enabling a **decentralized architecture** for **data marketplaces**

TU Delft

master thesis report
2021

ENABLING A **DECENTRALIZED ARCHITECTURE** FOR **DATA MARKETPLACES**

An exploratory study:
Investigating the impact of a decentralized architecture on the business models of data marketplaces

Master thesis submitted to Delft University of Technology
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in Management of Technology
at the Faculty of Technology, Policy and Management

by

EUGEN GAVRILUȚA

Student number: 4927435

To be defended in public on November 18th, 2021

Graduation committee

Chair and First Supervisor:	Dr.ir. GA (Mark) de Reuver Section of Information and Communication Technology
Second Supervisor:	Dr. B. (Ben) Wagner Section of Organisation and Governance
Supervisor:	Dr. H.A. (Hosea) Ofe Section of Information and Communication Technology

Preface

It requires some effort to write an MSc thesis, and it is not quite easy, surprisingly. Although, during the conceptualization of this thesis, I learned a lot about conducting scientific research as well as the three areas of investigation: decentralized architecture, business modeling, and data marketplaces. I could also say that the thesis research was enjoyable, sometimes. For certain they are moments in any longer projects when you want to stop and forget about it. From time to time, this is a good thing to consider. More specifically, to take breaks. Thereafter, as one of my supervisors mentioned, when coming back to the thesis work you would have a different look at it. Keep in mind to not take too long breaks. With this in mind, I could confirm that being committed pays off. At least the end of this journey comes with the personal and professional satisfaction of the completed studies. So, hopefully, time for a longer break, but not too long.

In the following, I would like to express my thanks to the people who have been very helpful to me during the time of writing the thesis. I have been very fortunate with my supervisors. Mark, Ben, and Hosea were more than anything fantastically normal support for improving the thesis. And most importantly, my gratitude for their motivation that kept me going with the research. From Mark, I learned the significance of asking the right questions. When properly framed, questions may trigger deep thoughts and are useful in exposing those requirements that could lie beneath the surface. So, to my future colleagues, be prepared for many questions. From Ben, I learned the implication of constructive feedback. Formal and informal feedback is always appreciated especially when addressed not as remarks but as recommendations for making a project stronger or more understandable. So, to my friends, get ready for some constructive feedback. From Hosea, I learned the meaning of focusing on the Why. The reason for frequently asking Why is to help pursue beyond basic ideas and already researched concepts. So, to Delft University, should I continue with a PhD degree? Why or why not?

There is the rumor that nothing is harder than a short call with another human being. Although this statement is an exaggeration, it is also on the pulse of something true about how our online schedule has changed - a meeting request is now a big ask. In this regard, thanks are owed to researchers and industry experts that accepted to have an interview with me. Their knowledge and experience in the subject areas of data marketplaces reduced possible errors and clarified misunderstandings in the research process. For privacy-perceiving reasons, the names or any other personal information are anonymized. Thanks to GDPR. Although, I will make sure to send a separate email with my gratitude for their time shared for a nice discussion with an MSc student.

Last but not least, and I think the most important, many thanks to my family for their unconditional love as well as to my friends for their encouragement. Here words are needless to express my appreciation for them being next to me through this master program. Even if sometimes it was hard to explain what I am researching, I could not have done it without you all!

Eugen Gavriluța

Delft, November 2021

Contents

List of Abbreviations

GDPR	General Data Protection Regulations
EU	European Union
EC	European Council
EP	European Parliament
EDS	European Data Strategy
DGA	Data Governance Act
TRUSTS	Trusted Secure Data Sharing Space
DSM	Digital Single Market
DSR	Design Since Research
HLA	High-Level Architecture
BM	Business Model
IS	Information Systems
FEDS	Framework for Evaluation in Design Science
TEE	Trusted Execution Environment
DLT	Distributed ledger technology

List of Figures

List of Tables

Executive Summary

Data Marketplaces are emerging digital platforms in data management. These platforms enable the infrastructure for data exchange, offering the opportunity for providers to commercialize their datasets and save time for consumers in identifying appropriate datasets. Conceptually, data marketplaces provide data exchange services as their core activity. However, in practice, organizations may be reluctant to exchange data through intermediaries as marketplaces instead of exchanging data bilaterally (Stahl et al., 2015). The bilateral relationship gives more control over the process to providers and consumers. For this purpose, a decentralized architecture could represent the peer-to-peer technical solution for the bilateral exchange. Therefore, what was previously called intermediaries, in the decentralized environment, data marketplaces are communication structures facilitating the commercial exchange of data between organizations (Koutroumpis et al., 2017). In the decentralized design, each transaction is recorded in a distributed ledger that is transparently revealed to the network, and participants are actively validating its contents (Zheng et al., 2018). In this regard, decentralization excludes by design the need for intermediaries as marketplaces to collect, host, or sell data. Although, with the exclusion of intermediaries, actors engaging in data exchange through decentralized marketplaces could be directly responsible for the commercial valorization of their assets.

Both from technical and business perspectives, studies on data marketplaces as digital platforms have already been conducted. However, in most scientific publications, data marketplaces are researched as a general phenomenon, defining the elementary differences between centralized and decentralized platform infrastructures. Comprehensive explanations of how the technical functionalities of a decentralized architecture are implemented in the context of data marketplaces, in some cases, are weakly described and more research is required for their inclusive understanding. Subsequently, for the business properties, the centralized approach is commonly considered as the primary research option with fewer assertions regarding decentralization. Although, decentralization introduces clear technological and business model distinctions for data marketplaces. More specifically, sustaining the business operations of data exchange infrastructure differs from controlling the data flow in the case of centralized marketplaces to transactions being validated by network participants in the case of decentralized marketplaces. Further, with no intermediaries, data providers and consumers exchange data in peer-to-peer transactions. Therefore in decentralized marketplaces, the relationships between actors are achieved through smart contracts establishing the bilateral agreement terms. Lastly, with less influence on data collection, hosting, or selling methods, the integration of supporting technologies may represent an imperative practice in the context of a decentralized architecture. Thus, based on the described distinctions, in this thesis, a specific focus on a decentralized architecture, namely blockchain technology, was considered for data marketplaces. The overarching objective was *to investigate the impact of a decentralized architecture on the business models of data marketplaces*. For this purpose, the main research question that represents the academic challenge and structures the research into an exploratory study was addressed as follows: *What is the impact of a decentralized architecture on the business models of data marketplaces?*

Two variables can be distinguished in the main research question: the *decentralized architecture* and the *business models*. To investigate the impact of the decentralized architecture on the business models, the two variables are constructed in the form of artifacts using the Design Since Research (DSR) methodology as described by Hevner (2007). More specifically, a *High-Level Architecture (HLA)* was designed to comprehensively outline the technical properties of decentralized data marketplaces. Subsequently, to schematically classify their business model characteristics a *Taxonomy of Business Models (BMs)* was established. Moreover, besides designing the artifacts, Hevner (2007) stated that it is also essential to maintain the balance between the construction and the evaluation of artifacts, as it provides input for further development and research rigor. Grounded arguments could be insufficient with a weak evaluation (Hevner, 2007). For this purpose, the well-cited Framework for Evaluation in

Design Science (FEDS), as described by Venable et al. (2016), was used. In FEDS, two main distinctions regarding the functional purpose of evaluation are described. These distinctions rely on *formative* and *summative evaluations*. The purpose of the *formative evaluation* was to produce empirical-based interpretations that provide the basis for improving the characteristics of the research artifacts. Thereafter, with the improved artifacts, the purpose for *summative evaluation* was to produce empirical-based interpretations that provide a basis for creating shared meanings about the designed artifacts (Venable et al., 2016). In both evaluation processes, researchers and industry experts in the subject areas of data marketplaces were consulted to gather their recommendations for improving the artifacts as well as their critical and supportive viewpoints for creating shared meaning about the artifacts. Therefore, by adding the different viewpoints of experts, the interviews reduced possible errors, clarified misunderstandings, and strengthen the generalized interference of the research.

