*, 2012), and this survey showed that it is indeed preferred among providers. (Stahl *et al.*, 2017)

The decision about the price is also influenced by the present competition. Data markets can be divided into three categories according to their pricing structure, that highly depends on market situation, amount and strength of competitors. Those categories are *monopoly*, *oligopoly*, and *strong competition.* More details about this division are given below in Table 6. (Muschalle *et al.*, 2012)

Market structure	Description
Monopoly	Suppliers have no relevant competition, which allows them to set prices to a level that will maximize profits, by selective pricing for different demanders with different preferences (price discrimination).
Oligopoly	The market is dominated by a few competing suppliers, who are constantly challenging each other, fighting and/or compromising for market share. Monopolistic prices are out of the game in oligopoly; prices are adjusted competitively and by complex analyses and forecasts of competitor's behavior.
Strong competition	Market prices of offered goods tend to align with marginal costs because individual suppliers are not powerful enough to set prices for profit-maximization. It can create a dangerous situation where the overall costs, that have to be covered, are not relevant for short term decisions. The only way out of profit loss is squeezing overall costs.

Table 6 – Typology of markets structure by pricing (own presentation based on Muschalle et al. (2012))

Just like in the typology of electronic marketplaces (Figure 10), the typology provided in Table 6 is not only characteristic for data marketplaces, but for marketplaces in general. But when it comes to data marketplaces, according to the interviews conducted by Muschalle *et al.* (2012) among data marketplace providers, the perceived competition differs significantly. Some interviewees are instantaneously capable of naming their strongest competitors, some of them are aware of similar products and offerings on the market, but not particularly bothered by them, while others claim that they do not have any known direct competitors.

2.2.3.2 Data economy and regulations

Continuity and variety of digital innovations introduce fast pace changes in a broad spectrum of markets, including the data market as well. However, data market is influenced by inequality of data innovation distribution among the different parts of the world and different countries, as well as by the

introduction of regulations that are being made with the goal of protecting data market stakeholders from the loss of control over data being exchanged/traded (e.g. GDPR – General Data Protection Regulation in European Union countries). (International Data Corporation (IDC) and the Lisbon Council, 2018)

The International Data Corporation (IDC) and the Lisbon Council (2018) presented a report consisted of the main indicators of the data market in 2016 and 2017, as well as potential directions and changes that could occur in the data market in 2020 and 2025, for the countries of European Union. They considered three possible scenarios:

- *Baseline scenario:* mostly expected and realistic scenario, where the growth of the data market follows already known pace, without high dominations among data owners or high differences in the distribution of data innovations among the society.
- *High growth scenario:* a scenario that could take place if the data innovations rise drastically, with open and transparent data governance and high data sharing. This scenario can also be called *Data-driven reality.*
- *Challenge scenario:* this scenario predicts a low and poorly distributed data innovation benefits among the society, due to digital markets fragmentation. It can also be called *Digital maze*

Table 7 presents the key indicators of the data market situation in 2016 and in 2017 in the EU, including the number of data suppliers and data users, expressed in the scale of thousands, as well as the value of data market and data economy, expressed in Euros, in the scale of billions. Furthermore, the three scenarios are presented as well.

	Facts		***Baseline Scenario***		***High Growth Scenario***		***Challenge Scenario***	
	2016	*2017*	*2020*	*2025*	*2020*	*2025*	*2020*	*2025*
Data suppliers (Thousand)	261	272	294	339	294	389	294	320
Data users (Thousand)	676	691	721	778	721	844	721	750
Value of Data Market (€ billion)	59.5	65.2	77.7	106.1	77.7	141.6	77.7	93
Value of Data Economy (€ billion)	300	336.6	477.2	680.1	477.2	1053.8	477.2	545.5

Table 7 – Data market facts and predictions (own presentation based on Cattabei et al. (2019))

The predictions presented in Table 7 show that the number of data suppliers and users, as well as the value of the data market and data economy, is going to increase already in 2020. The increase is expected to be equal in any case scenario (baseline, high growth or challenge). Furthermore, the prediction for 2025 shows that the number of data suppliers and data users is expected to reach a slow increase in any case scenario. The value of the data market and the value of the data economy is expected to have a strong increase until 2025. In the high growth scenario, data market value is expected to reach double value comparing with the value from 2017. Data economy value is expected to reach a triple of the amount from 2017. (Cattabei *et al.*, 2019)

Li *et al.* (2018) made a comparison of the data market situation existing in the European Union and the situation of the data market in China. It is estimated that the value of the data market in China reached $70 billion (~ 62€ billion) in 2015, while the estimated value by 2020 is $155 billion (~137.2€ billion). This high difference between Chinese and European data market values comes from the difference in data

exchange regulations. Unlike GDPR – regulation among European Union countries, China has no national regulation on data protection, data sharing and trading, which makes data market open for experimentations. The only regulations that exist in China, considering the data aspects, are those that protect national security and secrets, copyrights and prevention of illegal content propagation (e.g. terrorism or fake news).

2.2.3.3 Infrastructure

Data marketplaces can have different data-trading infrastructure framework – centralized or decentralized. The choice of infrastructure influences the business model of the company since all the aspects are built on top of it. (Koutroumpis *et al.*, 2017)

2.2.3.3.1 Centralized infrastructure

Centralized infrastructure is a simple and easy-to-use solution for multi-sided marketplaces in general, due to its common characteristics with any typical electronic marketplace infrastructure. Therefore, the earliest data marketplaces have had implemented a centralized infrastructure. As with digital platforms in general, data platforms provide products and/or services, for its participants who are interested in gathering on the platform and buying them. When more than one platform is offering the same or similar products or services, users can choose which one to use for trading. Thus, benefits in terms of cost, variety of suppliers or efficiency are necessary to attract users to prefer the platform over the ones from competitors. The positive involvement of participants brings benefits for the platform operator, but only when assuming that transaction costs will prevent trading parties from trading outside of the platform. (Koutroumpis *et al.*, 2017)

Centralized infrastructure design is the design that facilitates economies of scale and network effects, and which can lead to winner-take-all dynamics (as mentioned in 2.2.1). Users of the platform with centralized infrastructure are not in the position to negotiate licensing rules, terms and conditions, because those are sealed by the owner of the platform. To ensure a trustable platform, where datasets provenance is not lost once the dataset is delivered to a buyer, and transactions are traceable back to original suppliers, relatively strong entrance boundaries are needed. In other words, the platform needs to ensure a mechanism of controlling who exactly has access to the data marketplace trading activities (e.g. by implementing a strict procedure for creating user accounts and identifying users). On the other hand, those boundaries have to be carefully set, due to the risk of losing growth of the market. Boundaries can be safely omitted only in case of public data, or time-sensitive data, while the supplier of private data can be at risk once that data comes into the hands of those with bad intentions. (Koutroumpis *et al.*, 2017)

In Figure 12, the basic design of a centralized marketplace is illustrated. Data suppliers, data consumers, and data marketplace platform are the main stakeholders of the system. The data marketplace platform is taking care of data tagging, aggregating and combining, it offers ways of searching and discovering potentially interesting datasets, providing dataset validation, payment services and browsing through history. Those are the main parts of a typical centralized data marketplace, which can be extended individually by platform providers. Data suppliers visit the platform with a goal of offering potentially interesting dataset, while data consumers are in demand for some specific datasets which they eventually find in the marketplace offer. This type of design can support the first or second requirement of data marketplaces (as explained in 2.2.3), but not both at the same time. The marketplace can be seen as a trusted third party, that guarantee fairness, but in that case, the marketplace has to get insight into the dataset, which breaks the requirement of privacy. If the seller and buyer sides both unconditionally trust the marketplace

provider, this can be a basis for fraud and lack of transparency. (Koutroumpis *et al.*, 2017; Banerjee and Ruj, 2019)

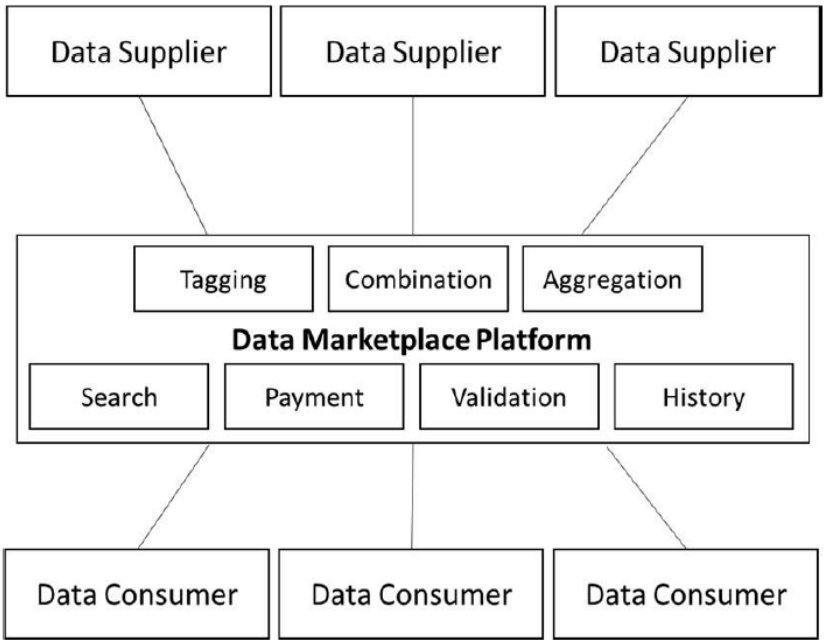

Figure 12 – Centralized marketplace design (Koutroumpis et al., 2017)

2.2.3.3.2 Decentralized infrastructure

With the goal to overcome the problems and challenges of centralized data marketplace infrastructure (like traceability of transactions, data provenance, fraud prevention, etc.), a new type of infrastructure has been developed. Some data marketplaces adopted the technology for underlying infrastructure, which was first introduced as an infrastructure of digital currencies. The new technology is called *distributed ledger technology*, also known as the *blockchain* (Iansiti and Lakhani, 2017). It is a *tamper-proof append-only* technology, which means that every transaction executed within the system is appended to the existing records of transactions, together with the details of transactions. Once the transaction is recorded on the blockchain, it is not possible to tamper it in any way. Information about the type of data, the trustworthiness of participants, quantity, the value of transactions and timing of transactions is decentralized and traceable back through the whole history of all transactions. The storage of all the information is distributed to the private storages of the participants on the blockchain, who always have a copy of blockchain where transactions are recorded. The efficiency and security of the blockchain are assured by performing complex algorithms executed on every transaction. (Koutroumpis *et al.*, 2017)

A decentralized design replaces the conventional centralized system with communication structures and algorithms because all the information stays recorded in the blockchain. With this design, proof of data origin of any data set can be independently verified. Additionally, transactions can be directly executed and verified between participants, without the need for a central authority. These complex algorithms have fewer challenges to satisfy the requirements for data marketplaces (fairness, transparency, privacy, and

security, as listed in 2.2.3). However, the optimized solutions for regulation and efficiency requirements are partly satisfied by decentralized infrastructure. (Koutroumpis *et al.*, 2017; Banerjee and Ruj, 2019)

As shown in Figure 13, a decentralized marketplace design shares many features of a centralized one. Data suppliers, data consumers, as well as data marketplace platform, are still the main part of the structure. However, in decentralized infrastructure, data suppliers keep their datasets stored on their private storage, instead of sharing them on one central platform. A decentralized data platform has to provide fewer services than a centralized data platform. Those services include matching of buyers and sellers with search/discovery functionality, combining and aggregating data into dataset products and handling payments. The processes of transaction validation, history of transactions and tagging is independent of a data platform with decentralized infrastructure. (Reuver, 2019; Koutroumpis *et al.*, 2017)

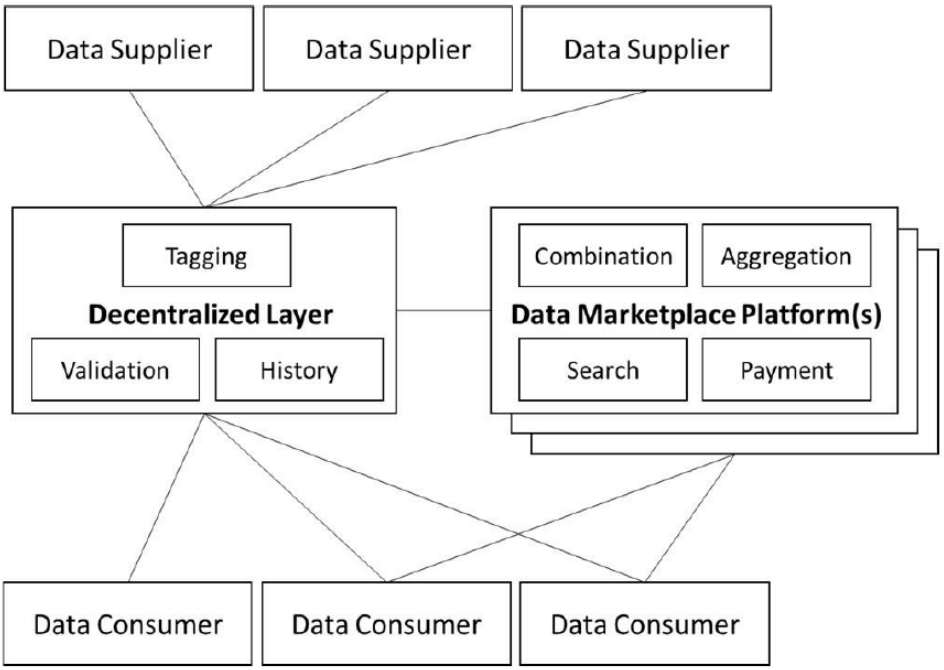

Figure 13 – Decentralized marketplace design (Koutroumpis et al., 2017)

Figure 13 is a generalized form of decentralized marketplace. Many papers propose a custom-designed decentralized data marketplace (Banerjee and Ruj, 2019; Nasonov *et al.*, 2018; Zyskind *et al.*, 2015; Ozyilmaz *et al.*, 2018; Hynes *et al.*, 2018; Dao *et al.*, 2018). All of the proposed designs use blockchain as the underlying technology, remove the need for a central authority and implement payment handling via cryptocurrency. Moreover, a specific characteristic of a decentralized data marketplace design is a possibility of smart contracts. A smart contract is a computer protocol which enables privacy-preserving, safe payments, enforcement of data provider's terms of use and general trust between buyers and sellers (Hynes *et al.*, 2018; Banerjee and Ruj, 2019). However, the inner logic of each proposed decentralized data marketplace design is different. Every proposed design implements unique algorithms with high complexity, in order to provide transparency, security, efficiency, fairness and trusted environment.

2.3 DATA-DRIVEN BUSINESS MODELS

This chapter introduces the business model as a concept, including a well know business model canvas for the creation of new or analysis of existing business models. Understanding of what a business model is and how it applies to companies is important in this thesis for two reasons. The first reason is to make discovering business models of data marketplaces possible, in order to answer the research question. The second reason is understanding of data-driven business models, which are a specific type of business models involved in the data ecosystem.

2.3.1 Introduction to business models

Business models can be defined in many ways, and researchers who provided definitions so far do not always completely overlap with each other (Teece, 2010; Zott *et al.*, 2011). Also, business model as a concept is often used without providing its clear definition in advance, which is the reason why some great innovations fail commercially due to a misunderstanding (Zott *et al.*, 2011; Teece, 2010). Nevertheless, a general explanation of what a business model is, and an overview of its characteristics will be presented in this chapter.

The business model of a company is a structure of creating and delivering value to its customers, with the aim of capturing profit from it. It represents how different variables, in terms of strategy, architecture, and economics, are set in order to guide decisions and create a sustainable competitive advantage for the company in its domain market. It should give a clear explanation about who are the customers of the company, and what are the needs and demands of those customers, as well as how does the company deliver value to its customers, and how the profit is gained back. Therefore, it is a process of creating value for companies as well as customers and society. (Osterwalder *et al.*, 2010; Morris *et al.*, 2005; Teece, 2010)

The business model concept emerged as a known framework during the middle of the 1990s (Zott *et al.*, 2011). Prior to this period, business models did not involve any other activities except selling products and earning money proportionally to the number of products being sold. However, the development of technologies and industries, as well as the increase of Internet popularity bring advancements and complexity into businesses. These advancements introduce new ways of bringing value to the customers for companies, while individuals and businesses have a possibility of accessing data and information in an easy way. It became easier for customers to investigate a product, inform themselves about it and compare it with those from the competition. Therefore, committing time and effort into creating a good business model is a necessary step for companies that are aiming to succeed. Nevertheless, once the business starts following the planned business model, it becomes easy for competitors to imitate it, and good business models usually get its copies in a few years or even months. Consequently, the business model becomes "shared" within competitors. (Teece, 2010)

In order to provide a deep insight into the nature of business models, a set of guidelines to be followed on the way of systematically defining, designing and implementing a new or analyzing current business models was proposed by Osterwalder *et al.* (2010). The set of guidelines is called the *business model canvas*. The proposed canvas consists of nine basic building blocks, that show the logic of how a company intends to make money and capture value. Those are (Osterwalder *et al.*, 2010):

Aspect	Description
Customer segments	This segment consists of different groups of people or organizations, who have common needs, demands, behaviors, and that company is trying to reach, serve and satisfy. Customer segments part is the main part of any business model because the whole process of creating value is directed into this segment.
Value Propositions	It includes products and/or services that the company offers to its customers. Company has to discover what are the difficulties or demands targeted customers have, and solve those difficulties or fulfill those demands by proposing a value.
Channels	These are the ways of reaching customers, which include promoting the business, offering an efficient way of making purchases, delivering the value, offering a system for evaluation and post-purchase support.
Customer Relationships	Every company maintains relationships with its customers. The relationships can be of the same type for all customers, or different for different groups of customers. The type of relationship can vary from completely personal communication and assistance to a completely automated one.
Revenue Streams	This segment describes the customer's willingness to pay for provided value, and how the company is handling the payments in order to contribute to its revenues. There are two different types of Revenue Streams that a business model can involve. The first one is the one resulting from one-time customer payments. The second one is the one from ongoing payments of the regular delivery of a Value Proposition to the customer or of post-purchase support.
Key Resources	In order to achieve all previously described segments, the company has to have a number of resources available. Types of resources differ among different business models and can be physical, financial, intellectual, or human.
Key Activities	These are the most important operations that the company has to perform in order to create value, maintain relationships with customers and profit throughout the business. It can be categorized as the production of products, solving problems or providing a networking platform.
Key Partnerships	It includes other key organizations who are either suppliers or partners to the company, and who are representing important support to the company's business model. This is a strategic decision that can happen between the company and supplier organizations, its non-competitors, or even with its competitors (so-called "coopetition").
Cost Structure	Conducting the whole business model with its segments leads to certain costs for the company. Those can be calculated by defining Key Resources, Key Activities, and Key Partnerships, and it is a company's decision how much will it be guided by those costs; the company can decide whether to be cost-driven or value-driven.

Table 8 – Business model canvas (own presentation based on Osterwalder et al. (2010))

2.3.2 Data-driven business models

In order to provide a better understanding of why the popularity of data marketplaces continuously increases and why they are necessary for companies, a general overview of data-driven type of businesses will be presented in this section. Furthermore, the reasons for an organization to support its business model with data, and proposed guidelines that should be taken in order to incorporate data-supported decisions into the business, are presented as well.

A data-driven business model is a business model of a company which uses data as a key resource. This includes companies that collect, aggregate or analyze data, in order to make a profit of that data or information, either by selling it or by selling any other product or service that relies on that data as a key resource. A data-driven business model can have assets other than data as key resources as well. Utilization of data as a key resource of the business model means learning about customers, business flaws and business opportunities. It is a step towards improved business. (Hartmann *et al.*, 2016)

Source of data for a data-driven business model can be internal or external. Internal sources are those that can be obtained from the company's internal IT systems, self-generated sensor data or crowdsourced data. When it comes to self-generated data, some companies gather only the data that they are aware they need, and for which they have the exact way of utilizing. The other group of companies measures and collects needed data, as well as data about everything they are able to collect, with the goal and hope that they can extract additional useful information from it, and solve undiscovered problems. External data includes datasets that can be purchased from external data providers (data marketplaces) or obtained from public resources that do not require any costs. Furthermore, data, from either external or internal (or combined) sources, is being cleaned, aggregated, processed and analyzed with the aim of gathering valuable and useful knowledge about the business (See 2.1.1). (Hartmann *et al.*, 2016; Spijker, 2014)

According to the research conducted on 179 large businesses, results have shown that data-driven businesses have 5-6% higher output and productivity than what would be expected if they would not be implementing a business model that is data-driven (Brynjolfsson *et al.*, 2011). Another research, done by Hartmann *et al.* (2016), based on 100 start-up companies, showed that the majority of companies use external data sources (73%), 11% use internal sources, while the rest of the companies (16%) combines data from internal and external sources. This information shows the importance of well-designed business models for data marketplaces, due to the need for external data among many companies (Hartmann *et al.*, 2016).

Companies implement data-driven business models in order to bring more value to their customers and create more benefits for both sides – better satisfaction on the customer's side and better profit on the provider's side. But, since this means that data is the key asset of a business, it is a common understanding that it should not be shared or exposed in any way to competitors, in order to protect competitive advantage over them. However, the behavior of restricting access to data by a dominant company, both in terms of scale and scope, can lead to unwanted harm to consumer welfare. Therefore, organizations that care to protect consumer rights, work on an assessment of dominant players in the data-driven economy. Examples of the organizations who actively act to protect involved parties from the misuse of data in the European Union are the Bundeskartellamt (German Federal Cartel Office) and the European Commission. (Kathuria, 2019; Moody and Walsh, 1999)

Due to emerging changes in the economy, influenced by data, conventional businesses see a necessity in developing new business models in order to stay competitive on the market. In other words, companies need to reinvent their business model, in order to extract, refine and capitalize on data. Such innovation is everything but uncomplicated, and require long-term preparation. Proposed guidelines for

reinventing existing business model into a data-driven business model, or creating a completely new one, involves answering the following six questions (Brownlow *et al.*, 2015):

1. What do we want to achieve by using data?
2. What is our desired offering?
3. What data do we require and how are we going to acquire it?
4. In what ways are we going to process and apply this data?
5. How are we going to monetize it?
6. What are the barriers to us accomplishing our goal?

Not only are conventional organizations in the position that they should incorporate data-supported decision making instead of intuition-supported, but they are forced to do that in order to remain competitive on the market. Data is not only the key raw material for data-driven innovations, but it is the impulse that triggered the innovations. However, despite the challenges of reinventing existing business models that companies face, data-utilization is improving at a constant pace. The improvement would not be possible if the utilized asset would not have the characteristics which data has. Data is easy to transport, to share or copy, and it can be equally good used in any location or environment (as discussed in 2.1.2). Additionally, understanding and learning from data is satisfying and addictive, and it influences changes and evolution of the modern world. (Spijker, 2014; Brownlow *et al.*, 2015; Zikopoulos *et al.*, 2012)

2.4 CHAPTER SUMMARY

Chapter 2 gives a theoretical basis for conducting the practical analysis in order to answer the research question of this thesis. The amount of available data is constantly increasing, and its utilization is challenging. The volume, variety, and velocity make a precise amount of data infeasible to estimate. However, data is the ultimate basis for the advancement of the modern economy, and creating value from data is of high importance for businesses in order to stay competitive. A data-driven business creates benefits to both customers and its profit – customers are more satisfied, while the profit within the business raises. The source of data for data-driven businesses can be internal or external. If a business does not have the possibility or resources for collecting the needed data within its inner processes, it has to reach out for external data. Therefore, the need for efficient collaboration between data-driven businesses and external data sources is increasing. Data marketplaces, a software-based data trading solution, emerge to respond to this need. They provide a platform infrastructure that enables interested buyers and sellers of data to discover each other and perform data trading throughout pricing strategies. Data marketplaces environment is challenging and in the need for improvements. Those improvements include fairness, transparency, privacy, safeness and security, regulations, liquidity and efficiency of the data trading system. Additionally, data platforms are susceptible to the "winner-take-all" phenomenon. This could make the first players, who provide the most acceptable solution for a data marketplace and reach the market leadership, reach high economies of scale and set too high boundaries for other players. At the same time, the value of the data economy and the data market is increasing fast, and the estimations are that the increase will become even faster in the next years.

3 Methodology

To answer the research question of this thesis *("What are the characteristic elements of data market business models?")*, two major steps have been performed – taxonomy building and pattern matching.

Before diving into the processes, it was necessary to build a stable basis of knowledge about the topic. The preparation was done by an exhaustive literature review as well as the discovery of relevant objects for the analysis and resulted in an overview of the current market situation. It is important to emphasize that those two preparation steps have been executed in parallel. The reviewing of the literature was done with the aim of discovering relevant keywords that were further used for the search. Also, both the steps were performed with the aim to discover and understand the characteristics of data marketplaces and to identify the concrete definition of a data marketplace. Since the definitions of data marketplaces between different researchers do not always completely overlap, this thesis follows one specific definition (see section 2.2.2.2), that is used for identification of representative data marketplaces set for analysis. Therefore, the "relevant objects" in this phase of analysis meant the set of internet platforms and websites, which eventually could be classified into a data marketplaces definition, after a general overview of their web presence was done. Once the clear and final definition of data marketplace was discovered, the set of previously discovered objects was filtered in order to create a set of platforms which exactly fit the definition of a data marketplace, which makes them suitable for the analysis (see APPENDIX 2).

3.1 TAXONOMY BUILDING

Taxonomy building was performed by following the method suggested by Nickerson *et al.* (2013). This method provides guidelines for taxonomy building, that involve combining knowledge obtained from literature with the knowledge obtained by directly analyzing objects of interest. It is used to classify, clarify, understand and systematically analyze complex problems or domains. Taxonomy building in general well-accepted and a commonly used approach in the area of business models for Information Systems, like car-sharing business models (Remane *et al.*, 2016), digital business models (Bock and Wiener, 2017), FinTech business models (Eickhoff *et al.*, 2017) or business models of startups in the financial sector using blockchain (Beinke *et al.*, 2018). Taxonomies are also a frequently used approach in biology, social sciences and various other scientific disciplines (Nickerson *et al.*, 2013).

The terminology used for taxonomy is not unique among different researchers, but the one that is most commonly used to describe its elements is *dimension* and *characteristic*, which was adopted for this thesis as well. Other possibilities include *variable* and *values (domain)*, *character* and *character state*, *attribute* and *value*, *dimension* and *variable*, or *category* and *capability*. Taxonomy can often consist of dimensions that change over time, and there is never a guarantee that a complete set of dimensions or their characteristics has been discovered. Consequently, taxonomies are built to be as effective, useful and optimal as possible, but they should never be adopted as "best" or "correct" ones. Furthermore, useful taxonomy should contain a limited number of dimensions and their characteristics, but also enough of them to provide a clear separation of involved concepts. It should also be open for an extension when and if necessary afterward. (Nickerson *et al.*, 2013)

The suggested process of taxonomy building from Nickerson *et al.* (2013) consists of seven steps that lead to its development. The process is visualized in Figure 14 and explained in the following.