As already noted, the thesis research follows an association of three areas of investigation: decentralized architecture, business modeling, and data marketplaces as the primary study interest. Although, before researching the decentralized architecture and business models, data marketplaces had to be explicitly understood. Moreover, for the design of the HLA, their infrastructure components were studied according to the current scientific literature. Therefore, considering the characteristics of viewing data as economic assets and the technical functionalities of data trading platforms, in the thesis, the following generic infrastructure components were defined: *data process and analytics*; *data quality assessment*; *pricing mechanism*; *query & search*; *data exchange*; *data storage, custody, manipulation*; and *identity management*. These infrastructure components could be summarized as follows. Data are generally regarded as intermediate goods that *need to be further processed and analyzed* with complementary technologies for consumers to gain utility (Koutroumpis et al., 2017). Although considering their nonexistent property rights as intangible products, the *pricing models* are important mechanisms in the commercialization process (Teece, 2010). Subsequently, for *data quality*, metadata is an essential reference point both for providers and consumers. Through metadata, providers can describe the supplied datasets, and consumers can trace these specifications, thereby enabling the *query & search* functionalities. Technically, data marketplaces provide the *data exchange* infrastructure for finding, buying, and selling data (Fricker & Maksimov, 2017). Additionally, besides the data exchange infrastructure, the regulatory processes secured through different licensing models are also required to ensure *data storage, custody, and manipulation* (Stahl et al., 2015). Lastly, actors as data providers or consumers should have appropriate access rights to their resources through *identification, authentication, and authorization mechanisms*.

On the basis of the identified infrastructure components, their technical specifications were analyzed in the context of blockchain technology with the ultimate goal to design a decentralized HLA for data marketplaces. For this purpose, the Enterprise Application Architecture was used to integrate the infrastructure components into a layered architecture. After the evaluation phase, seven layers were defined in the final HLA: *Access Layer*, *Application Layer*, *Integration Layer*, *Data Layer*, *Smart Contract Layer*, *DLT Layer*, and *Security Layer*. These layers could be summarized as follows. The *Access Layer* is designed to ensure the ownership, sovereignty, and control of data through tokenization processes. Further, the *Application Layer* is the conceptual layer that gathers the infrastructure components with which actors directly interact. Next, the *Integration Layer* was included to expand the architecture capabilities acting as a middleware for third-party services. Moreover, to outline the efficient management of data assets, the *Data Layer* was composed of internal and external infrastructure components. Further, by applying similar principles as conventional licensing models, in the *Smart Contract Layer,* the constraints and the regulatory processes for data transactions were described. With the established data constraints, the *DLT Layer* defined the on-chain and off-chain environments for storing the information about the performed transactions and to handle the actual data transfers from providers to consumers. Lastly, the *Security Layer* characterized the asymmetric cryptographic mechanisms for the unique identification of actors. As a result, the eight layers with aligned infrastructure components constructed the HLA, which represented the first designed artifacts.

In the thesis, the firm belief is followed that both the technical and business specifications are essential in understanding the data marketplaces with blockchain architecture. In particular, research on business models is required to advance the development of these emerging data platforms. Therefore, the second designed artifact was a classification scheme referred to as a Taxonomy of BMs. For this purpose, a case study was conducted on five active data trading platforms by examining their whitepapers through desk research. The final taxonomy comprised *3 meta-characteristics* according to the rationale of how data marketplaces *create* and *deliver* value to customers and then convert payments received to profits, describing the *captured* value (Teece, 2010). Thus, using the meta-characteristics of value *creation*, *delivery*, and *capture*, *8 compositional* elements were distinguished from the business model canvas as the leading framework in taxonomy construction (Osterwalder & Pigneur, 2010). Moreover, to accomplish a focused analysis of data marketplaces, the *compositional elements* were divided into *18 business dimensions* subtracted from previous studies and aligned according to the specification of blockchain technology (Bergman, 2020, M. van de Ven et al., 2021, and Fruhwirth et al., 2020). Lastly, with the subtracted dimensions, *46 business model characteristics* were identified in the whitepapers of the data trading platforms. Therefore, the *3 meta-characteristics*, *8 compositional* elements, *18 business dimensions*, and *46 business model characteristics* constructed the Taxonomy of BMs, which represented the second designed artifact.

The designed artifacts provided both technical and business insights that were utilized in investigating the application of blockchain technology in data trading platforms as well as in evaluating their business model characteristics. From a technical perspective, the primary role of data marketplaces in the decentralized environment is offering the tools for actors as providers and consumers to operate independently in the data exchange. This means that the efficiency of the decentralized design may directly depend on the quality of their tools. In this sense, the primary technical recommendation for practitioners is to examine the possibility of establishing the operation base for the development and maintenance with the primary focus on increasing the overall platform and tools quality. Moreover, considering that in a decentralized architecture, there are no intermediaries, growing the user base and developing a functional ecosystem of data service providers could represent a fundamental requirement in the inception phase of the decentralized markets. For this purpose, the HLA offers a structural framework to define the architectural layers and their infrastructure components that are easily understandable and would not require much effort to grasp how everything is technically tied together. From a business perspective, data marketplaces with blockchain architecture use cryptocurrency rewards to sustain the data exchange infrastructure. More specifically, marketplace owners substitute the initial investment costs or maintain the development process by leveraging their cryptocurrency resources. Although, to benefit from desirable revenue models a business recommendation for practitioners is to cautiously determine the crypto-economic strategies that raise the cryptocurrency value. This may incentivize the participation of complementary technology providers or knowledgeable actors to enhance the marketplace usability. Moreover, when defining the strategies for growing the data marketplaces as commercial organizations, the Taxonomy of BMs could function as a map of business dimensions to analyze their market position and to support the decision-making processes.

This thesis research makes several scientific contributions. Firstly, a decentralized architecture in the context of data marketplaces may represent a complex technological topic. Therefore, the researched infrastructure components could offer the fundamental knowledge required in studying the technical specifications of decentralized marketplaces as well as the domain-specific application of blockchain technology in different use cases for data exchange. Secondly, previous studies on the business models of data marketplaces seemed to cover only the general properties of decentralization. Thus, the analyzed business characteristics contribute to a more focused and inclusive understanding of the business specification for decentralized data platforms. Thirdly, the constructed artifacts could be regarded as frameworks suitable for practical use. Thereby, marketplace owners may evaluate their platform architecture or analyze their data marketplaces using business modeling as a viewing lens.

"Is data the new oil?"

- European Parliament

1
Introduction

Over the last decades, the number of growing technologies shaped the modern economy. With society becoming increasingly connected, technologies are transforming all sectors of activities. Nowadays, meaningful information generated from data is crucial in almost all business areas (Stahl et al., 2014a). To remain competitive - digitalization is the differentiator in the emerging digital economy. The center of transformation relies on data and data-driven innovations. However, finding the necessary data that is the most suitable for someone's requirements, considering the prolific generation and supply of data, is a challenging task (Stahl et al., 2014b). This challenge led to the inception of specialized marketplaces where data is exchanged as a commodity.

The digital platforms offering data exchange services, as their core activity, are characterized as data marketplaces. Conceptually, data marketplaces are intermediary platforms that connect data providers, consumers, and other complementary technology providers (Koutroumpis et al., 2017). Thus, on the one hand, data providers have the opportunity to commercialize their data commodity through these platforms, saving effort in the negotiation process. On the other hand, data consumers have a substantial access range of data, saving time in finding the necessary datasets. However, difficulties emerge in large-scale data exchange. Both for data providers and consumers, privacy and security concerns may discourage participation in data marketplaces (Koutsos et al., 2020). Mechanisms that allow the disclosure of personal data with intermediaries are difficult to be achieved under the General Data Protection Regulations (GDPR). Moreover, businesses are lacking economic incentives for exchanging data, especially with regard to losing the competitive edge. Numerous attempts to establish data marketplaces failed. For instance, after seven years of poor performance, the Microsoft Azure DataMarket closed down. Thus, it appears challenging to set up a large-scale open marketplace for trading data (Koutroumpis et al., 2020).

In practice, commercial data marketplaces purchase data and further sell data on demand via a negotiated contractual agreement (Bergman, 2020). In this context, data marketplaces directly control the data flow from providers towards consumers and could be classified as centralized multilateral data platforms. According to Koutroumpis et al. (2017), a centralized design could only work when the offered services are more beneficial than the bilateral data exchange. This means that centralized multilateral data platforms have to provide clear incentives for exchanging data in terms of competitive prices or efficient services. Consequently, these incentives should be more favorable than exchanging data directly between providers and consumers. Previous studies have shown that hierarchical relations dominate the data market (Stahl et al., 2015). The term hierarchical defines the providers with only a single data domain offering. This implies that hierarchical and bilateral relations between the data providers and consumers are preferred over intermediary relations through marketplaces. Although, the most common architecture for data marketplaces still has the conventional characteristics of a centralized approach despite the dominant hierarchical relations in the data market (Fruhwirth et al., 2020). Decentralization could represent an alternative to the centralized approach. In the decentralized architecture, the data providers and consumers can exchange data in peer-to-peer transactions. This peer-to-peer data exchange is accomplished through decentralized transaction storage. More specifically, transactions in a decentralized architecture are not stored on a central entity but encrypted and distributed to actors sustaining the network. Technically, the decentralized transaction storage could be achieved through blockchain technology with smart contracts to establish the peer-to-peer agreement terms (Hynes et al., 2018). In this context, any direct control of intermediaries on the data flow is excluded while the marketplace benefits for data exchange are perceived (Koutroumpis et al., 2017). The perceived benefits indicate that decentralized data marketplaces share the same facilities as centralized marketplaces in terms of the substantial range of data or the time saved for finding datasets. Additionally, decentralized marketplaces alleviate some of the centralized limitations regarding privacy, transparency, and data ownership (Koutroumpis et al., 2017). However, excluding intermediaries from the data flow assigns more control in the exchange process to data providers and consumers. In this regard, more control in the process could also imply more responsibility in the data management process. Thus, on the one hand, marketplaces may offer the infrastructure and technical tools for decentralized data exchange. On the other hand, actors using decentralized data marketplaces would be directly responsible for the commercial business valorization of their assets. In this context, in understanding the data marketplaces with decentralized architecture, both the technical and business specifications are essential. Therefore, the decentralized architecture, as well as the business models for data marketplaces, are researched in this thesis.