First, *meta-characteristics* have to be defined, which make a solid ground and limitation for the discovery of characteristics. From those meta-characteristics, characteristics are derived accordingly. Without this step, there is a risk of performing an unstructured analysis by generating too many important, as well as unimportant characteristics, with the hope that meaningful results will suddenly appear. (Nickerson *et al.*, 2013)

Furthermore, after defining meta-characteristics, *ending conditions* have to be established. This is a very important step because the whole process is of iterative nature, which requires clearly established termination constraints, which can be objective or/and subjective ones. (Nickerson *et al.*, 2013)

Third, the process gets to the junction point, where one of two paths are possible to take for the continuation of analysis – *conceptual-to-empirical* approach or *empirical-to-conceptual* approach. The former one is a method of discovering dimensions and their characteristics from relevant literature, without investigating the objects of interest individually. In other words, dimensions and characteristics are supposed to be discovered deductively. The latter one includes investigation of objects that are of interest for the whole classification, by observing and investigating them and identifying their characteristics. In other words, dimensions and characteristics are supposed to be discovered inductively. (Nickerson *et al.*, 2013)

The two approaches merge into the step of *checking the ending conditions* that were previously defined. If the ending conditions are not met, the process goes back to the junction point and continues with one of the two possible paths. Otherwise, when all ending conditions are met, the process terminates and taxonomy development is considered to be finished. (Nickerson *et al.*, 2013)

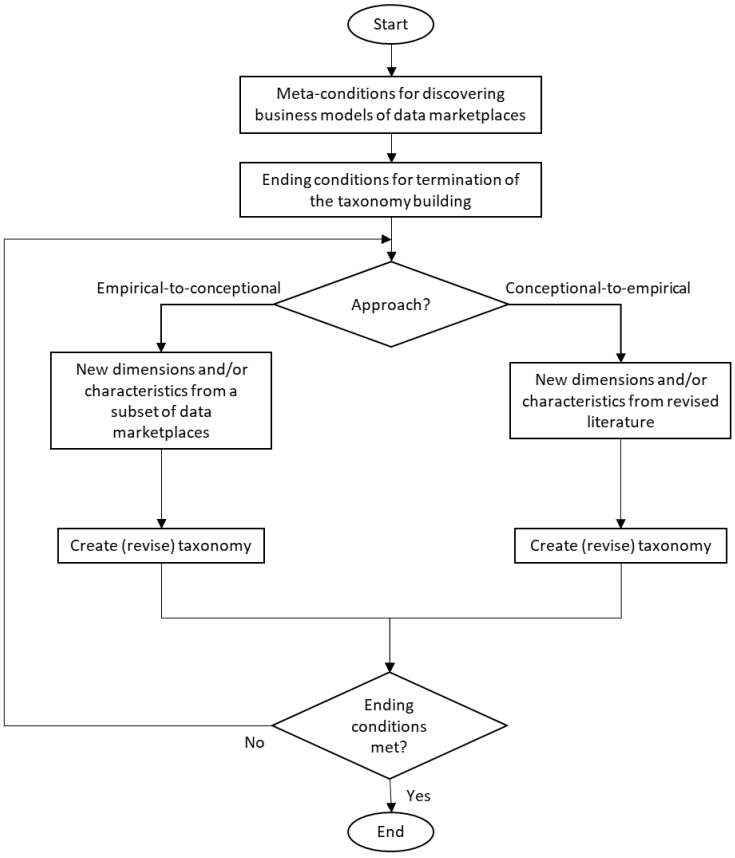

Figure 14 – Process of taxonomy development (own illustration based on Nickerson et al. (2013))

Figure 14 presents the explained steps used for the development of data platforms and data marketplaces taxonomy, by following the guidelines suggested by Nickerson *et al.* (2013). Each of the steps is adapted to fit the purpose of answering the research questions of this thesis, and in the following, they are described in more details.

Meta-characteristics

The goal of this thesis is to observe data platforms and data marketplaces from the business model perspective, and therefore to discover relevant characteristics of their business models. Defined meta-characteristics have to provide a strong and clear starting point for that purpose (Nickerson *et al.*, 2013). Therefore, meta-characteristics, in this case, are the commonly used framework for business model analysis: v*alue creation, value proposition, value delivery and value capture* (Remane *et al.*, 2017; Clauss, 2017; Osterwalder *et al.*, 2010). Consequently, every dimension and its characteristic discovered in the further process has to belong to one of those three components of the business model framework.

End conditions

Nickerson *et al.* (2013) proposed a set of eight objective ending conditions and four subjective ending conditions, that are also applicable for the need of this analysis, and all of them were followed exactly as proposed. In the following, all the ending conditions are listed:

1. *Objective ending conditions:*
    a. All objects from the representative sample were examined
    b. No object was merged or split in the last iteration
    c. Every characteristic of every dimension describes at least one object
    d. No new dimensions or characteristics were added in the last iteration
    e. No dimensions or characteristics were merged or split in the last iteration
    f. Every dimension is unique
    g. Every characteristic is unique within its dimension
    h. Each cell is unique and is not repeated

2. *Subjective ending conditions*
    a. Concise – taxonomy is meaningful but not overwhelming
    b. Robust – dimensions and characteristics differentiate sufficiently
    c. Comprehensive – all dimensions of interest identified and possible to reuse
    d. Extendible – easy to add new dimensions/characteristics
    e. Explanatory – dimensions/characteristics can describe an object from the domain

Select approach

The process of taxonomy development consisted of two conceptual-to-empirical iterations, followed by five empirical-to-conceptual iterations. This approach was chosen in order to first extract all relevant dimensions and their characteristics for business models of data marketplaces, that were already recognized by other researchers who have dealt with similar challenges, and to further enrich those dimensions and characteristics, by investigating data marketplaces directly. The activities in both approaches are described in the following.

Conceptual-to-empirical approach: In the first two iterations, relevant characteristics from existing literature were integrated. The literature included previous works, researches, and surveys on data marketplaces (Fricker and Maksimov, 2017; Koutroumpis *et al.*, 2017; Muschalle *et al.*, 2012; Stahl *et al.*,

2016) as well as electronic marketplaces and platforms business models in general (Täuscher, 2016; Täuscher and Laudien, 2017).

Empirical-to-conceptual approach: The following five iterations involved the classification of business models of data marketplaces from empirical data. To systematically identify the relevant objects for the inductive iterations, a rigorous procedure to identify relevant publications for a literature review suggested by Brocke *et al.* (2009) and adopted by Szopinski *et al.* (2019) was used. The steps of this process are explained in the following.

a. Tool search:

The most widely used search engine *Google* was used for the searching process. The engine was set to incognito mode in order to avoid the mutual influence of results and the possible effects from previous searches. The results of the search, which have been examined as a potential fit for the analysis set, were those yielded by Google in the first three pages of the search engine.

b. Identification of keywords:

The applied keywords for the search have been as following: “data marketplace”, “data market”, “data trading platform”, “data platform”, “buying data” and “data brokers”. Moreover, data marketplaces already surveyed in the inspected literature in the background section were included in the data collection for the analysis. However, the keywords like “database”, or “repository” are not included in the search, following the same argument like Fricker and Maksimov (2017) followed in their research on data marketplaces, that the interest of this research is considering businesses with data, and not data warehousing or businesses with data warehousing.

c. Selection of data sources:

The information that was extracted and used for the analysis from empirical point of view, was the information found in sources that included company websites, their respective data trading platform, white paper (if it was available for download), and enriched with the information found on other informative websites that provided articles mentioning investigated marketplace (if one or more existed). Those companies, whose information were insufficient, were not included in the data collection for the analysis. If available, possible and needed, an account for investigated data marketplace was created, with the goal of finding out if it is possible to observe its functionalities and offerings in more details, and consequently extracting those details for the analysis purposes.

d. Inclusion criteria:

The collection of data for the analysis included the companies whose physical location was anywhere in the world, without exception. However, independent from the company’s origin, only those companies whose online presence has been available in the English language at the moment of analysis, have been considered for including into the set for analysis.

e. Exclusion criteria:

To create a representative set of platforms for the analysis, the initial set of discovered objects have been filtered in two iterations. Every iteration included a specific set of criteria for exclusion used for the filtering, which is listed in the following:

1. Criteria for exclusion prior to taxonomy development:

    a. The object must exactly fulfill the definition of a data marketplace that this research follows (see section 2.2.2.2)

2. Criteria for exclusion during the taxonomy development:

    a. There is no sufficient public information available about the data marketplace - marketplace operators (employees or owners) have not been contacted in any way (e.g. direct emails exchange or phone call) for further information, and subscription or buying of datasets on the market with the goal of examining the process in more details was omitted due to the financial limitations (while some datasets that were offered for free have been downloaded for the sake of getting more information from it).
    b. Technical failures of the data marketplace preventing information retrieval (e.g. not possible to create an account), or location issues (e.g. unsupported area)
    c. The platform exists, but it is under a construction or in a testing phase

The complete set of the platforms which have been excluded during the two filtering iterations is presented in APPENDIX 2, together with the exact reason for not including it into the taxonomy development process.

Checking end conditions

After every iteration, end conditions have been checked, in order to determine if at least one more iteration is necessary. Once every end condition was met, the taxonomy development process was stopped and the generated set of dimensions and their characteristics was considered to be final and useful output that was further used as an input for the second major step – pattern matching.

3.2 PATTERN-MATCHING

The last step of the research involved using the developed taxonomy and analysis results for the identification of possible patterns within analyzed data marketplaces. The identification of patterns was performed following the guidelines suggested by Yin (2009). It involved recognizing similarities and dissimilarities within the cases, which progressed into the separation of groups between them, with mutually similar characteristics. As suggested from the guidelines, the number and variety of cases for this analysis contributed to the higher validity of the results. Consequently, the analyzed data marketplaces have been grouped according to their characteristics, and meaningful archetypes of data marketplaces have been identified. This process was guided by adopting the typology development guidelines provided by (Doty and Glick, 1994).

4 RESULTS

4.1 TAXONOMY AFTER CONCEPTUAL-TO-EMPIRICAL APPROACH

After the first two iterations of building the taxonomy, conceptual-to-empirical approach, a set of eleven dimensions with their respective characteristics have been acquired. Table 9 gives an overview of the results of this identification. Due to the lack of information found in the revised literature, at this point, it was not clear if all of the dimensions consist of mutually exclusive and collectively exhaustive characteristics, which would break the core condition of taxonomy. Each dimension with its characteristics discovered through this iteration is explained in the following. Sources of information for every dimension are specified in the separate column of Table 9. Additionally, all of the identified dimensions were aggregated according to the value proposition, value delivery, value creation, and value capture aspects of business models, which have also been used as meta-characteristics set prior to taxonomy development.

1. *Platform infrastructure*

As elaborated in more detail in 2.2.3.3, data platform can have different infrastructure designs – *Centralized* or *Decentralized*. *Centralized* design is a type of infrastructure that is conventional, and includes storage of datasets in one predefined place once the datasets are offered for trading, and accessing it from this storage once they have been traded. *Decentralized* infrastructure includes an emerging blockchain technology for storage of datasets in a decentralized manner with the aim of improving privacy, safety, and provenance of data and its traders. (Koutroumpis *et al.*, 2017)

2. *Data origin*

In order to trade data, suppliers have to get it from some type of source. This dimension includes the following categories: *Internet, Self-generated, User, Community, Governments, Authority*. If data is collected from online sources, either manually or automatically, and sold either aggregated or processed, this data belongs to the *Internet* data source type. *Self-generated* is the type of source that includes data generation either from private services and analysis methods or to those gathered from other data sources that are not included in the rest of the listed types of data sources. Moreover, *User* characteristic includes data that is collected directly from user inputs, who did that by using services that provide them with an output, and it is always mapped to some level of personal information about the user. *Community* as a data source characterize data that is collected from marketplaces or crowdsourcing services, or when users can edit the data. Furthermore, a data source categorized as *Governments* is the one collected from any highly trustable sources – government agencies or ministries. Last but not least, the *Authority* includes data provided by an expert or an institution. (Stahl *et al.*, 2017)

3. *Review system*

One key factor that influences the efficiency of transactions on digital marketplaces is the amount of trust between participants. Consequently, platform providers have to offer a mechanism to enhance trust, and for many platforms. This is done through a review system. Reviews can be exchanged between buyers and sellers – *Reviews by users*, or from the marketplace provider – *Reviews by marketplace*. Data

marketplace that does not provide the possibility of reviews belongs to characteristic *None*. (Täuscher, 2016; Täuscher and Laudien, 2017)

4. *Time relevancy*

This dimension refers to a period of time data is valid, and the frequency of updates needed to keep the validity. There are two distinguished types of datasets for this dimension: *Static* – data consists of facts that are not subject to frequent changes, and *Dynamic* – in order to keep it valid, datasets need regular updates. (Stahl *et al.*, 2017)

5. *Type of access*

A method of uploading and/or downloading datasets on the platform can be distinguished into several types. One of those is *API*, which is a predefined communication definition and protocol and can be accessed by attaching a software component to it, that is developed for the purpose of interacting with this interface (Souza *et al.*, 2004). *Download*, unlike APIs, does not require specialized software component, but data is rather accessed in the form of a downloadable file. Furthermore, some data marketplaces offer their own *Specialized Software* which should be used for examination, analysis, or visualization of data. Another possible dimension is *Web Interface*, which is a type of data access through a Web browser directly, where data is shown and can be explored and used. (Stahl *et al.*, 2017)

6. *Data output type*

As explained previously in 2.2.2, data, manipulated on data platforms, has one of the available machine-readable formats: *XML*, *CSV/XLS*, *JSON*, *RDF*, *Report*. *XML* and *JSON* are data formats that are used for semi-structured data, while *RDF* is often used in Semantic Web. Furthermore, *CSV/XLS* format is used by standard spreadsheet software tools for tabular data. Finally, *Report* as a characteristic includes all visualized data formats like PDF, DOC or JPEG. (Stahl *et al.*, 2014)