The remaining sections of this chapter are structured as follows. The research context regarding the European Union (EU) strategies for the data economy is discussed in section 1.1. Next, the research problem is presented in section 1.2. Subsequently, the knowledge gap, the research objective, and the main question are specified in section 1.3. The research approach is described in section 1.4. Thereafter, the research process and the sub-questions are explained in section 1.5. Moreover, the scope of the research is outlined in section 1.6. Further, the research overview is established in section 1.7. Lastly, the chapter concludes with a reading guide for the entire thesis in section 1.8.

1.1. Research Context

1.1.1. General Research Context

Organizations have been trading data for a very long time. Companies like Reuters and Dun & Bradstreet have been selling data since the mid-19th century (Carnelley et al., 2013). With increasing volume and value perceived - data is an important asset in the current digital economy. The value of data for the 27 Member States in the digital economy is estimated to reach an amount of 829 billion Euros by 2025 (European Commission, 2020b). In this context, few tech giants have a significant part of the world's data. According to Synergy Research Group, Amazon and Microsoft are worldwide market leaders in cloud services, followed by Google, Alibaba, and Tencent (srgresearch, 2020). These tech giants and private non-EU companies have numerous opportunities to manipulate the data market into their interest. But the EU also has the potential to be a strong competitor in the data-driven market. Thus, in March 2019, the European Council (EC) announced that *"The EU needs to go further in developing a competitive, secure, inclusive and ethical digital economy […]. Special emphasis should be placed on access to, sharing of and use of data […]"* (European Council, 2019). The announcement followed the European Data Strategy (EDS) to establish the framework in ensuring attractive policies for businesses, researchers, and public administrations to cooperate in data sharing. EU aims to become a leading data-driven society and create a single European data market where industrial and sensitive personal data are securely regulated and used (European Commission, 2020a). In this regard, clear steps have been taken. The European Parliament (EP) and the Council released the regulation for the free flow of non-personal data in 2018 (European Parliament, 2018). Thereafter, the Open Data Directive has been in force since 2019 to allow the re-use of data held by national public administrations (European Parliament, 2019). Lastly, in November 2020, in compliance with the European Data Strategy, the EP and the Council elaborated the proposal for the Data Governance Act (DGA). The DGA aims *"at facilitating data sharing including by reinforcing trust in data sharing intermediaries that are expected to be used in the different data spaces"* (European Parliament, 2020). At the moment of writing the thesis, the DGA is in the development process to *"interplay with the legislation on personal* *data"*. The legislation on personal data, namely the GDPR, must be considered a foundation in the EDS and directly applied to the DGA (European Parliament, 2020).

Digital platforms providing data-sharing services represent an attractive research topic considering the EU strategy in placing a special emphasis on data access, sharing, and use. Thus, data marketplaces could have an essential role in the growing European digital economy. Furthermore, regarding the fundamental regulations on the use of personal data, a decentralized architecture may ensure the peer-to-peer exchange of data, thereby potentially enhancing privacy in collaboration between organizations.

1.1.2. Project-Specific Context

The current thesis research is carried out as part of the TRUSTS project. Trusted Secure Data Sharing Space (TRUSTS) is an EU-funded project focusing on bringing together highly interdisciplinary partners to improve the technical aspects of trust, security, and privacy in data exchange practices. Trust, security, and privacy could facilitate faster growth of the data economy in Europe, thereby creating incentives for data owners to valorize their data assets. The aim of the project is *"researching and developing technology, business models, ethical and legal guidelines to enable the promise of the Digital Single Market (DSM)"* (trusts-data, 2020). Subsequently, the project objective is to develop a data-sharing platform complying with GDPR that could integrate future platforms in different jurisdictions.

1.2. Research Problem

The topic of data marketplaces has been broadly investigated over the last period in Europe. Joint efforts of the EU, researchers and businesses focus on improving the data market through multisided platforms for sharing and re-using data. Studies distinguished different characteristics of the architectures used in data marketplaces. For example, Koutroumpis et al. (2017) suggested three data-trading frameworks: a centralized platform, a decentralized platform, and a consortium-like collective governance model. In a more recent study, Fruhwirth et al. (2020) described four differing archetypes in the business models of data marketplaces, centralized data trading, centralized data trading with smart contracts, decentralized data trading, and personal data trading. However, with regard to the platform architecture, a clear distinction can be made between the centralized and decentralized approaches (Spiekermann, 2019). In centralized platforms, data providers and consumers can exchange data via marketplaces that are directly responsible for storing and managing datasets. These private or collectively governed platforms have their own or collective commercial interests. Thus, with the significant power of storing and managing datasets, centralized marketplaces may manipulate the data exchange process into their interest, which could differ from the interest of actors participating in the exchange. This introduces the problem in the alignment of interests for all actors. The alternative for the centralized architecture is the decentralized approach. A decentralized architecture based on blockchain technology with terms of transaction agreed between providers and consumers through smart contracts may ensure the preservation of privacy and ownership of data (Hynes et al., 2018; see also Spiekermann, 2019). Although, for a comprehensive understanding of potential opportunities with the decentralized architecture, namely blockchain technology, two aspects are essential in the research. First is the decentralized architecture itself, and second is the business models for data marketplaces with blockchain technology. The importance of the two aspects is explained in the following sections.

1.3. Knowledge Gap, Objective, Question

From technical and business perspectives, studies on data marketplaces have already been conducted. For instance, Koutroumpis et al. (2017) classified the data-trading platforms, described the design of the centralized and decentralized data marketplaces and introduced the benefits of decentralization. In more specific terms, Spiekermann (2019) researched the trends of data monetization and pointed out the use of blockchain technology as a decentralized architecture to provide high-level transaction integrity and data sovereignty. Moreover, both within the TU Delft repository and in the databases for scientific articles, research can be found on the business models for data marketplaces. For example, Bergman (2020) and van de Ven (2020) studied the business models of data marketplaces and created a taxonomy for the analysis of business model patterns. In the same context, Fruhwirth et al. (2020) researched the characteristics of data marketplaces from a business perspective. In particular, research on the business models in the context of data marketplaces is required to advance the development of these emerging data platforms. Business model frameworks may contribute to both theoretical and practical understanding of data marketplaces, enabling the transformation of technical ideas into functioning value propositions as well as the strategic design of business decisions for sustainable growth. Moreover, considering the special emphasis of EU authorities on access to, sharing of and use of data, described in section 1.1, the research on data marketplaces is quickly evolving in Europe. Therefore, scholars and practitioners require comprehensive business models to derive the value created for actors to further establish a competitive, secure, inclusive, and ethical digital economy.

The previously described studies are examples of available scientific papers on data marketplaces. In these examples, most of the research focuses on data marketplaces as a general phenomenon describing the differences between centralized and decentralized approaches. Thus, the research on decentralized architecture in the context of data marketplaces is rather fragmented among studies. For instance, in most of the technical papers, authors refer to the decentralization benefits with arguments based on the immutability, transparency, or security properties. Although, comprehensive explanations of how these properties are technically implemented, in some cases, are weakly described and more research is required for their inclusive understanding. Further, from the business perspective, scholars mainly note the difference between centralized or decentralized infrastructures for data marketplaces. However, the business activities of decentralized data marketplaces could have their particularities, considering that in the decentralized infrastructure, providers and consumers control the exchange process. Moreover, the elementary distinction of scholars in the data marketplace infrastructures indicates that in most of the studies, the centralized approach is commonly considered as the primary research option with fewer assertions regarding decentralization. Therefore, this thesis aims to have a more focused approach to decentralization in the subject areas of data marketplaces covering the described fragmentations that may also be considered a **knowledge gap**.

In the decentralized design, each transaction is recorded in a distributed ledger that is transparently revealed to the network, and participants are actively validating its contents (Zheng et al., 2018). In this regard, what was previously regulated by a centralized marketplace, now is a communication structure that facilitates the operation of transactions by offering technical tools in the data market (Koutroumpis et al., 2017). Thus, decentralization excludes by design the need for intermediaries as marketplaces to collect, host, and sell data. This introduces clear distinctions in the business models of data marketplaces with a decentralized architecture. Firstly, *sustaining the business operations of data exchange infrastructure* differs from controlling the data flow in the case of centralized marketplaces to transactions being validated by network participants in the case of decentralized marketplaces. Secondly, with no intermediaries, data providers and consumers exchange data in peer-to-peer transactions. Therefore in decentralized marketplaces, the *relationships between actors could be achieved through smart contracts* establishing the bilateral agreement terms. Thirdly, with less influence on data collection, hosting, or selling methods, the *integration of supporting technologies* may represent an imperative practice in the context of a decentralized architecture. Moreover, data marketplaces could also *require seeking new sources of revenue* if earlier described data services are provided by third-party supporting technologies. Therefore, in sustaining comprehensive business models, decentralized marketplaces differ in some instances from centralized marketplaces. For this purpose, the **research objective** of the thesis is: *To investigate the impact of a decentralized architecture on the business models of data marketplaces*.

Based on the knowledge gap and considering the research objective, the **main research question** that represents the academic challenge and structures the research into an exploratory study is: ***What is the impact of a decentralized architecture on the business models of data marketplaces?***

1.4. Research Approach

The research approach is intended to provide an explicit and structured framework for the research process. Furthermore, on the basis of the research approach, the research activities and the sub-research questions can be methodologically organized to answer the main research question. Two variables can be distinguished in the main research question: the *decentralized architecture* and the *business models*. In investigating the impact of the decentralized architecture on the business models, in the thesis, the two variables are constructed in the form of artifacts using the Design Since Research (DSR) methodology as described by Hevner (2007). More specifically, a High-Level Architecture (HLA) is designed to establish the comprehensive technical characteristics of data marketplaces with a decentralized architecture. Subsequently, a Taxonomy of Business Models (BMs) is established to identify the suitable characteristics in analyzing the decentralized data marketplaces using business modeling as a viewing lens. In the following, the DSR methodology is explained in compliance with the two research artifacts.

The DSR methodology provides specific guidelines for building and evaluating design artifacts in the Information Systems (IS) field, supporting the so-called construction problems (Hevner, 2007). The methodology can be characterized as a practical and less philosophical approach, which is also the preferred outcome for the research process in this thesis. According to Hevner (2007), DSR comprises three complementary research cycles: the *relevance cycle*, the *rigor cycle*, and the *design cycle*. First, in the *relevance cycle*, the problems and opportunities are introduced, thereby defining the contextual environment of the research project (Hevner, 2007). This cycle is driven by the desire to improve the environment through the innovative introduction of artifacts and the processes used to create them. Next in the *rigor cycle*, the research activities are connected with scientific theories, methods, and expertise to establish the knowledge base and the theoretical foundation. This cycle provides past knowledge to ensure that the designed artifacts deliver a research contribution and not a routine design based upon applying well-known processes (Hevner, 2007). Lastly, the *design cycle* represents the central research activity iterating between building and evaluating the design artifacts. In this cycle, the focus relies on the construction of artifacts, their evaluation, and gathering feedback to refine further the design. Hevner (2007) stated that it is essential to maintain the balance between the construction and the evaluation of artifacts. Relevance and rigor must conclusively underpin both activities. Moreover, in the construction of the artifacts, grounded arguments could be insufficient with a weak evaluation (Hevner, 2007). Although, for the evaluation process itself, few guidelines are offered in the paper of Hevner (2007). For this purpose, the Framework for Evaluation in Design Science (FEDS), as described by Venable et al. (2016), is used. FEDS is a well-cited framework that consists of an explicit process of why, when, how, and what to evaluate.

In FEDS, two main distinctions are described regarding the functional purpose of evaluation. These distinctions rely on formative and summative evaluation. First, for *formative evaluation*, the purpose is to produce empirical-based interpretations that provide the basis for improving the characteristics of the research artifacts (Venable et al., 2016). As earlier noted, the research artifacts are the *HLA* and the *taxonomy of BMs*. To improve the characteristics of these research artifacts, semi-structured interviews are conducted with researchers and industry experts in the subject areas of data marketplaces, where various recommendations for the improvement of artifacts are gathered. Subsequently, with the improved artifacts, the purpose of *summative evaluation* is to produce empirical-based interpretations that provide a basis for creating shared meanings about the research artifacts (Venable et al., 2016). In this regard, the same researchers and industry experts are consulted to subtract both critical and supportive viewpoints, thus creating the shared meaning about the application of decentralized architecture, namely blockchain technology, in data exchange. Finally, the viewpoints of experts are aligned in the current scientific literature, concluding the development of the artifacts, which as a result, provide the technical and business knowledge to answer the main research question.

1.5. Research Process and Sub-Research Questions

With the established design since research approach, the thesis follows an association of three areas of investigation: decentralized architecture, business modeling, and data marketplaces as the subject of primary interest. Before researching the decentralized architecture and business models, data marketplaces have to be explicitly understood. This research phase entails the *technical background* that provides the knowledge base for the research of the design artifacts. Moreover, considering that the HLA is subsequently created, the ultimate goal of this phase is to define the infrastructure components that describe the primary functionalities of data trading platforms. For this purpose, the first sub-research question is:

What are the generic infrastructure components for data marketplaces?

Based on the infrastructure components from the *theoretical background*, their relationships in the decentralized architecture are analyzed in the following sub-research question. The deliverable of this question is the HLA, which represents the first research artifact that helps to understand the blockchain architecture and facilitates the research in distinguishing the functional requirements for data marketplaces. According to DSR, this phase marks the start of the *design cycle* with the creation of the HLA. For this purpose, the second sub-research question is:

How does a decentralized architecture for data marketplaces look like?

The research strategy applied in answering the second sub-research question is desk research. The desk research strategy aims to analyze the different literature sources on blockchain technology and data marketplaces to strengthen the causal inferences of the study.

Next, data marketplaces with a decentralized architecture also require sustaining comprehensive business models to maintain and create value for actors engaging in data exchange. As earlier noted, the second researched artifact in the thesis is the taxonomy of BMs for data marketplaces with a decentralized architecture. In compliance with DSR, this phase continues the *design cycle* by creating the taxonomy of BMs. The taxonomy aligns the business model characteristics of decentralized data marketplaces and provides the insights necessary to answer the third sub-research question:

What are the characteristics elements in a taxonomy of business models for data marketplaces with a decentralized architecture?

The case study is the research strategy applied in answering the third sub-research question. This strategy is used to have a practical approach in analyzing the business model characteristics for data marketplaces with a decentralized architecture. The case study is conducted on several market-active data trading platforms by examining their websites, whitepapers, platform blogs, terms and conditions through desk research.

With the designed HLA and the taxonomy of BMs from the second and third sub-research questions, these artifacts are evaluated in compliance with the FEDS framework (Venable et al., 2016). According to the research approach, the first evaluation process is defined as *formative evaluation*. This evaluation process focuses on producing empirical-based interpretations that provide the basis for improving the characteristics of the research artifacts (Venable et al., 2016). For this purpose, researchers and industry experts in the subject areas of data marketplaces are consulted to gather their recommendations on improving the artifacts. Thereby, first, the HLA is evaluated on *comprehensibility, logical structure, and completeness*. Subsequently, the taxonomy of BMs is evaluated on *completeness* and *detailedness*. For this purpose, the fourth and fifth evaluation questions are introduced as follows:

Is the decentralized architecture for data marketplaces evaluated as comprehensible, logically structured, and complete?

Is the taxonomy of business models for data marketplaces with a decentralized architecture evaluated as detailed and complete?

The research method chosen in answering the fourth and fifth evaluation questions is semi-structured interviews. Two aspects describe the semi-structured approach. The first *structured* aspect relies on pre-defined questions. Thereafter, the second *semi*-structured aspect represents the follow-up questions that appeal to an open discussion to clarify the responses of interviewees. Furthermore, these interviews are used to discuss possible incorrectly structured elements in the HLA, incomplete taxonomy characteristics, or vague research concepts. Therefore, by adding the different viewpoints of experts, the interviews reduce possible errors, clarify misunderstandings, and strengthen the generalized interference of the research.

According to the research approach, the last evaluation phase is the *summative evaluation*. The functional purpose of this evaluation phase is to produce empirical-based interpretations that provide a basis for creating shared meanings about the research artifacts. Critical and supportive viewpoints regarding the application of the decentralized architecture in data exchange are evaluated with the same researchers and industry experts. Furthermore, their viewpoints are further discussed and aligned in the current scientific literature, thereby creating shared meanings about the research artifacts. For this purpose, the final sixth evaluation question that concludes the development of the artifacts is:

What value is created for actors in data exchange with the decentralized architecture?