7. *Pre-purchase testability*

These characteristics distinguish whether the customer is allowed to access datasets before paying for them – *Complete access*, only part of it, enough to get an idea what to expect – *Restricted access*, or not at all – *None*. (Stahl *et al.*, 2014)

8. *Domain*

Data marketplaces can be open to any kind of datasets, that are covering any domain users are interested in selling or buying. In that case, this data marketplace attribute fits characteristic *All*. However, some data marketplaces are focusing only on one or a few data domains, and with this approach targeting only a specific niche of customers. If datasets provided on the marketplace consist of data about stock markets, company developments, product information like pricing, or about specific economic sectors, this marketplace domain belongs to the *Finance/Economy* group. *Scientific* data includes data that describes environmental, pharmaceutical, medical, or scientific work or research. Moreover, *Social Media* data domain includes posts, tweets, opinions, and trends collected from social media networks. The *Geo* data domain includes data about maps, landscapes, geographical position of business or individuals expressed

in coordinates. Furthermore, data that is in form of lists including addresses, emails, or customer information, belongs to the *Address* data group. Finally, if data is generated by using sensors from Internet of Things devices, or if it acts like a basis for building systems for smart decision-making and control, it belongs to the *Sensor* data group. (Fricker and Maksimov, 2017; Stahl *et al.*, 2014)

9. Marketplace participants

Platform providers can choose an aim for users of their business model, whether it is directed to private individuals or businesses. Therefore, data marketplace can decide for business-to-customer (B2C), business-to-business (B2B), or customer-to-customer (C2C) orientation. (Täuscher and Laudien, 2017)

10. Pricing model

This topic is elaborated in details in 2.2.3.1, and the same classification is used as possible characteristics of data marketplace attributes *Free, Usage-based, Package pricing, The flat fee tariff, Two-part tariff, Freemium*. (Muschalle *et al.*, 2012; Stahl *et al.*, 2014)

11. Price discovery

This dimension describes how the price for datasets or subscription is determined prior to the transaction. Prices can be set from one side of the market participants (*Set by sellers* or *Set by users*), and the other side can accept or decline it. Furthermore, prices can be fixed by marketplace provider, and in this case, both buyers and sellers have to agree with it before starting any kind of transactions – *Fixed prices*. The remaining two options are *Auction* – buyers or sellers are bidding against each other, and the other side chooses the most acceptable price, and *Negotiation* – a communication between buyer and seller in order to agree on the most acceptable price for both parties. (Täuscher and Laudien, 2017)

	Dimension	Characteristics						Source
Value Creation	**Platform infrastructure**	Centralized			Decentralized			Koutroumpis *et al.* (2017)
	Data origin	Internet	Self-generated	User	Community	Government	Authority	Stahl *et al.* (2017)
	Review system	User reviews		Review by marketplace		None		Täuscher (2016), Täuscher and Laudien (2017)
Value Proposition	**Time relevancy**	Static			Dynamic			Stahl *et al.* (2017)
	Type of access	API		Download	Specialized software		Web Interface	Stahl *et al.* (2017)
	Data output type	XML	CSV/XLS	JSON	RDF		Report	Stahl *et al.* (2014)
	Pre-purchase testability	None		Restricted access		Complete access		Stahl *et al.* (2014)
Value Delivery	**Domain**	All	Finance/ Economy	Scientific	Social media	Geo	Address	Sensor	Fricker and Maksimov (2017), Stahl *et al.* (2014)
	Marketplace participants	C2C		B2C		B2B		Täuscher and Laudien (2017)
Value Capture	**Pricing model**	Free	Usage-based	Package pricing	The flat fee tariff	Two-part tariff	Freemium	Muschalle *et al.* (2012), Stahl *et al.* (2014)
	Price discovery	Fixed prices	Set by sellers	Set by buyers	Auction	Negotiation		(Täuscher and Laudien, 2017)

Table 9 - Deductively acquired set of key business model attributes of data marketplaces (own presentation)

4.2 REVISED TAXONOMY AFTER EMPIRICAL-TO-CONCEPTUAL APPROACH

Furthermore, the following five iterations, that were done by empirical-to-conceptual approach, resulted in a set of characteristics that expanded the original set, as well as some changes of the existing dimensions and their characteristics. Table 10 presents the changes that have been introduced after the iterations – the characteristics that are excluded from the taxonomy (cells colored dark gray) as well as dimensions and characteristics that were added through the inductive approach (cells colored light gray). In the following, all the changes that have been done in comparison to the first version of taxonomy generated inductively (see Table 9) will be presented with the explanation of the reason for their alteration.

In order to satisfy the first key requirement of taxonomy building – taxonomy should be "collectively exhaustive", the following changes were performed:

- Due to the lack of accessibility to some information about analyzed marketplaces, some of the dimensions (*Review system, Type of access, Data output type, Pricing model*) were extended with *No info* characteristic.

Similarly, the "mutually exclusive" restriction is satisfied with the following changes:

- Dimension *Type of access* was extended with *API/Download* characteristic because there was an occurrence of marketplaces which provide a choice for customers between *API* access and simple download of datasets.

- *Government* as characteristic of *Data origin* dimension has been excluded since there was no marketplace that provided that kind of datasets for sale.

- None of the analyzed marketplaces have a possibility of exploring data through a web interface, which was a reason for exclusion of the *Web interface* characteristic from the *Type of access* dimension.

- Dimension *Data output type* was extended with *Multiple options* characteristic, while *XML* and *RDF* characteristics have been excluded from this dimension. The important is to note that *XML* and *RDF* did occur in some of the analyzed marketplaces, but those marketplaces offer other output types as well. Therefore, *Multiple options* characteristic consists of *XML* and *RDF*, among other characteristics.

- *Marketplace participants* dimension was extended with *Any* characteristic, while *C2C* and *B2C* characteristics were excluded. Characteristic *Any* can include *C2C* and/or *B2C* characteristic, among other characteristics.

- *Time relevancy* dimension was extended with *Both* characteristics, since some of the cases did indeed provide both possibilities, depending on the dataset that is being offered.

- None of the analyzed marketplaces provided the possibility of complete access to datasets before buying them. Therefore, the *Complete access* characteristic was excluded from *Pre-purchase testability* dimension.

- In the case of *Domain* dimension, there was no marketplace that provided data domain that would fit under *Scientific* characteristic.

- *Social media* characteristic was replaced with another, broader characteristic – *Personal*. This characteristic includes any data that is in general related to an individual and can consist of location, physical, physiological, genetic, mental, economic, cultural or social data (Opiria, 2019). Therefore, it implicitly includes social media data, which was the reason for previously mentioned exclusion.

- When it comes to *Pricing model* dimension, *Free* and *Two-part tariff* characteristics are excluded due to their lack of occurrence in the analyzed set.

- Also, a minor change was done to the first characteristic of "Domain" dimension – "All" was changed into "All/Any", in order to achieve better descriptiveness of the characteristic, but without changing its meaning.

In the following, new dimensions and their characteristics, which have been added, will be explained:

- *Privacy:* Due to many privacy concerns among data marketplace stakeholders, this dimension was chosen to describe what kind of privacy protection does the marketplace claim to have. Therefore, the data traded on the marketplace can be *Anonymized, Encrypted* or *Both (Anonymized and Encrypted).* A marketplace was classified in one of these three characteristics only if it openly claimed to support them. Otherwise, the marketplace that did not have this information belongs into *No info* characteristic, which can mean two things – the marketplace does anonymize and/or encrypt the data, but the marketplace provider did not choose to present this information openly, or the marketplace does not anonymize and/or encrypt the data.

- *Payment currency:* Since operating with money is one aspect of data marketplaces, there were three possible options recognized during the analysis – marketplace can operate with standard money - *Fiat money* (Mankiw, 2014), with *Cryptocurrency* (Chohan, 2017), or provide users with the possibility of choice between the two – *Both*.

- *Data quality guarantee:* Some marketplaces openly claim that data traded on their platform has a guaranteed quality. This type of claim can influence buyers and sellers whether or not they are willing to do data trading on a certain platform. Therefore, the dimension *Data quality guarantee* is added with the characteristic *Yes*, in case marketplace claims to have it, or *No info*, which can mean that marketplace did not mention anything about it, or simply do not guarantee data quality at all.

- *Smart contract with blockchain:* This dimension describes if the marketplace operations are followed by smart contracting or not. More details about what smart contract is provided in 2.2.3.3.2.

- *Additional purchase support:* Creating additional value and enhancing purchasing experience can be done by providing a specific type of product or service that is not directly offered through the business. In the business model canvas (see 5.3.1), these specific products and services belong to Channels segment. In the case of data marketplaces, it can be specific datasets that are created and delivered after the customer requests it, or offered support by in-house developers for data analysis services. Therefore, the dimension can have characteristic *With additional costs* - additional services are available with extra payment, *Included in price* – additional services are possible with no extra payment or *None* – data marketplace does not provide any additional services.

Dimension	Characteristics							
Platform infrastructure	Centralized				Decentralized			
Data origin	Internet	Self-generated	User		Community	Government	Authority	
Review system	User reviews		Reviews by marketplace		None		No info	
Time relevancy	Static			Dynamic			Both	
Type of access	API	Download	Specialized software		Web interface	API/Download	No info	
Data output type	XML	CSV/XLS	JSON		RDF	Report	Multiple options	No info
Pre-purchase testability	None		Restricted access			Complete access		No info
Domain	All/Any	Finance/ Economy	Scientific	Social media	Geo	Address	Sensor	Personal
Marketplace participants	C2C		B2C		B2B		C2B	Any
Pricing model	Free	Usage-based	Package pricing		The flat fee tariff	Two-part tariff	Freemium	No info
Price discovery	Fixed prices		Set by sellers		Set by buyers		Auction	Negotiation
Privacy	Anonymized		Encrypted		Both		No info	
Payment currency	Crypto			Fiat			Both	
Data quality guarantee	Yes				No info			
Smart contract with blockchain	Yes				No			
Additional purchase support	With additional costs			Included in price			None	

Table 10 – Intermediate set of business model attributes of a data marketplace (own illustration)

Legend: excluded | added

The intermediate step of checking if all of the ending conditions have been met, was performed after every iteration. The results of the checks are demonstrated in Table 11. As shown, uniqueness of dimensions and uniqueness of all cells have been satisfied throughout the whole process. Also, there was no splitting of objects of analysis at any point of the analysis, while the platforms analyzed represent completely separated instances, and each of the companies implements their business model individually. Throughout the whole process, the subjective condition of extendibility of the taxonomy has been satisfied as well. After the second conceptual-to-empirical iteration, no dimensions or characteristics have been merged or split, and no new dimensions or characteristics were added to the taxonomy set, which was an indicator that literature sources have been exhausted and that it is potentially beneficiary to move on to the empirical-to-conceptual iteration. However, those two conditions became unmet again, once the set was growing and changing with information found by direct objects observations, as well as the condition that each characteristic is unique within its dimension. However, later in empirical-to-conceptual iterations, all the three conditions have been met. The subjective ending condition that considers if the dimensions and characteristics are explanatory, has not been met in the first iteration, unlike in all other following iterations. The rest of the subjective conditions, including robustness, comprehensiveness, and conciseness, have not been met until the end of the fifth iteration for the first two, and the last iteration for the third one, respectively. All of the objects for analysis have been considered at the fifth iteration. Last but not least, exclusion of the characteristics from the dimensions set, that did not describe at least one object from the set of objects for analysis, have been done in the seventh iteration, which was also the last iteration performed.

	Ending condition	**1st (c2e)**	**2nd (c2e)**	**3rd (e2c)**	**4th (e2c)**	**5th (e2c)**	**6th (e2c)**	**7th (e2c)**
Objective	All objects from the representative sample were examined					**x**	**x**	**x**
	No object was merged or split in the last iteration	**x**	**x**	**x**	**x**	**x**	**x**	**x**
	Every characteristic of every dimension describes at least one object							**x**
	No new dimensions or characteristics were added in the last iteration		**x**				**x**	**x**
	No dimensions or characteristics were merged or split in the last iteration		**x**					**x**
	Every dimension is unique	**x**	**x**	**x**	**x**	**x**	**x**	**x**
	Every characteristic is unique within its dimension	**x**	**x**			**x**	**x**	**x**
	Each cell is unique and is not repeated	**x**	**x**	**x**	**x**	**x**	**x**	**x**
Subjective	Concise – taxonomy is meaningful but not overwhelming							**x**
	Robust – dimensions and characteristics differentiate sufficiently					**x**	**x**	**x**
	Comprehensive – all dimensions of interest identified and possible to reuse					**x**	**x**	**x**
	Extendible – easy to add new dimensions/characteristics	**x**	**x**	**x**	**x**	**x**	**x**	**x**
	Explanatory – dimensions/characteristics can describe an object from the domain		**x**	**x**	**x**	**x**	**x**	**x**

Table 11 – Ending conditions checked after each iteration (own presentation based on Nickerson et al. (2013))

The complete overview of dimensions with their characteristics, including the short description for every characteristic, indicator whether the characteristic was identified conceptually or empirically, and the source of knowledge in case of conceptual characteristics, is provided in APPENDIX 3.

The complete set of dimensions and characteristics, that resulted after all of the iterations of taxonomy building have been done, and all of the ending conditions have been met, is shown in Table 12, which includes only the dimensions and characteristics that are part of the final taxonomy. Also, included dimensions of the taxonomy have been reorganized into one of the meta-characteristics defined prior to the start of the analysis – *Value Creation*, *Value Proposition*, *Value Delivery*, and *Value Capture*. The resulting group of dimensions that belongs to *Value Creation* set, consists of three dimensions, which consist of eleven characteristics altogether. Furthermore, *Value Proposition* set consists of six dimensions, while twenty-two characteristics belong to it. Four dimensions belong to *Value Delivery* set and consist of fourteen characteristics. Finally, *Value Capture* has three discovered dimensions, with a total of thirteen characteristics belonging to it. Therefore, the resulting set consists of sixteen dimensions and sixty characteristics altogether.

	Dimension	Characteristics					
Value Creation	**Platform infrastructure**	Centralized			Decentralized		
	Data origin	Internet	Self-generated	User	Community	Authority	
	Review system	User reviews	Reviews by marketplace	None	No info		
Value Proposition	**Privacy**	Anonymized	Encrypted	Both	No info		
	Data quality guarantee	Yes		No info			
	Time relevancy	Static	Dynamic	Both			
	Type of access	API	Download	Specialized software	API/Download	No info	
	Data output type	CSV/XLS	JSON	Report	Multiple options	No info	
	Pre-purchase testability	None	Restricted access	No info			
Value Delivery	**Additional purchase support**	With additional costs	Included in price	No			
	Domain	All/Any	Finance/ Economy	Geo	Address	Sensor	Personal
	Marketplace participants	B2B	C2B	Any			
	Smart contract with blockchain	Yes	No				
Value Capture	**Pricing model**	Usage-based	Package pricing	The flat fee tariff	Freemium	No info	
	Price discovery	Fixed prices	Set by sellers	Set by buyers	Auction	Negotiation	
	Payment currency	Crypto	Fiat	Both			

Table 12 – Final taxonomy for business model attributes of data marketplaces (own presentation)

4.3 FREQUENCY ANALYSIS

Classification of the set of twenty final data marketplaces was done based on developed taxonomy, which is shown in APPENDIX 4. Furthermore, frequency analysis was derived from the classification, as presented in Table 13. The results indicate that approximately two-thirds of investigated platforms operated in a centralized manner while only one third operated decentralized. Self-generated data from private sources represents the most prominent origin of data. More than half of the investigated data marketplaces offer additional purchase support, however, predominantly for an extra charge.

Eighteen of the twenty data marketplaces either do not have a review system or offer no information on a review system. Similarly, half of the investigated data marketplaces did not offer any way of protecting privacy, while the other half guarantees that data is being encrypted or anonymized or both, within their marketplace. However, with fourteen of the twenty data marketplaces, a majority guaranteed the quality of their data. Also, seventeen of the investigated twenty data marketplaces offered both regularly updated datasets in combination with static datasets or exclusively regularly updated datasets. Fifteen of the twenty data marketplaces relied on APIs or downloads to offer access to data. Data from the investigated twenty data marketplaces did not reveal a concrete focus regarding the domain of offered datasets. Also, while almost half of the data marketplaces laid on a B2B focus or no focus at all, with three of the twenty data marketplaces only a minority used a C2B model.

With nine out of twenty, almost half of the investigated data marketplaces offered smart contracts with the blockchain while only six offered payments via cryptocurrency. Fiat money was the most prominent payment currency. Only one data marketplace offered both payments with cryptocurrency and fiat money. The most prominent pricing models were usage-based models (used by seven out of twenty data marketplaces), a flat fee tariff (used by five data marketplaces) and the Freemium-model (used by four out of twenty data marketplaces). Further, eleven out of twenty data marketplaces relied on fixed prices while six data marketplaces relied on prices set by sellers.

Dimension	Characteristics				
Platform infrastructure	Centralized (65%)			Decentralized (35%)	
Data origin	Internet (5%)	Self-generated (50%)	User (15%)	Community (10%)	Authority (20%)
Review system	User reviews (10%)	Reviews by marketplace (10%)	None (45%)	No info (35%)	
Privacy	Anonymized (30%)	Encrypted (10%)	Both (10%)	No info (50%)	
Data quality guarantee	Yes (70%)		No info (30%)		
Time relevancy	Static (15%)		Dynamic (55%)	Both (30%)	
Type of access	API (35%)	Download (20%)	Specialized software (15%)	API/Download (20%)	No info (10%)
Data output type	CSV/XLS (30%)	JSON (20%)	Report (5%)	Multiple options (20%)	No info (25%)
Pre-purchase testability	None (60%)		Restricted access (35%)	No info (5%)	
Additional purchase support	With additional costs (40%)		Included in price (15%)	No (45%)	
Domain	All/Any (25%)	Finance/ Economy (10%)	Geo (10%)	Address (10%)	Sensor (20%)	Personal (25%)
Marketplace participants	B2B (45%)		C2B (15%)	Any (40%)	
Smart contract with blockchain	Yes (40%)		No (60%)		
Pricing model	Usage-based (35%)	Package pricing (15%)	The flat fee tariff (25%)	Freemium (20%)	No info (5%)
Price discovery	Fixed prices (55%)	Set by sellers (30%)	Set by buyers (5%)	Auction (5%)	Negotiation (5%)
Payment currency	Crypto (30%)		Fiat (65%)	Both (5%)	

Table 13 – Frequency analysis for business model attributes of data marketplaces (own presentation)

Legend: **1-25%** | **26-50%** | **51-75%** | **>76%**

4.4 ARCHETYPES

Following the recognized patterns, data marketplaces have been derived into four groups - archetypes. Every archetype is named with self-explainable name, which is demonstrated in Table 14, together with its representative data marketplace and its main characteristics. Each of the archetypes is in more details described in the following.

1. **Centralized data trading:** This archetype represents data marketplaces that show very similar characteristics to traditional marketplaces, with the goal of making data trading as simple and as quick as possible. Eleven marketplaces, out of twenty, belongs to this archetype, and all of them have a centralized infrastructure. Eight of them do not provide any information about privacy protection, which can be seen as a disadvantage of centralized infrastructure. None of the included data marketplaces provide privacy protection with a smart contract, and all of them accept only fiat money as currency for transactions, whose prices are mostly fixed by the marketplace in advance (seven out of eleven). These types of data marketplaces can take any characteristic of data origin, data domain, or data output type, as well as the pricing model.

2. **Centralized data trading with smart contract:** Only one marketplace from the analysis set fit to this archetype, but even though it also has centralized infrastructure, there are two dimensions that make it stand out – encryption of data and smart contact with blockchain. This specific case of data marketplace makes trading easy and understandable – due to its centralized characteristic but tend to solve security and legal issues of standard centralized data marketplaces.

3. **Decentralized data trading:** This archetype implements decentralized infrastructure, which is by default followed by smart contracting. Five out of eleven data marketplaces fit into this category. These marketplaces sell self-generated, dynamic data and none of them provide any additional services. Due to the smart contract, all marketplaces guarantee data quality. Four out of five of the cases accept only the cryptocurrency, while only one data marketplace from the set provides the possibility of choice for users – crypto or fiat.

4. **Personal data trading:** Three out of twenty marketplaces stood out with their common mission – to bring personal data trading control back to its owners. This type of platform makes it possible for anyone to make available his/her data for trading, in exchange for money. Therefore, this archetype has a consumer-to-business characteristic, it operates with user-generated personal data, and the trading is done through specialized software, with the assumption that it is the easiest and the most user-friendly solution to make usage of the marketplace possible for anyone. These data marketplaces propagate the message that personal data is continuously being stolen from individuals without their knowledge, by giant companies, which thrive from collected data through their data-driven businesses, without providing anything valuable in return. Therefore, the data marketplaces which belong to personal data trading archetype, present themselves as the way of taking back the control over personal data to their owners. Any user who wants to join the network signs up into the data marketplace, and through a specialized software shares his/her data, for a predefined amount of money in return.

4.5 ILLUSTRATIVE EXAMPLES OF THE DISCOVERED ARCHETYPES

Table 14 shows the representative data marketplaces for each discovered archetype (discussed in 4.4). Data marketplaces that were chosen as a representative case for the archetypes are Quandl, Dawex, IOTA and Datacoup. Also, Table 14 shows the main characteristics of every representative data marketplace, with the reason for the comparison between distinguished types of business models.

Data marketplace archetype	*Centralized data trading*	*Centralized data trading with smart contract*	*Decentralized data trading*	*Personal data trading*
Representative platform	Quandl	Dawex	IOTA	Datacoup
Value creation	- Centralized - Anonymized	- Centralized - Encrypted	- Decentralized - Encrypted	- Decentralized - Anonymized
Value proposition	- Dynamic datasets	- Static and dynamic datasets	- Dynamic datasets	- Dynamic datasets
Value delivery	- API or download - Restricted access to data samples - B2B - No smart contract	- API or download - Restricted access to data samples - B2B - Smart contract	- API - No test data samples - B2B - Smart contract	- Specialized software - No test data samples - C2B - Smart contract
Value capture	- Freemium pricing - Prices set by sellers - Fiat currency	- Usage based pricing - Prices set by sellers - Fiat currency	- The flat fee pricing - Price set by sellers - Cryptocurrency	- Usage based pricing - Fixed prices - Cryptocurrency

Table 14 - Data marketplaces archetypes with representative case (own presentation)

5 Summary and discussion

This research has been performed by taking a perspective of business models to propose a taxonomy for data marketplaces. The dimensions and characteristics have been derived conceptually from sources including Fricker and Maksimov (2017), Koutroumpis *et al.* (2017), Muschalle *et al.* (2012), Stahl *et al.* (2017), Stahl *et al.* (2014), Täuscher (2016) Täuscher and Laudien (2017), and further enriched and extended with empirical material from a sample that includes twenty data marketplaces. APPENDIX 3 outlines all of the conceptual and empirical dimensions and characteristics used for the development of the taxonomy, including their short explanation and origin of the information in case of conceptual result.

With the goal of building a valid taxonomy, as recommended Nickerson *et al.* (2013), two main principles of taxonomy development have been strictly followed – all the elements of taxonomy have to be collectively exhaustive and mutually exclusive. Therefore, the process of building the taxonomy consisted of seven iterations, within which many changes occurred in order to satisfy the two principles as well as the ending conditions predefined prior to the taxonomy development. Consequently, only conceptual characteristics that did occur in the empirical material, have been included in the final result of the taxonomy. From the objective perspective, and having that all of the objective conditions have been satisfied at the end of the taxonomy development, the developed taxonomy consists of unique dimensions, that further consists of characteristics that are unique within its dimension. Similarly, each combination of dimensions is indeed unique within the developed taxonomy. Furthermore, all of the objects from a representative sample were examined and classified within the taxonomy. All of the subjective ending conditions are met, which implies that the developed taxonomy is concise, robust, comprehensive, extendible and explanatory.

5.1 DISCUSSION OF SOURCES FOR EMPIRICAL DATA

The information for empirical enrichment of conceptually identified dimensions and characteristics of the taxonomy, includes the information gathered from websites of companies, their respective data trading platform, white paper (if one was available for download), and enriched with the information found on other informative websites that provided articles mentioning investigated marketplaces (if one or more existed). Companies that provide a downloadable white paper, which explains the procedure of trading within their marketplace, intend to clarify the approach they are following and attract clients by explaining it in details. However, a website of the companies is the source of information about business that makes the first impression; companies intend to present themselves in the best light, while potentially omitting any information that could influence the opinion of potential clients visiting the website, who are in the process of deciding whether to use the services of the marketplace or not. Since some marketplaces have a separated internet domain, where their platform for trading is provided, or specialized software for the trading activities, those platforms provided this analysis with a deeper insight into the trading itself. Analysis of each data marketplace included investigation of available web articles that mention and describe the platform. This approach of gathering information provided a general impression about maturity and popularity of individual data marketplaces. However, the web articles that mention data marketplaces turned out to be the least resourceful for the purpose of providing additional information; the articles mostly present the same information which is already available on the website of the respective data marketplace.

5.2 DISCUSSION OF TAXONOMY

As expected, the separation between centralized and decentralized marketplaces is indeed noticeable within competitors. The former one provides a simple solution for trading, known to anyone who has ever done any transaction of goods by using an electronic marketplace, and easy to understand and handle. The latter seems to be more complex and eventually harder to understand for individuals who do not have high technical literacy. However, since a uniform design of decentralized marketplaces is in general not yet discovered (as discussed in section 2.2.3.3.2), understanding the complete underlying concept of one decentralized data marketplace, is not a guarantee of immediate understanding of some other considered decentralized data marketplace. The differences outnumber the similarities between them, which can be the reason for each of the marketplaces to separately and individually present their underlying idea. Anyhow, decentralized marketplaces within the sample of this analysis take approximately one-third of the whole set. That makes them an important player of the market despite their complexity, since the reasons for the complexity, in general, are privacy-preserving, security and increased trust (as discussed in section 2.2.3.3.2).

Data origin as a dimension does not play a major role in data marketplaces with centralized infrastructure and did not appear to be interdependent on any other characteristics. On the other hand, data marketplaces with decentralized infrastructure have either self-generated or user-generated origin of data. Those kinds of data origins imply that data is either collected directly from sources that produce them (and the infrastructure supports the automatic registration of those sources onto the platform), or the data is directly created by the user who is being aware of providing the data into the trading system.

It is challenging to make conclusions about the investigated data marketplaces which do not implement a reviewing system or those which do not provide any information about it. The minority (20%) of the investigated data marketplaces have a reviewing system. It is possible that the rest of the data marketplaces do not find it useful or helpful. However, since the blockchain technology provides a way of raising the trust towards the marketplace, decentralized marketplaces could be in less need for a complex reviewing, unlike the centralized data marketplaces.

Privacy-preserving issues within data marketplaces are highly discussed topic (see section 2.2.3). The consideration of privacy protection, either by anonymizing the data traded within the marketplace, encrypting the stored data or preferably both of the options, can be an influencing point for making the impression about the business model of a data marketplace. However, many data marketplaces from the analyzed sample do not offer a specific statement about the ways they are preserving privacy issues within the trades. It can be assumed that they do not offer any kind of privacy protection for the datasets of their users, or that they did not find necessary to specify this particular detail since they do not find the existence of these characteristics questionable.