Similar to the formative evaluation, the research method used to answer the sixth evaluation question is semi-structured interviews. Although, in this evaluation phase, the interviews are structured based on an open discussion with pre-defined topics on decentralized architecture and data exchange. Therefore, in compliance with the sixth evaluation question, experts are consulted whether decentralized architecture provides more security for actors in data exchange, whether decentralized architecture provides more trust for actors in data exchange, and whether decentralized architecture provides more means of control in data exchange.

Finally, by answering each sub-research question and evaluation question, the necessary insights can be subtracted to answer the main research question in the conclusion of the current thesis.

1.6. Research Scope

The underlying thesis scope is to provide the fundamental knowledge for innovation and technology management practices. Thus, no in-depth technical details are analyzed. The research scope is limited to the technological level of analysis for the HLA and the business level of analysis for the BM taxonomy.

For a clear position of the current research, several aspects are essential to be mentioned in advance. These aspects define the boundaries and the scope of the study. First, throughout the research, *data is viewed as intermediate goods* that may hold sensitive information or not. The analysis regarding the classification of data is out of scope. Second, data marketplaces are analyzed as *multisided marketplaces* where more than a single provider or consumer can exchange data. This exchange process is later described as a many-to-many matching mechanism for the demand and supply of data. Third, an actor participating in data exchange could be characterized as a commercial business entity, private non-commercial organization, or governmental institution. The research focuses only on *Business-to-Business (B2B)* data transactions. Other relations as Business-to-Government (B2G) or Business-to-Customer (B2C) are out of scope. Forth, the analysis of decentralized architecture focuses on *blockchain technology* where the agreements between data providers and consumers are specified in self-executable smart contracts. Other decentralized or distributed architectures are out of scope.

1.7. Research flow diagram

According to the research approach from section 1.4 and in compliance with the research process from section 1.5, the research flow is described in figure 1. On the basis of DSR, the construction activities for the two artifacts are structured following the relevance, design, and rigor cycles. Therefore, the primary components in the research flow diagram are the main design process that outputs the design artifacts. Moreover, as earlier mentioned, another key activity in DSR is the evaluation of the design artifacts, as it provides input for further development; thereby, the formative and summative evaluations are included according to the FEDS framework.

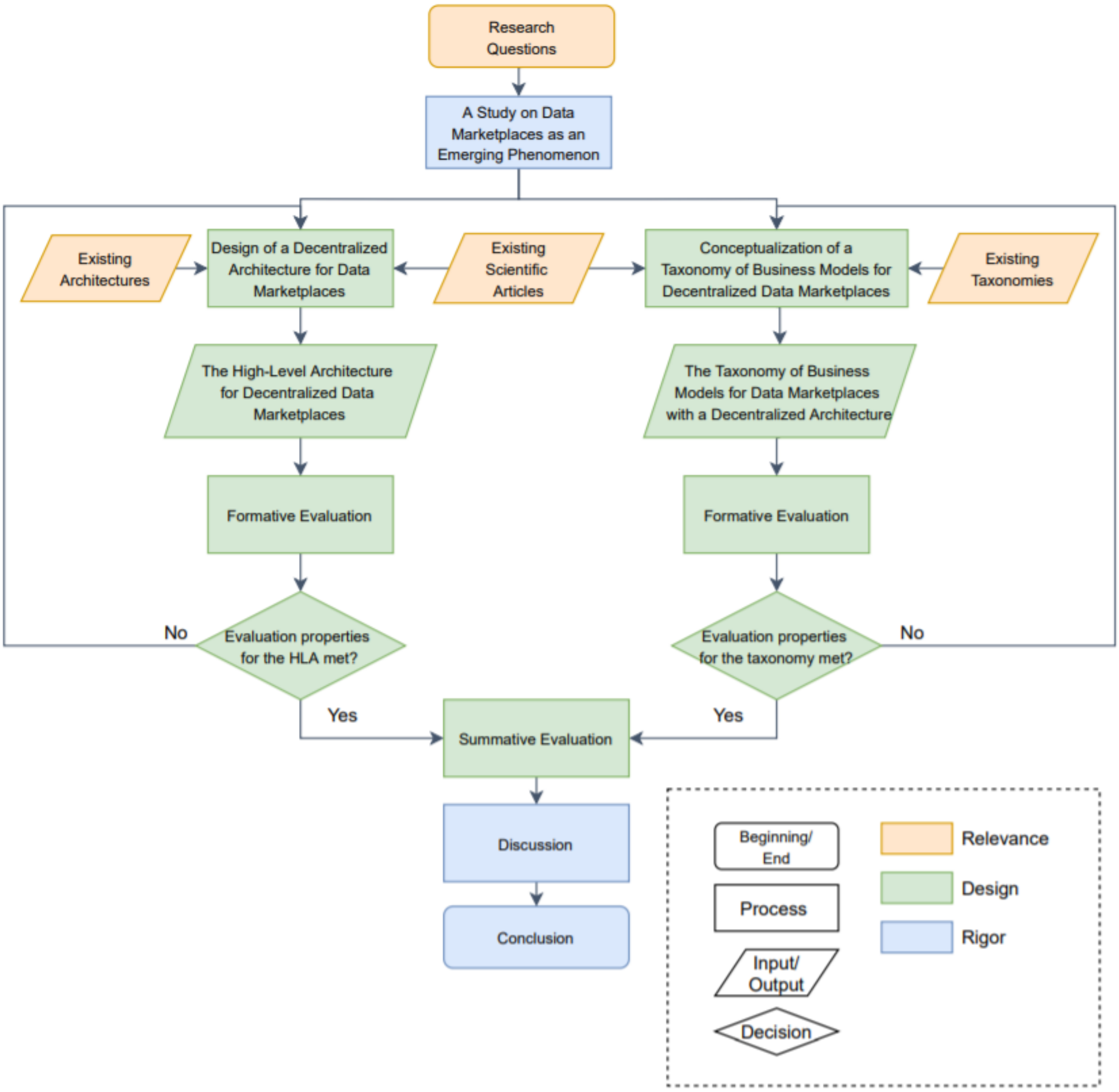

Figure 1: Research flow diagram

1.8. Reading guide

The remainder of the thesis is structured as follows. A study on data marketplaces as an emerging phenomenon is introduced in chapter 2. In this chapter, the theoretical background regarding data marketplaces is described to establish the theoretical foundation for data marketplaces. Next, according to the theoretical foundation from chapter 2 in the following chapter 3, the technical specifications of the decentralized architecture are analyzed. The ultimate objective of this chapter is to design the HLA, thereby creating the first research artifact. In chapter 4, the research is focused on the business specification for data marketplaces. For this purpose, previous studies on the taxonomy of business models for generic marketplaces are used to conceptualize a taxonomy for decentralized marketplaces. Furthermore, in this chapter, the business model characteristics are explored in the documents of market-active data trading platforms. As a result, this chapter delivers a taxonomy of business models for data marketplaces with a decentralized architecture. Further, on the basis of the designed HLA and the taxonomy of BMs, these two artifacts are evaluated with researchers and industry experts in chapter 5 to gather their recommendations on improving the artifacts. Subsequently, in chapter 6, with the improved artifacts, the same researchers and industry experts are consulted to create the shared meaning about the application of the decentralized architecture in data exchange practices. Lastly, in chapter 7, in compliance with the designed and evaluated artifacts, the research results are discussed by answering each sub-research question leading to the conclusion in chapter 8, where the main research question is answered. Thereafter, in chapter 9, a reflection of the thesis is presented with the scientific contribution, research limitation, and recommendations for future research.

2
A Study on Data Marketplaces as an Emerging Phenomenon

The thesis research follows an association of three areas of investigation: decentralized architecture, business modeling, and data marketplaces as the primary study interest. Data marketplaces are still emerging digital platforms in the data industry. Thus, before researching the decentralized architecture and business models, data marketplaces have to be explicitly defined.

Numerous scholars have studied the possibility of trading data assets through intermediary digital marketplaces. Thus, considering their work, a comprehensive understanding of data marketplaces and the working definition of these species of digital platforms are established in this chapter. Moreover, considering that the HLA is subsequently created in the following chapter 3, the ultimate goal is to determine the infrastructure components that describe the primary functionalities of data trading platforms, thereby answering the first sub-research question: *What are the generic infrastructure components for data marketplaces?*

The chapter is structured as follows. The literature search conducted on data marketplaces is introduced in section 2.1. Next, the concept of data as economic assets is characterized in section 2.2. Thereafter, data marketplaces are defined in section 2.3. And lastly, the classifications of data marketplaces and the data-trading frameworks are described in sections 2.4 and 2.5. The chapter concludes in section 2.6 with the definition of the generic infrastructures components for data marketplaces.