Data quality guarantee as a dimension can be considered as weak, due to the fact that it is often used as a part of marketing representation of a company, but without a clear explanation what exactly are the measures the data marketplace implements, that support this claim. However, considered from the perspective of a potential client, it can be an influential factor onto the general impression about the marketplace and the decision about performing trades within it. Therefore, it is not a surprising outcome that the majority of the marketplaces (70%) claim that the data, that is being traded on their platform, is guaranteed of good quality.

Time relevancy dimension has been considered with the aim of understanding whether the data marketplaces that provide only static or only dynamic datasets, also follow some other distinguishable patterns. The only identified pattern is that all the decentralized data marketplaces from the sample, that provide data with the self-generated origin, provide only dynamical datasets. The reason for this could be

that self-generated data origin includes data collected from sensors and sources that are regularly updating their values.

Dimensions that consider the type of access to the datasets and type of data output have been expected to uncover interdependency between each other, but the analyzed sample did not show any significant pattern between them. However, due to the diversity of data access as well as data output type possibilities within the marketplaces, another conclusion can be derived – the access and output of data are not standardized. This fact can produce unnecessary complications. It can happen, for example, that one marketplace, that has the dataset buyer needs, offers only download as data access type, and only JSON output type, while the other marketplace has similar dataset, that user wants to combine with the first one, but it is only possible to access it by implementing specific API for that marketplace and to get the dataset in the form of report. The buyer needs to combine those two datasets into one that follows only one format, which is not impossible, but it can take additional time.

As discussed in section 2.2.3, due to the "Arrow's information paradox", it was not expected that data marketplaces are enthusiastic about offering access to datasets prior to purchase. Indeed, most of the data marketplaces (60%) from the analyzed sample do not offer any pre-purchase testability, while 35% marketplaces offer restricted access to datasets, usually by providing a representative sample.

Majority of data marketplaces claim to offer additional services (55%), in case the offer of datasets within the marketplace is not satisfying for client's need, if additional cleaning or processing of data is needed, a client wants to buy a combination of available sets as one single set, etc. From the information extracted from chosen sources for this analysis, it is not possible to identify all the services that could be offered additionally from data marketplaces. The reason for that is because a contacting of the responsible person from the company was needed, either by email or a phone call, in order to agree on the matter. However, this dimension describes the possibility of having benefits in addition to predefined ones within a data marketplace, which can be an additional source of profit for a data marketplace business.

Domain as a dimension has been considered in order to discover whether the nature of domain for datasets offered within the marketplace have any role in shaping the business model, as similar was done with the dimension that considers time relevancy of datasets. The only significant pattern that stood out from this dimension, is the trading of personal data, and the marketplaces who are hosting this type of trades have been distinguished from others in terms of their business model type.

45% of data marketplaces analyzed from the sample emphasize that their business model type, in terms of marketplace participants, supports exclusively B2B trading, while 15% have exclusively C2B as a target participants relation within their marketplaces. On the other side, 40% of them do not bother about specifying to whom their platform should serve. However, not specifying it does not necessarily mean that platform business targets anyone who wants to use the platform, but it could be under the assumption that the target group is self-recognizable.

The clear separation has been recognized between data marketplaces in terms of providing the smart contract with blockchain within the business. This dimension turned out to be closely connected with the platform infrastructure, but the two dimensions are not completely interdependent. All except one case followed the same combination of characteristics – centralized infrastructure paired with no smart contracting and decentralized infrastructure paired with smart contracting. However, the special case, that is a combination of centralized infrastructure and smart contracting (See APPENDIX 4) was the reason for this dimension to stand out as a separate dimension from the one that considers infrastructure.

The usage-based pricing model is the simplest solution for marketplaces, but in the case of data marketplaces, as discussed in section 2.2.3.1, it is not the best option. Consequently, it was expected that the analysis will confirm the empirical findings and that data marketplaces will offer pricing models other than usage-based pricing. However, the usage-based pricing took the lead in the pricing models

implemented within data marketplaces from the analyzed sample (see Table 13). As it is the simplest solution for the matter of pricing models, this could be the reason for data marketplace providers embracing it.

The developed taxonomy showed that prices of datasets on the marketplaces are usually fixed in advance by the marketplace providers. However, the investigation of sources showed that the determination of prices is not openly discussed by the marketplaces. In addition, selling data on the marketplace requires contact by email with the marketplace providers. The consequence of the limitations regarding the pricing of datasets is that it is not clear if the marketplace earns money, and how much it earns, by reselling data through the platform.

Only one platform from the analyzed sample allows its users to choose between trading data with fiat money or with crypto-currency (see Table 13). All the other platforms follow the exact same pattern – centralized infrastructure allows only fiat currency payments, while decentralized infrastructure means that the payments are possible only with cryptocurrency. This is an expected outcome. However, one data marketplace from the analyzed set made an exception, by allowing users to choose between the two currency types. This possibility makes the decentralized marketplace an acceptable solution for the users who do not have an interest in trading with crypto-currencies.

5.3 DISCUSSION OF ARCHETYPES

Four business model archetypes of data marketplaces were derived from the established taxonomy and from the final representative sample of twenty data marketplaces:

- Centralized data trading
- Centralized data trading with smart contract
- Decentralized data trading
- Personal data trading

The main criterion that separated data marketplaces was whether data marketplaces stored their data in a centralized or decentralized manner. In that regard, anonymity and encryption of data are major aspects of differentiating data marketplaces.

The *Centralized data trading* archetype does not significantly differ from conventional marketplaces for other goods, which is an outcome that was expected. It is indeed the most understandable and the easiest option for an individual or business eager to buy or sell data on a platform. However, data marketplaces that belong to this archetype do not provide information about privacy preserving. Centralized infrastructure is known for lack of security and privacy protection, and that is their biggest weakness (as discussed in section 2.2.3.3.1). This type of data marketplaces could face the alarming importance to reconsider their business model aspects and to introduce innovations for solving data governance and privacy issues, in order to survive the emerging environment.

Even though only one data marketplace from the analyzed sample belongs to the *Centralized data trading with smart contract* archetype, it stood out because of the characteristic that made it distinguishable from other data marketplaces with centralized infrastructure. This marketplace made the attempt of solving typical problems with security, privacy, and transparency the others struggle with. Due to the single example of such a data marketplace, it is possible that similar examples will arise in the near future. It is not feasible to discuss whether this kind of business model really decreases or eliminates the issues with security, privacy, and transparency. However, solving these issues is the possible reason why this data marketplace provider introduced smart contracting in the centralized trading system in the first place. It is

important to emphasize that this type of data marketplaces did not occur in literature revised prior to the analysis.

The *Decentralized data trading* archetype is also an example of a data marketplace business model that is already discussed among some other researches in the existing literature. As stated in the literature, there is no universal pattern that can be recognized among the samples. The common characteristics are, as expected, the elimination of central authority, cryptocurrency, and smart contracts. However, payment mechanisms, reviewing and quality control of data, and accessing mechanisms differ from platform to platform. It is certain that the complexity of such marketplaces is justified, and that data marketplaces will develop further in the direction of decentralization.

A special type of data marketplaces stood out as a separate archetype – *Personal data trading*. The distinct characteristic of these platforms is the idea they promote - that personal data gets stolen from individuals without their knowledge, by giant companies who do not provide anything valuable in return. This is the reason why data marketplaces which belong to personal data trading archetype, present themselves as the way of taking back the control over personal data to their owners. Any user who wants to join the network signs up into the data marketplace, and shares his/her data through specialized software, for a predefined amount of money in return. However, information gathered from sources different from those created by the marketplaces themselves, relived other types of concerns and discussion among potential users – that these data marketplaces are not reliable. Users are not able to retrieve the money they earned from selling data into a fiat currency, usually with the promise of the marketplace that the functionality is not yet implemented. Nevertheless, there are existing data marketplaces which implement personal data trading business model, and it is yet to discover whether they will survive or not.

5.4 LIMITATIONS AND FUTURE RESEARCH

The developed taxonomy contains dimensions and characteristics that provide a clear separation of involved concepts. However, this research is subject to several limitations.

First, limiting the investigation to data marketplaces with the English language holds a potential bias in data selection. This goes hand in hand with the low number of twenty data marketplaces in the final selection. Further research could address this issue by also investigating data marketplaces on a broader basis, for example through including languages other than English (e.g. Mandarin or Russian). In addition, increasing the number of investigated cases could open up the opportunity for further evaluation of the established taxonomy as well as a quantitative cluster analysis of data marketplaces.

Second, although strictly following the process of taxonomy building proposed by Nickerson *et al.* (2013), future research might uncover additional dimensions and characteristics of data marketplaces. This is especially the case since, it can be assumed that business models are subject to change, for example through interactions with their environment (Dellyana *et al.*, 2018). Further, due to the emerging nature of data marketplaces, it is clear that the set of marketplaces for this analysis just captured the moment, rather than created a stable set for long-term observation. It is to be expected that new players will join the market, while others will not be able to cope with challenges, even in the near future.

Third, the possibility of misinterpretation of empirical data cannot be ruled out. However, in order to minimize their probability, multiple iterations on the collected information on data marketplaces were performed.

Fourth, this research did not involve direct communication with marketplace providers (e.g. via emails or phone calls). There is a possibility that involving more types of contact would change an outcome, by at least lowering the amount of "No info" results within the analysis. However, it can be assumed that investigating only the public presentation of data marketplace providers could offer valuable insights into the design of business models for data marketplaces. Future research could address this issue through in-depth case studies of data marketplaces and their respective providers. Subsequently, the proposed system of dimensions and characteristics as well as the established taxonomy form a solid basis for future research on business models of data marketplaces.

The analysis is a possible subject to human error. This is due to the subjective understanding and interpretation of reviewed sources done by the author of this thesis, which involved investigating textual documents written by marketplace providers, that could be followed by misinterpretation of claims. Therefore, it is possible that the results of this thesis would be of higher value if the analysis would be done by more than one person. However, the assumption is that multiple iterations over documents and websites minimized the number of possible mistakes.

6 REFERENCES

Agarwal, A., Dahleh, M. and Sarkar, T. (2019), *A Marketplace for Data: An Algorithmic Solution*, In: *ACM EC '19: Proceedings of the 2019 ACM Conference on Economics and Computation*, Phoenix, AZ, USA, available at: http://arxiv.org/pdf/1805.08125v4.

Banerjee, P. and Ruj, S. (2019), *Blockchain Enabled Data Marketplace - Design and Challenges*, Cryptology and Security Research Unit, Kolkata, India, available at: http://arxiv.org/pdf/1811.11462v2.

Beinke, J., Teuteberg, F. and Nguyen Ngoc, D. (2018), "Towards a Business Model Taxonomy of Startups in the Finance Sector using Blockchain", paper presented at Thirty Ninth International Conference on Information Systems, San Francisco.

Bock, M. and Wiener, M. (2017), "Towards a Taxonomy of Digital Business Models - Conceptual Dimensions and Empirical Illustrations", paper presented at Thirty Eigh Thirty Eighth International Conference on Information Systems, Seoul, South Korea.

Brocke, J. vom, Simons, A., Niehaves, B., Riemer, K., Plattfaut, R. and Cleven, A. (2009), "Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process", paper presented at European Conference on Information Systems (ECIS), AIS Electronic Library (AISeL).

Brownlow, J., Zaki, M., Neely, A. and Urmetzer, F. (2015), *Data-Driven Business Models: A Blueprint for Innovation*, Working paper, available at: DOI: 10.13140/RG.2.1.2233.2320.

Brynjolfsson, E., Hitt, L.M. and Kim, H.H. (2011), "Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?", DOI: 10.2139/ssrn.1819486, *SSRN Electronic Journal*.

Carnelley, P., Schwenk, H., Cattaneo, G., Micheletti, G. and Osimo, D. (2016.), *Europe's Data Marketplaces: Current Status and Future Perspectives, SMART 2016/0063*, available at: http://datalandscape.eu/data-driven-stories/europe%E2%80%99s-data-marketplaces-%E2%80%93-current-status-and-future-perspectives (accessed 9 June 2019.).

Cattabei, G., Micheletti, G., Osimo, D. and Jakimowicz, K. (2019), *The European Data Market Study: Second Report on Facts and Figures*, Updating The European Data Market Monitoring Tool, *SMART 2016/0063*, available at: http://datalandscape.eu/sites/default/files/report/EDM_D2.4_2ndReport-FactsFigures_26032019.pdf (accessed 20 June 2019).

Chakrabarti, A., Quix, C., Geisler, S., Pullmann, J., Khromov, A. and Jarke, M. (2018), "Goal-Oriented Modelling of Relations and Dependencies in Data Marketplaces", paper presented at 11th International i* Workshop co-located with the 30th International Conference on Advanced Information Systems Engineering (CAiSE 2018), June 12, 2018, Tallinn, Estonia, available at: http://ceur-ws.org/Vol-2118/iStar2018_paper_4.pdf (accessed 30 June 2019.).

Chohan, U.W. (2017), "Cryptocurrencies: A Brief Thematic Review", Chohan, Usman W., Cryptocurrencies: A Brief Thematic Review (August 4, 2017). Available at SSRN: https://ssrn.com/abstract=3024330 or http://dx.doi.org/10.2139/ssrn.3024330, *SSRN Electronic Journal*.

Clauss, T. (2017), "Measuring business model innovation: conceptualization, scale development, and proof of performance", *R&D Management*, Vol. 47 No. 3, pp. 385–403.

Dao, D., Alistarh, D., Musat, C. and Zhang, C. (2018), *DataBright: Towards a Global Exchange for Decentralized Data Ownership and Trusted Computation*, ICLR 2018, available at: http://arxiv.org/pdf/1802.04780v1.

Deichmann, J., Heineke, K., Reinbacher, T. and Wee, D. (2016), "Creating a successful Internet of Things data marketplace", *McKinsey Digital*, October 2016, pp. 1–8, available at: https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/creating-a-successful-internet-of-things-data-marketplace (accessed 11 February 2019.).

Dekkers, M. (2013), "Asset Description Metadata Schema (ADMS)", available at: https://www.w3.org/TR/vocab-adms/ (accessed 14 March 2019.).

Dellyana, D., Simatupang, T. and Dhewanto, W. (2018), "Managing the actor's network, business model and business model innovation to increase value of the multidimensional value networks", *International Journal of Business and Society*, Vol. 19, pp. 209–218.

Domo, Inc. (2018), *Data Never Sleeps 6.0*, 6th ed., available at: https://www.domo.com/solution/data-never-sleeps-6?aid=DPR060518 (accessed 20-Feb-19).

Doty, D.H. and Glick, W.H. (1994), "Typologies as a Unique Form of Theory Building: Toward Improved Understanding and Modeling", *The Academy of Management Review*, Vol. 19 No. 2, p. 230.

Eickhoff, M., Muntermann, J. and Weinrich, T. (2017), "What do FinTechs actually do? A Taxonomy of FinTech Business Models", paper presented at International Conference on Information Systems 2017, South Korea.

Eisenmann, T., Parker, G. and van Alstyne, M.W. (2006), "Strategies for two-sided markets", *Harvard business review*, Vol. 84 No. 10, p. 92.

El Arass, M. and Souissi, N. (2018), "Data Lifecycle: From Big Data to SmartData", *2018 IEEE 5th International Congress on Information Science and Technology (CiSt)*, pp. 80–87.

Fricker, S.A. and Maksimov, Y.V. (2017), "Pricing of Data Products in Data Marketplaces", in Ojala, A., Holmström Olsson, H. and Werder, K. (Eds.), *Software Business*, *Lecture Notes in Business Information Processing*, Vol. 304, Springer International Publishing, Cham, pp. 49–66.

Gantz, J. and Reinsel, D. (2013), *The digital universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East*, IDC Digital Universe Study 2012, available at: https://www.emc.com/collateral/analyst-reports/idc-digital-universe-united-states.pdf (accessed 21 April 2019.).

Gawer, A. (2014), "Bridging differing perspectives on technological platforms: Toward an integrative framework", *Research Policy*, Vol. 43 No. 7, pp. 1239–1249.

Günther, W.A., Rezazade Mehrizi, M.H., Huysman, M. and Feldberg, F. (2017), "Debating big data: A literature review on realizing value from big data", *The Journal of Strategic Information Systems*, Vol. 26 No. 3, pp. 191–209.

Hartmann, P.M., Zaki, M., Feldmann, N. and Neely, A. (2016), "Capturing value from big data – a taxonomy of data-driven business models used by start-up firms", *International Journal of Operations & Production Management*, Vol. 36 No. 10, pp. 1382–1406.

Haupt, M. (2016), ""Data is the New Oil"- A Ludicrous Proposition. Natural resources, the question of ownership and the reality of Big Data", available at: https://medium.com/project-2030/data-is-the-new-oil-a-ludicrous-proposition-1d91bba4f294 (accessed 5 May 2019.).

Hendler, J. and Pardo, T.A. (2012), "A Primer on Machine Readability for Online Documents and Data", Data.gov, available at: https://www.data.gov/developers/blog/primer-machine-readability-online-documents-and-data (accessed 18 March 2019.).

Hynes, N., Dao, D., Yan, D., Cheng, R. and Song, D. (2018), "A demonstration of sterling. A Privacy-Preserving Data Marketplace", *Proceedings of the VLDB Endowment*, Vol. 11 No. 12, pp. 2086–2089.

Iansiti, M. and Lakhani, K.R. (2017), "The Truth About Blockchain", available at: https://hbr.org/2017/01/the-truth-about-blockchain (accessed 23 April 2019.).

International Data Corporation (IDC) and the Lisbon Council (2018), *European Data Market Study: First Report on Facts and Figures: Updating the European Data Market Monitoring Tool*, SMART 2016/0063.

Kathuria, V. (2019), "Greed for data and exclusionary conduct in data-driven markets", *Computer Law & Security Review*, Vol. 35 No. 1, pp. 89–102.

Khan, N., Yaqoob, I., Hashem, I.A.T., Inayat, Z., Ali, W.K.M., Alam, M., Shiraz, M. and Gani, A. (2014), "Big data: survey, technologies, opportunities, and challenges", *The Scientific World Journal*, Vol. 2014, p. 712826.

Koutroumpis, P., Leiponen, A., Thomas, L.D.W. and others (2017), *The (unfulfilled) potential of data marketplaces*, available at: http://pub.etla.fi/ETLA-Working-Papers-53.pdf.

Kushal, A., Moorthy, S. and Kumar, V. (2011), *Pricing for Data Markets*, Technical Report, available at: https://pdfs.semanticscholar.org/93c7/0f67a329279afe66a753252f546642a0e1aa.pdf (accessed 26 March 2019.).

Li, X.-Y., Qian, J. and Wang, X. (2018), "Can China lead the development of data trading and sharing markets?", *Communications of the ACM*, Vol. 61 No. 11, pp. 50–51.

Liang, F., Yu, W., An, D., Yang, Q., Fu, X. and Zhao, W. (2018), "A survey on big data market: Pricing, trading and protection", *IEEE Access*, Vol. 6, pp. 15132–15154.

Loebbecke, C. and Picot, A. (2015), "Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda", *The Journal of Strategic Information Systems*, Vol. 24 No. 3, pp. 149–157.

Mankiw, N.G. (2014), *Brief Principles of Macroeconomics*, 7th ed., Cengage Learning.

Manyika, J., Chui, M., McKinsey Global Institute, Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, A.H. (2011), *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, McKinsey.

McAfee, A. and Brynjolfsson, E. (2012), "Big Data: The Management Revolution", *Harvard business review*, Vol. 90, 60-6, 68, 128.

Moody, D. and Walsh, P. (1999), "Measuring the Value Of Information - An Asset Valuation Approach", paper presented at Seventh European Conference on Information Systems, ECIS, Copenhagen, Denmark.

Morris, M., Schindehutte, M. and Allen, J. (2005), "The entrepreneur's business model: toward a unified perspective", *Journal of Business Research*, Vol. 58 No. 6, pp. 726–735.

Muschalle, A., Stahl, F., Löser, A. and Vossen, G. (2012), "Pricing Approaches for Data Markets", paper presented at International Workshop on Business Intelligence for the Real-Time Enterprise.

Nasonov, D., Visheratin, A.A. and Boukhanovsky, A. (2018), “Blockchain-Based Transaction Integrity in Distributed Big Data Marketplace”, in Shi, Y., Fu, H., Tian, Y., Krzhizhanovskaya, V.V., Lees, M.H., Dongarra, J. and Sloot, P.M.A. (Eds.), *Computational Science – ICCS 2018*, *Lecture Notes in Computer Science*, Vol. 10860, Springer International Publishing, Cham, pp. 569–577.

Nickerson, R.C., Varshney, U. and Muntermann, J. (2013), “A method for taxonomy development and its application in information systems”, *European Journal of Information Systems*, Vol. 22 No. 3, pp. 336–359.

Opiria (2019), *Opiria & PDATA token: Whitepaper*, available at: www.opiria.com (accessed 29 April 2019.).

Osterwalder, A., Pigneur, Y. and Clark, T. (2010), *Business model generation: A handbook for visionaries, game changers, and challengers / written by Alexander Osterwalder and Yves Pigneur ; editor and contributing co-author, Tim Clark*, John Wiley & Sons, Hoboken, N.J.

Ozyilmaz, K.R., Dogan, M. and Yurdakul, A. (2018), “IDMoB: IoT Data Marketplace on Blockchain”, in *2018 Crypto Valley Conference*, pp. 11–19.

Philip Chen, C.L. and Zhang, C.-Y. (2014), “Data-intensive applications, challenges, techniques and technologies: A survey on Big Data”, *Information Sciences*, Vol. 275, pp. 314–347.

Piazza, M. and Pedicini, M. (2018), “What Arrow’s Information Paradox Says (to Philosophers)”, in *Handbook of Experimental Pharmacology*, Vol. 84, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 83–94.

Remane, G., Hanelt, A., Tesch, J.F. and Kolbe, L.M. (2017), “The Business Model Pattern Database — A Tool For Systematic Business Model Innovation”, *International Journal of Innovation Management*, Vol. 21 No. 01, p. 1750004.

Remane, G., Nickerson, R., Hanelt, A., Tesch, J. and M. Kolbe, L. (2016), “A Taxonomy of Carsharing Business Models”, paper presented at 37th International Conference on Information Systems (ICIS 2016), Dublin.

Reuver, M. de (2019), “Finding a Business Model for Decentralized Data Marketplaces”, available at: https://safe-deed.eu/finding-a-business-model-for-decentralized-data-marketplaces/ (accessed 23 March 2019.).

Schmid, B.F. and Lindemann, M.A. (1998), “Elements of a reference model for electronic markets”, in *Proceedings of the Thirty-First Hawaii International Conference on System Sciences, Kohala Coast, HI, USA, 6-9 Jan. 1998*, IEEE Comput. Soc, pp. 193–201.

Schomm, F., Stahl, F. and Vossen, G. (2013), “Marketplaces for data”, *ACM SIGMOD Record*, Vol. 42 No. 1, p. 15.

Schüritz, R., Seebacher, S. and Dorner, R. (2017), “Capturing Value from Data: Revenue Models for Data-Driven Services”, paper presented at 50th Hawaii International Conference on System Sciences.

Shapiro, C. and Varian, H.R. (2010), *Information rules: A strategic guide to the network economy*, 19th ed., Harvard Business School Press, Boston.

Souza, C.R.B. de, Redmiles, D., Cheng, L.-T., Millen, D. and Patterson, J. (2004), “How a Good Software Practice thwarts Collaboration. The Multiple Roles of APIs in Software Development”, in *12th International Symposium on Foundations of Software Engineering (FSE 2004), Newport Beach, CA, USA*, ACM Press, New York, NY, pp. 221–230.

Spiekermann, M., Tebernum, D., Wenzel, S. and Otto, B. (2018), “A Metadata Model for Data Goods”, paper presented at Multikonferenz Wirtschaftsinformatik 2018, Lüneburg, Germany.

Spijker, A. van 't (2014), *The new oil: Using innovative business models to turn data into profit*, Technics Publications, Basking Ridge, NJ.

Stahl, F., Schomm, F., Vomfell, L. and Vossen, G. (2017), “Marketplaces for Digital Data: Quo Vadis?”, *Computer and Information Science*, Vol. 10 No. 4, pp. 22–37.

Stahl, F., Schomm, F. and Vossen, G. (2014), *The Data Marketplace Survey Revisited*, Working Paper, available at: http://hdl.handle.net/10419/94187 (accessed 12 April 2019.).

Stahl, F., Schomm, F., Vossen, G. and Vomfell, L. (2016), “A classification framework for data marketplaces”, *Vietnam Journal of Computer Science*, Vol. 3 No. 3, pp. 137–143.

Szopinski, D., Schoormann, T., John, T., Knackstedt, R. and Kundisch, D. (2019), “Software tools for business model innovation: current state and future challenges”, *Electronic Markets*, Vol. 60 No. 11, p. 2794.

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

Täuscher, K. and Laudien, S.M. (2017), “Understanding platform business models: A mixed methods study of marketplaces”, *European Management Journal*, Vol. 36 No. 3, pp. 319–329.

Teece, D.J. (2010), “Business Models, Business Strategy and Innovation”, *Long Range Planning*, Vol. 43 No. 2-3, pp. 172–194.

Thomas, L.D.W. and Leiponen, A. (2016), “Big data commercialization”, *IEEE Engineering Management Review*, Vol. 44 No. 2, pp. 74–90.

Turner, V., Reinsel, D., Gantz, J. and Minton, S. (2014), “The Digital Universe of Opportunities. Rich Data and the Increasing Value of the Internet of Things”, available at: https://uk.emc.com/collateral/analyst-reports/idc-digital-universe-2014.pdf (accessed 14 May 2019.).

Vega-gorgojo, G., Fjellheim, R., Roman, D., Akerkar, R. and Waaler, A. (2016), “Big Data in the Oil & Gas Upstream Industry - A Case Study on the Norwegian Continental Shelf”, *Oil Gas European Magazine*, Vol. 42, pp. 67–77.

Woerner, S.L. and Wixom, B.H. (2015), “Big Data: Extending the Business Strategy Toolbox”, *Journal of Information Technology*, Vol. 30 No. 1, pp. 60–62.

Yin, R.K. (2009), *Case Study Research: Design and Methods*, 4th ed., SAGE Publications.

Zikopoulos, P., Eaton, C., Deroos, D., Deutsch, T. and Lapis, G. (2012), *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*, McGraw-Hill, New York, NY.

Zott, C., Amit, R. and Massa, L. (2011), “The Business Model: Recent Developments and Future Research”, *Journal of Management*, Vol. 37 No. 4, pp. 1019–1042.

Zuiderwijk, A., Loukis, E., Alexopoulos, C., Janssen, M. and Jeffery, K. (2014), “Elements for the Development of an Open Data Marketplace”, paper presented at Conference for e-Democracy and Open Government, Krems an der Donau, Austria.

Zyskind, G., Nathan, O. and Pentland, A. (2015), *Enigma: Decentralized Computation Platform with Guaranteed Privacy*, available at: http://arxiv.org/pdf/1506.03471v1 (accessed 25 April 2019.).