2.1. Literature Search and Selection Criteria

The literature search for this chapter was focused on retrieving the scientific articles of different scholars that researched the main characteristics of data marketplaces. Thus, in compliance with the chapter objective, three aspects are essential in the literature search. First, to identify the challenges of exchanging data. Second, to investigate the data trading mechanisms. And third, to describe the different data-trading frameworks for data marketplaces.

The topic of data marketplaces is a known research concept within TU Delft. Therefore, the initial literature search was performed on the university repository. The search listed five relevant master theses. These theses revealed engaging insight on business model taxonomies, multi-party computation for the privacy-preserving exchange of data, data trading structures, and the conceptual architecture for data marketplaces. After acquiring the fundamentals of data marketplaces from previous master theses, the literature search followed with exploring the scientific databases. Several databases for scientific articles were queried. Among these are Scopus, Google Scholar, ResearchGate, and IEEE Xplore. Furthermore, different combinations of structured queries were performed to identify the whitepapers focusing on researching the data marketplaces. The searching queries used a sequence of keywords and corresponding synonyms.

The first searching iteration of scientific articles with the keyword "data marketplaces" revealed a significant amount of 175 papers on Scopus. Logical operators in combination with keywords were applied to narrow the query results. For instance, several combinations used were: ("data marketplace" AND "data economy"), ("data marketplace" AND "data goods"), ("data trading" AND "data market"), ("data-driven business models" AND "platforms"), etc. This approach produced a lower number of publications with more specific results making it possible to estimate the relevance by scanning through the titles and abstracts of publications. Eight scientific articles were selected for more detailed analysis. The literature search continued on Google Scholar to identify the articles that applied a case study research on the current markets for data. The articles found described the challenges of open markets in trading data as intangible goods. Furthermore, these articles revealed several market frameworks with real examples of matching mechanisms for data trading. With Google Scholar search, three more scientific articles were added to be examined. Subsequently, ResearchGate has been queried for additional publications. The articles found researched the application of data marketplaces in various fields, such as the Internet of Things, Smart Cities, Smart Healthcare, and Enterprise Supply Chain. These articles were useful in understanding the opportunities which arise with open data-driven platforms. Additionally, two more articles were considered to be investigated. The final literature search was based on the engineering aspect of data marketplaces. For this purpose, the IEEE Xplore digital library was queried to retrieve technical articles for data marketplaces. Different studies described the concepts of data-intensive platforms, digital markets for trading data, and decentralized data marketplace. In this context, three articles were selected. In total, sixteen papers were chosen after the completion of search iterations. According to the process previously described, these papers were found by filtering through scientific databases using various combinations of keywords. The inclusion and exclusion of papers were based on carefully scanning the publications according to their title and abstract to estimate the relevance. In cases where more riding was required, the introduction with the main findings was considered. Six papers were chosen as the main references in this chapter. The remaining papers were allocated to complete the literature search of the following chapters 3 and 4, where the decentralized architecture and the business models are investigated. The chosen articles are described as follows.

- The first chosen article represents the work of *Schomm et al. (2013)* as one of the first scientific research papers on data marketplaces. This scientific paper provides a taxonomy for classifying different data vendors.
- The second paper is *Stahl et al. (2015)*, one of the pioneers analyzing data marketplaces through a case study that describes the common principles of exchanging data.
- The third and fourth papers are *Koutris et al. (2012)* and *Muschalle et al. (2012)*, focusing on the data pricing models to reach most consumers.
- The fifth selected paper is based on the research of *Koutroumpis et al. (2017)*, which characterizes the challenges in data quality assessment, describes the market design principles, and analyzes the data-trading frameworks by introducing their technical characteristics.
- The sixth paper constitutes the work of *Spiekermann M. (2019)*. This author conducted a case study on data marketplaces to describe a taxonomy for the business model analysis and illustrated the position of different actors engaging in data exchange through a role model schema.

2.2. Data as Economic Assets

Data are rarely valuable alone and are generally input into analytics to produce insights that can be expressed as content-based information goods (Koutroumpis et al., 2017). For instance, companies such as Amazon or Google collect and aggregate data to further commercialize it through their services to third parties. These third parties may use the insights from the aggregated data to analyze market trends or perform other business activities. Therefore, data can be viewed as intermediate goods that need to be further processed and analyzed with complementary technologies for consumers to gain utility (Koutroumpis et al., 2017). Although creating utility with data requires shifting from assessing data as products to assessing data as problem solutions. Numerous aspects are essential to be researched for this purpose. In particular, considering the nonexisting property rights of intangible products like data, the quality assurance and pricing models are two governance mechanisms important in commercializing data assets (Teece, 2010).

2.2.1. Assessing the Quality of Data Assets

As intermediate and intangible goods, scholars described the challenges arising in verifying the quality of data. Koutroumpis et al. (2017), expressed that the quality of intangible assets as data is not observable before consumption. The quality could be assessed only after a longer period of usage or by comparing statistical properties against similar datasets. In estimating the data quality, traditional mechanisms such as reputation systems may be insufficient. A factor for assessing the data quality could be accomplished through comprehensive information about the *provenance*, *collecting methods*, and *data sources*. These benchmarks can be integrated into the metadata information as a proxy factor for quality evidence (Koutroumpis et al., 2017).

In the same context, Sharma P. et al. (2020), suggested that metadata represents the description of the data but does not completely characterize the quality of datasets. The data quality can be divided into two criteria: *objective criteria* and *subjective criteria*. On the one hand, *objective criteria* were already specified as provenance, collecting methods, and data sources. On the other hand, *subjective criteria* represent the relevancy of data according to consumer requirements (Sharma et al., 2020).

Breaking down the arguments of the two scholars, the data quality assessment as intangible assets is a complex task that may consider the following aspects. First, *metadata is an essential reference point* both for data providers and consumers. Through metadata, providers can describe the supplied datasets, and consumers can trace these specifications. Second, *data provenance*, *collection methods*, and *data sources* are essential properties of metadata. These properties represent the objective criteria for assessing the quality of data assets. Although, the quality of data could also be verified through *subjective criteria according to consumer requirements*. Lastly, *statistical quality assessment* of data against similar datasets could be used to enhance the quality assessment.

2.2.2. Pricing Data Assets

Another challenge in exchanging data as intermediate goods is pricing these intangible assets (Teece, 2010). Again, pricing mechanisms could be directly dependent on the quality of data. The principles for ensuring the primary characteristics of data quality were previously described (see section 2.2.1). Although, besides quality, data providers should also cautiously determine the pricing models that reach the most consumers. The straightforward pricing approach is assigning a *flat price*. However, enriched models for pricing data could also be used.

One of the pricing models is query-specific views. This model implies pricing data according to views consisting of selection queries based on a relation attribute (Koutris et al., 2012). This means that data providers can assign prices to specific attributes that are consequently used in querying datasets. Further, consumers can decide between purchasing entire datasets or only the necessary subsets derived according to selected attributes. This mechanism characterizes the *package pricing model*, where providers offer a certain amount of data for a predeterminate price. The *package pricing model* provides the flexibility for consumers to access only the necessary set of data by specifying the querying attributes. However, these attributes can become complex when consumers are given more choices to query the dataset (Koutris et al., 2012). Thus, providers or marketplace owners should specify clear guidelines needed in accessing datasets. Similar guidelines are required for setting prices to ensure good proprieties for querying attributes. Therefore a good practice for marketplace owners is to assist data providers in setting prices for attributes so that inconsistency in pricing does not result in higher-priced attributes compared to buying the whole dataset.

Another pricing model is providing data according to the time frame as the single attribute for costs calculation. For instance, data providers can specify monthly price subscriptions to their datasets. This approach is characterized as a *flat fee tariff model*, known for different services such as licensing or hosting (Muschalle et al., 2012). The model ensures providers and consumers with a simple schema for planning future revenues or expenses, considering that time is a linear calculation variable. However, due to the linear factor, the flat fee tariff model lacks flexibility. Therefore, consumers should bear in mind their preference for datasets to purchase only the required data so that they do not end on subscribing to a dataset that does not meet their requirements. To ensure more flexibility, data providers may offer short-term contracts with the *flat fee tariff model* c.

The combination of package and flat fee tariff models forms the *two-part tariff model*, where the costs are divided into fixed basic fees, and additional fees per unit consumed (Muschalle et al., 2012). For the *two-part tariff model*, the basic fees cover the fixed costs of the providers in generating data and squeeze profit from additional per unit fees. The *two-part tariff model* could be a flexible and convenient payment method for consumers. However, calculating the separation of the basic fees and additional per unit fees requires a comprehensive estimation of the number of consumers accessing the datasets.

A derived model from the two-part tariff is the *freemium model*. This model offers the basic services for free and charges for the premium services, which provide additional value to consumers. The *freemium model* could be used to ensure a large attendance of the data consumers. However, this model can only work if the marginal costs are minimal, otherwise per unit losing could get out of control (Muschalle et al., 2012).

The four models, *package pricing*, *flat fee tariff*, *two-part tariff*, and *freemium*, were observed in practice for multiple data vendors in a case study research by Muschalle et al. (2012). The study described two additional models as *free pricing* or *usage-based pricing*. The free pricing model implies offering data for free, and the usage-based pricing indicates that each unit of a commodity raises data costs. However, the *free pricing model* is commonly used by public authorities and does not provide commercial incentives for the B2B data exchange. In the same context, the *usage-based model* has difficulties generating profit for providers considering the falling marginal costs of consumed data. Thus, data providers may disregard the usage-based model as the first choice in the pricing strategy (Muschalle et al., 2012). Moreover, considering the described strengths and weaknesses, data providers should carefully choose the pricing models by analyzing the characteristics of their datasets. In case of uncertainties, competitive traditional *flat prices* can offer clear revenues or insights into the market by analyzing the responses of consumers to provided datasets.

2.3. Defining Data Marketplaces

The definition of data marketplaces may encounter changes with the advancement of data management and related technologies. Therefore, the current thesis establishes a generic definition to describe the main characteristics of data marketplaces. Fricker and Maksimov (2017) argued that data marketplaces are digital platforms on which datasets can be offered and accessed. Data marketplaces enable data exchange by providing the infrastructure for finding, buying, and selling data, as well as obtaining access to their providers (Fricker & Maksimov, 2017). Although in principle, this describes the general functionalities of data marketplaces, it does not include the regulatory aspect for data quality, ownership, and privacy. In this context, Schomm F. et al. (2013) stated that the access and manipulation of data could be regulated through different licensing models (Schomm et al., 2013).

Theoretically, every data set should fall under a specific license describing the terms and conditions that explicitly regulate how data may be used. Moreover, actors participating in data exchange should have appropriate access rights to their resources through identification, authentication, and authorization processes. In the same context, another important aspect for data marketplaces is the cooperation of different actors to enable the network effect. This would enhance the market efficiency and improve the match between the supply and demand of data (Koutroumpis et al., 2017). For a better description of cooperation among actors, Spiekermann M. (2019) divided them into four categories: data providers, data buyers (consumers), third-party service providers, and market owners. Providers offer their data to be commercialized by the consumers querying the marketplace willing to purchase necessary datasets. Third-party service providers are skilled actors able to deliver analytic services that add value to datasets. Lastly, the marketplace owners develop the infrastructure required for data exchange to occur. Therefore, by evaluating the different characteristics of the data marketplace, a common definition can be shaped as follows.

Data marketplaces can be defined as digital platforms that offer the infrastructure for data exchange among different actors. Based on their activities in the marketplace, actors are classified as data providers, data consumers, third-party service providers, and marketplace owners. The access, manipulation, and usage of data could be regulated through established licensing models. The final purpose of data marketplaces is to enable the network effect by facilitating the transaction mechanisms for an efficient and trustworthy match between the supply and demand of data.

2.4. The Classification of Data Marketplaces

According to the definition from the previous section, data marketplaces facilitate the transaction mechanisms for an efficient and trustworthy match between the supply and demand of data. Conceptually, matching data supply and demand is not different from any other type of digital market (Koutroumpis et al., 2017). Therefore, the analysis of different matching mechanisms may enable the classification of data marketplaces. This classification could provide an explicit structure to describe the characteristics of the different types of data marketplaces. In this section, the challenges of growing the user base and ensuring safety in data exchange are investigated according to each type of identified data marketplace. Koutroumpis et al. (2017) derived the matching mechanisms of supply and demand of data into four categories of data marketplaces. The categories with the terms of exchange and real marketplace examples are described in the following table 1.

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

Table 1: The classification of Data Marketplaces by matching mechanism (Koutroumpis et al., 2017)

According to Table 1, the first marketplace classification is the *ono-to-one matching* mechanism. The one-to-one matching mechanism involves a single provider and a single consumer exchanging data via marketplaces on bilateral negotiations. In this matching mechanism, marketplaces can focus on contracts enforcement through tight monitoring, ensuring the safe execution of data exchange (Koutroumpis et al., 2017). Although in addressing a large number of actors, the *one-to-one* *marketplaces* could be inefficient due to the extensive efforts necessary for monitoring the tight contracts or finding trading parties. As a result, the extensive efforts directly trigger higher costs, longer searching processes, and continuous relationship management.

The second matching mechanism describes the *one-to-many relationship*. In this classification, a single data provider exchanges data with multiple consumers. The safety of exchanging data could be ensured through standardized terms of exchange to avoid extra costs in managing the relationship with each consumer. Data marketplaces frequently use the *one-to-many matching* relationship (Koutroumpis et al., 2017). However, the standardization of these marketplaces may lack efficient monitoring of data quality. Furthermore, the *one-to-many marketplaces* require intensive branding efforts in the competitive environment to establish a large user base.

The third matching relationship involves multiple providers enabling their data to a single consumer. In this case, the consumer is the marketplace platform. The efficiency of the *many-to-one relationship* depends on the popularity of the data marketplaces. Popular many-to-one marketplaces could collect large amounts of data by providing free services for a large user base. However, free services usually involve the absence of negotiation and relationship management, thus affecting data quality. For instance, providers could supply biased or irrelevant data as a result of a deficient agreement.

The fourth *many-to-many matching* relationship describes the multisided marketplaces where data is offered by multiple providers and purchased by various consumers. Therefore, to ensure the safe exchange process, a regulatory environment is required in the case of multisided marketplaces (Koutroumpis et al., 2017). The regulatory environment could the accomplished through licensing models that specify the data usage terms (Stahl et al., 2015). Furthermore, data could have a domain-specific industry (e.g., automotive industry-specific datasets) or any industry domain in the *many-to-many marketplaces*. Therefore, considering that multiple actors can exchange data, marketplaces should balance the supply and demand of data. This could be accomplished through supporting long-tail sales. The concept of long-tail sales was introduced by Anderson (2006), who argued that products with low sales or supply could collectively make up the market share, directly rivaling well-established marketplaces that exclude these categories of products due to their low demand. For domain-specific data, long-tail sales could be achieved by matching a multitude of actors with time-stable datasets to increase usability and value creation. Subsequently, for all industry data, long-tail sales could be achieved by segmentation of data pools, thereby curating available datasets to create problem solutions in different use cases. These aspects led to the observation that for many-to-many data platforms to have a greater chance to succeed, it takes an ecosystem to be established. Moreover, the ecosystem could focus not only on data providers and consumers but also on the service providers as data storage, analytics, aggregation, or transformation, to enable the network effect.

According to the scope of the research, the thesis focuses on *multilateral data marketplaces* with a *many-to-many matching* mechanism. Although, for a comprehensive understanding of data marketplaces as an emerging phenomenon, the *one-to-one*, *one-to-many*, and *many-to-one* matching mechanisms also described fundamental insights for the research. The analysis shows that bilateral *one-to-one* marketplaces stand out with tight contracts that can ensure the safe execution of data exchange. However, monitoring the execution of tight contracts introduces high costs, long searching processes, and continuous relationship management. Standardized contracts described in the *one-to-many* relationship could reduce the costs in relationship managing while appealing to a larger user base. However, standardized contracts influence the efficient monitoring of data quality. Lastly, the absence of contracts in the *many-to-one* relationship can eliminate the barriers to harvest large amounts of data. However, providers could supply biased or irrelevant data due to deficient agreements affecting data quality again. The analysis of the *one-to-one*, *one-to-many*, and *many-to-one* matching mechanisms revealed that it is possible to achieve large markets with less control or small markets with greater control (Koutroumpis et al., 2017). Based on these analyses, in the case of *many-to-many marketplaces*, where multiple actors exchange data, there should be a degree of control ensuring data protection for providers and quality assurance for consumers. This control could be accomplished through *licensing models* that specify the data usage terms in the exchange process. Moreover, multilateral marketplaces should *incentivize the service providers that offer hard-to-find data* to contribute to the data platform. As a result, their collective data assets *would support the long tail of data sales and increase the data market share*. This approach may appeal to an extensive consumer base by introducing practical uses cases sustained with a large pool of data assets. In this context, marketplaces require a *well-established ecosystem of data service providers*. For this purpose, the trading frameworks are another essential aspect of research for multilateral data marketplaces to determine the contribution of different actors in the ecosystem.