7 APPENDICES

7.1 APPENDIX 1

The list provided in Table 15 presents the set of data marketplaces that have been completely included in the analysis, together with the reference to their online presence. Last access to all of the references was in *May 2019*.

Data Platform	***Reference***
Advaneo	https://www.advaneo.de/de/
Big	https://big.exchange/
Caruso dataplace	https://www.caruso-dataplace.com/
Databroker dao	https://databrokerdao.com/platform/
Datacoup	http://datacoup.com/
Datahub	https://datahub.io/
Datapace	https://www.datapace.io/
Dawex	https://www.dawex.com/en/
Fyscal	https://fysical.org/
IOTA	https://data.iota.org/#/
Marketscan	https://www.marketscan.co.uk/
OnAudience.com	https://www.onaudience.com/dataexchange?action=dataexchange
oneTRANSPORT	https://onetransport.io/#
Opiria PData	https://opiria.io
Pitney Bowes	https://www.pitneybowes.com/us/data/sdm.html
Quandl	https://www.quandl.com/
Red Lion Data	https://www.redliondata.com/
Streamr	https://www.streamr.com/
Wibson	https://wibson.org/
Xignite	http://www.xignite.com/

Table 15 - Data platforms included in the analysis

7.2 APPENDIX 2

Table 16 presented below consists of objects that were discovered during the preparation phase of the analysis, and have been included in the starting set of analysis objects. After the analysis was finished, the presented objects have been excluded, since the developed taxonomy could not be applied to them. The reasons for the performed filtering are presented in the first column of the table – *Reasons to exclude.* The term "object" is used because part of the set is indeed data platform or data marketplace, but have been excluded for other reasons, while some of them do not fit to the definition at all. Therefore, the generalized term "object" was chosen for the purpose of presenting all cases together. Furthermore, the reference for the online presence of objects is given in the last column - *Reference*. Last access to all of the references was in *May 2019.*

Reason to exclude	***Name of organization***	***Reference***
No enough info – direct contact needed	Acxiom	https://www.acxiom.com/what-we-do/data/
	BDEX	https://www.bdex.com/
	Data republic	https://www.datarepublic.com/
	Dex	https://www.dex.sg/
	Openprise	https://www.openprisetech.com/data-marketplace/
	Salesforce Datastudio	https://www.salesforce.com/products/marketing-cloud/data-sharing/
	Thinknum	https://www.thinknum.com/
No enough info – unable to create account	Otonomo	https://otonomo.io/
	owerIQ	https://www.owneriq.com/Second-Party-Data-Education
	Quadrant	https://www.quadrant.io
	Synchronicity	https://synchronicity-iot.eu/
The platform is under construction or in the testing phase.	netObjex	https://www.netobjex.com/data-marketplace/
	Bonseyes	https://www.bonseyes.com
	Datum	https://datum.org/
	Enigma	https://enigma.co/marketplace/
	Data Market Austria	https://datamarket.at/
Does not fit the definition - data related services	Bloomberg	https://www.bloomberg.com/professional/product/bloomberg-polarlake/
	Cognite	https://www.cognite.com/
	FactSet	https://www.factset.com/services/data-delivery
	Factual	https://www.factual.com/

	Kochava	https://www.kochava.com/kochava-collective/
	Radius	https://radius.com/
	Reply	https://reply.io/data-marketplace
	Sobloo	https://sobloo.eu/data
	Veracity	https://www.veracity.com/
	Cybernetica	https://cyber.ee/products/secure-data-exchange/
Does not fit the definition – provides only links to data	dmi	https://dmi.io/
One-way market	Knoema	https://knoema.com/
	Mmojo	https://mmojo.com/data-marketplace/
	Qlik	https://www.qlik.com/us/products/qlik-data-market
Open/free data	Amazon AWS	https://aws.amazon.com/opendata/
	ArcGIS Hub	https://hub.arcgis.com/pages/open-data
	Figshare	https://figshare.com/
	Google public data	https://www.google.com/publicdata/directory
	Mobility Data Marketplace	https://www.mdm-portal.de/en/
	OpenData – Socrata	https://opendata.socrata.com/
	OpenDataSoft	https://www.opendatasoft.com/
	OSDC	https://www.opensciencedatacloud.org/

Table 16 - Objects excluded from the analysis

7.3 APPENDIX 3

Table 17 provides a complete set of dimensions consisted in the developed taxonomy and the origin of every dimension, including the reference to the source of those dimensions which are conceptually collected. Furthermore, all characteristics for each dimension are presented, together with a short description of every characteristic.

Dimension	***Origin***	***Characteristic***	***Description of characteristic***
Platform infrastructure	Conceptual from Koutroumpis et al. (2017)	Centralized	Data are stored and accessed from predefined storage spaces
		Decentralized	Data are stored decentralized (e.g., using blockchain)
Data origin	Conceptual from Stahl et al. (2017)	Internet	Data are gathered from online sources (manually or automatically)
		Self-generated	Data are gathered from private sources
		User-generated	Data are collected from user-inputs (e.g., in exchange for using a service)
		Community	Data are collected from marketplaces and crowdsourcing services
		Governments	Data are provided by government institutions (e.g., ministries)
		Authority	Data are collected by institutions with a high level of proficiency
Additional purchase support	Empirical	With additional costs	Data marketplace charges for additional services
		Included in price	Data marketplace offers additional services for free
		No	Data marketplace does not provide additional services
Review system	Conceptual from Täuscher (2016), Täuscher and Laudien (2018)	Reviews by users	Reviews directly between buyers and sellers
		Reviews by marketplace	Data marketplace provides reviews
		None	Data marketplace does not provide reviews
Data quality guarantee	Empirical	Guaranteed	Data marketplace guarantees quality of purchased data
Time relevancy	Conceptual from Stahl et al. (2017)	Static	Offered data does not change after its creation
		Dynamic	Regular updates to dataset needed to keep data valid

	Empirical	Both static and dynamic	Data marketplaces offer both data that required frequent updated as well as data which does not change over time
Type of access	Conceptual from Stahl et al. (2017)	API	Use of a predefined protocol interface to access data
		Download	Data are accessed through the downloadable file
		Specialized Software	Data marketplace requires designated software to handle data
		Web Interface	Data are accessed through a web browser
	Empirical	API and Download	Data can be accessed via API as well as via download
Data output type	Conceptual from Stahl et al. (2017)	XML	Format for semi-structured data
		JSON	Format for semi-structured data
		RDF	Semantic Web-Format
		CSV/XLS	Tabular data
		Report	Visualized data formats (e.g., PDF, DOC, JPEG)
	Empirical	Multiple options	Data marketplace offers multiple options for data output types
Pre-purchase testability	Conceptual from Stahl et al. (2017)	Complete access	Customers have complete access before paying for data
		Restricted access	Customers can access only part of the data before prior purchase
		None	Customers cannot access data before paying for them
Domain	Conceptual from Stahl et al. (2017), Fricker and Maksimov (2017)	All/Any	Data marketplace not restricted to a certain domain
		Finance/Economy	Economics related data (e.g. stock market data or pricing data)
		Scientific	Research data
		Social Media	Data gathered from social media
		Geo	Geographical positions expressed in coordinates (e.g., businesses or individuals)
		Address	Lists of customer information (e.g., mail and E-mail addresses)

		Sensor	Data generated by or used for sensors (e.g., IoT data)
	Empirical	Personal	Data related to private information about individual (e.g. location, health, genetic, economic, cultural data)
Privacy	Empirical	Anonymized	Data marketplace stores anonymized data
		Encrypted	Data marketplace stores encrypted data
		Anonymized and Encrypted	Data marketplace stores anonymized and encrypted data
Marketplace participants	Conceptual from Täuscher and Laudien (2018)	C2C	Data marketplace operates exclusively C2C
		B2C	Data marketplace operates exclusively B2C
		B2B	Data marketplace operates exclusively B2B
	Empirical	C2B	Data marketplace operates exclusively C2B
		Any	Data marketplace not restricted in terms of buyers and sellers
Pricing model	Conceptual from Muschalle et al. (2013), Stahl et al. (2014)	Free	Selected datasets are offered for free (e.g., to increase awareness for the data marketplace)
		Usage-based	Customers pay proportionally for units (e.g., API-calls or time)
		Package pricing	A selected amount of data is offered for a fixed price
		Flat free tariff	Full access to the data marketplace is offered for a recurring fee
		Two-part tariff	Combines a flat fee tariff with additional usage-based pricing
		Freemium	Basic features can be used for free while additional features are unlocked for a charge
Price discovery	Conceptual from Täuscher (2016), Täuscher and Laudien (2018)	Fixed prices	Data marketplace has fixed prices
		Set by sellers	Prices are set by sellers
		Set by buyers	Prices are set by buyers
		Auction	Buyers and sellers are bidding against each other
		Negotiation	Buyer and seller agree on an acceptable price for both parties

Payment currency	Empirical	Crypto-currency	Data marketplace handles payment via crypto-currency
		Fiat-currency	Data marketplace handles payment via fiat currency
		Both Crypto and Fiat Currency	Data marketplace handles payment both with fiat currency and crypto-currency
Smart contract with blockchain	Empirical	Yes	Data marketplace offers an option for smart contracting
		No	Data marketplace does not offer an option for smart contracting

Table 17 – Complete description of taxonomy

7.4 APPENDIX 4

Table 18 gives a complete analysis of results from the taxonomy development process.

Value Creation

#	Marketplace	Platform infrastructure		Data origin					Review system			
		Centralized	Decentralized	Internet	Self-generated	User	Community	Authority	User reviews	Review by marketplace	None	No info
1	Advaneo	x			x				x			
2	Big	x			x							x
3	Caruso	x			x							x
4	Datahub	x					x					x
5	Marketscan	x					x				x	
6	OnAudience.com	x		x								x
7	oneTRANSPORT	x			x							x
8	Pitney bowes	x						x			x	
9	Quandl	x						x			x	
10	Red Lion Data	x						x			x	
11	Xignite	x						x				x
1	Dawex	x			x							x
1	Databroker dao		x		x				x			
2	Datapace		x		x					x		
3	IOTA		x		x						x	
4	Streamr		x		x						x	
5	Fysical		x		x						x	
1	Datacoup	x				x					x	
2	Opiria PData		x			x					x	
3	Wibson		x			x				x		

Value Proposition

#	Marketplace	Privacy				Data quality guarantee		Time relevancy			Type of access					Data output type					Pre-purchase testability		
		Anonymized	Encrypted	Both	No info	Yes	No info	Static	Dynamic	Both	API	Download	Specialized software	API/Download	No info	CSV/XLS	JSON	Report	Multiple options	No info	None	Restricted access	No info
1	Advaneo				x		x	x				x							x		x		
2	Big	x					x			x					x					x		x	
3	Caruso				x		x			x	x									x			x
4	Datahub				x		x		x					x					x		x		
5	Marketscan				x	x		x				x				x						x	
6	OnAudience.com	x				x				x					x					x	x		
7	oneTRANSPORT				x		x			x	x						x				x		
8	Pitney bowes				x	x			x					x					x			x	
9	Quandl	x				x			x					x		x						x	
10	Red Lion Data				x	x			x			x				x					x		
11	Xignite				x	x				x	x								x			x	
1	Dawex		x			x				x				x		x						x	
1	Databroker dao				x	x			x		x						x					x	
2	Datapace			x		x			x		x						x				x		
3	IOTA		x			x			x		x						x				x		
4	Streamr				x	x			x			x				x					x		
5	Fysical	x				x			x		x					x					x		
1	Datacoup	x					x		x				x							x	x		
2	Opiria PData			x		x		x					x					x			x		
3	Wibson	x				x			x				x							x	x		

Value Delivery

#	Marketplace	Additional purchase support			Domain						Marketplace participants			Smart contract with blockchain	
		With additional costs	Included in price	No	All/Any	Finance/Economy	Geo	Address	Sensor	Personal	B2B	C2B	Any	Yes	No
1	Advaneo			x	x								x		x
2	Big		x							x			x		x
3	Caruso			x					x		x				x
4	Datahub	x			x								x		x
5	Marketscan	x						x			x				x
6	OnAudience.com	x								x	x				x
7	oneTRANSPORT		x						x				x		x
8	Pitney bowes	x					x						x		x
9	Quandl	x				x					x				x
10	Red Lion Data	x						x			x				x
11	Xignite	x				x					x				x
1	Dawex	x			x						x			x	
1	Databroker dao			x					x		x			x	
2	Datapace			x	x								x	x	
3	IOTA			x					x		x			x	
4	Streamr			x	x								x	x	
5	Fysical			x			x						x	x	
1	Datacoup			x						x		x			x
2	Opiria PData		x							x		x		x	
3	Wibson			x						x		x		x	

Value Capture

#	Marketplace	Pricing model					Price discovery					Payment currency		
		Usage-based	Package pricing	The flat fee tariff	Freemium	No info	Fixed prices	Set by sellers	Set by buyers	Auction	Negotiation	Crypto	Fiat	Both
1	Advaneo	x					x						x	
2	Big					x				x			x	
3	Caruso			x				x					x	
4	Datahub				x		x						x	
5	Marketscan			x			x						x	
6	OnAudience.com		x				x						x	
7	oneTRANSPORT				x			x					x	
8	Pitney bowes			x			x						x	
9	Quandl				x			x					x	
10	Red Lion Data		x				x						x	
11	Xignite	x					x						x	
1	Dawex	x						x					x	
1	Databroker dao	x					x					x		
2	Datapace		x				x					x		
3	IOTA			x				x				x		
4	Streamr				x			x						x
5	Fysical			x							x	x		
1	Datacoup	x					x						x	
2	Opiria PData	x					x					x		
3	Wibson	x							x			x		

Table 18 - Data marketplaces analysis