2.5. Data-Trading Frameworks

Regarding multilateral data marketplaces, a clear distinction can be made between *centralized* and *decentralized* architectures (Spiekermann, 2019). Moreover, the data management responsibilities in the marketplace ecosystem differ due to the distinct characteristics of the trading frameworks. The differences could be generally associated with controlling the storage, custody, and manipulation of data assets. In the following, these differences are explained according to the centralized and decentralized multilateral data marketplaces.

Centralized multilateral data marketplaces reflect the standard design of multisided platforms. In this framework, marketplace owners are in direct control of the marketplace operations. The owners create the network effect through efficient services offered on their platforms to attract data providers and consumers. In this context, Spiekermann (2019) described the role model of the data marketplace ecosystems according to the following figure 2.

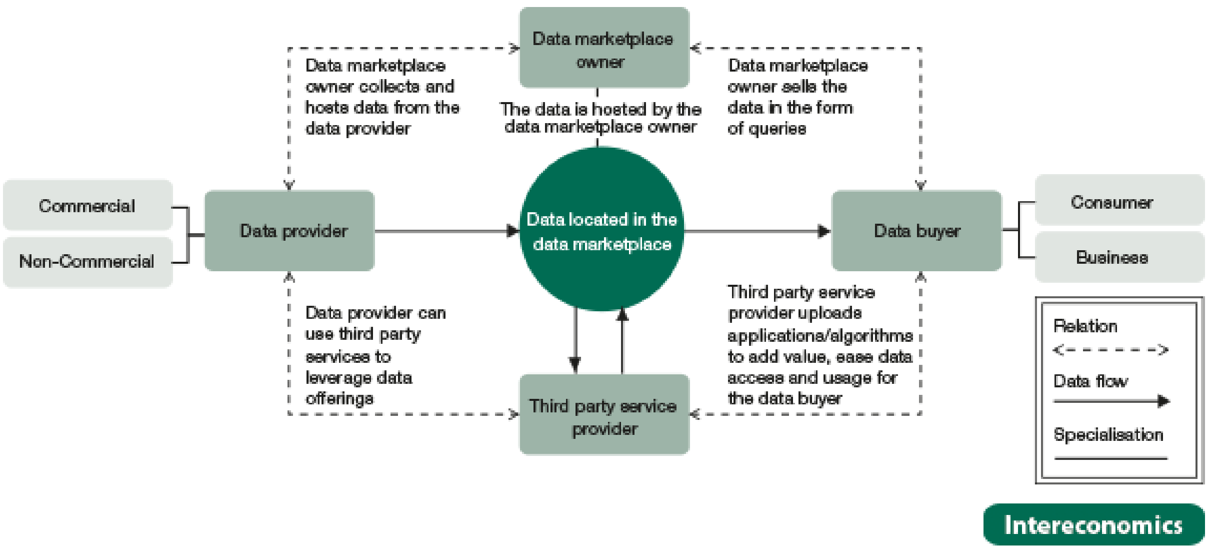

Figure 2: The role model of centralized data marketplace ecosystems (Spiekermann, 2019)

According to the role model described by Spiekermann (2019), data flows from providers to consumers through the marketplaces, where marketplace owners collect, host, and sell data. In this context, the marketplace owners are directly responsible for providing the services to incentivize different actors in exchanging data through their platforms. However, if services are considered disadvantageous or if the trading costs are higher than the market expectation, the marketplace owners would encounter difficulties operating the centralized multilateral data marketplaces (Koutroumpis et al., 2017). Thus, extensive efforts are required from the marketplace owners to accomplish the long tail data sale and establish the growing ecosystem of data providers, consumers, and third-party services, as previously explained (see section 2.4).

Decentralized multilateral data marketplaces are an alternative architectural approach for data trading platforms based on decentralized technologies. Contrary to the centralized data marketplaces in the decentralized architecture, data is exchanged without intermediaries as marketplace owners. According to the research scope (see section 1.6), blockchain technology is the researched decentralized architecture in the thesis. In the blockchain architecture, the exclusion of intermediaries is accomplished by communicating and validating transactions in blocks recorded on a distributed ledger (Zheng et al., 2018). This ledger is available to every participant in the marketplace. Thus, the decentralized architecture eliminates by design the need for a central entity to collect and host data, enabling peer-to-peer transactions. Furthermore, the information about the *timing*, *quantity*, *value of transactions* is also publicly registered, ensuring transparency in data transactions (Koutroumpis et al., 2017). Therefore, considering the technical characteristics of a decentralized architecture, figure 3 presents the changes in the role model of Spiekermann (2019) to reflect the described particularities of decentralized multilateral data marketplaces.

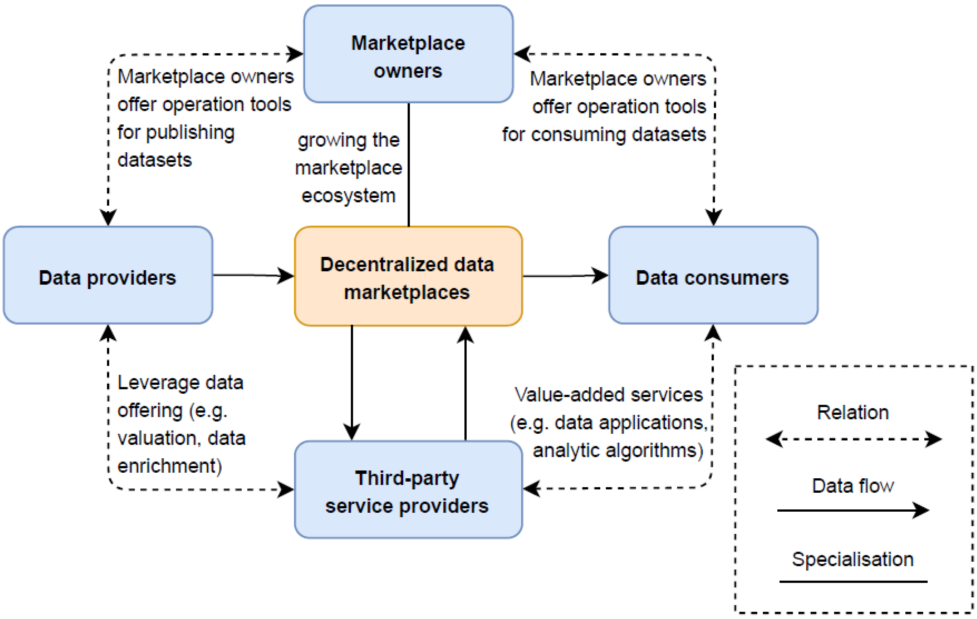

Figure 3: The role model of decentralized data marketplace ecosystems

According to Koutroumpis et al. (2017), what was previously called a centralized marketplace in the decentralized design is a communication structure that facilitates the operations of the decentralized market (Koutroumpis et al., 2017). Therefore, the activities of marketplace owners to collect, host, and sell data also shift towards offering operation tools in the data market. These operation tools are later explored in more detail in the following chapter 3. Similar to centralized marketplaces in the decentralized ecosystem, challenges in the commercialization of intangible products could be encountered. Consumers want more than just products; they want solutions to their needs (Teece, 2010). Therefore, the responsibility of the marketplace owners should also focus on continuously improving the platform and growing the data marketplace ecosystems.

2.6. Conclusion of Chapter 2

Data marketplaces are still emerging digital platforms in the data industry. Therefore, before researching the decentralized architecture and business models, data marketplaces had to be explicitly understood. For this purpose, this chapter researched several aspects applicable to the data industry and particularly to data trading platforms. These aspects are summarized as follows, thereby also describing the infrastructure components analyzed in chapter 3 according to blockchain technology.

In this chapter, the first aspect was related to viewing data as economic assets. Data are rarely valuable alone and need to be *further processed and analyzed with complementary technologies* for consumers to gain utility (Koutroumpis et al., 2017). Moreover, considering the nonexistent property rights of intangible products like data, the *data quality assessment* and *pricing models* are two governance mechanisms important in commercializing data assets (Teece, 2010). For *data quality assessment*, the provenance, collection methods, and data sources may represent an essential quality indicator for consumers (Koutroumpis et al., 2017). Subsequently, for *pricing models*, enriched models for pricing data could be used by data providers to reach the most consumers (Sharma et al., 2020). Next, through metadata, providers can describe the supplied datasets, and consumers can trace these specifications. In this regard, the metadata could enable the *query and search* functionalities for data marketplaces (Sharma et al., 2020). Therefore, concluding the first aspect related to viewing data as economic assets, the following generic infrastructure components could be defined for data marketplaces: *data process and analytics; data quality assessment; pricing mechanism;* and *query & search.*

The second aspect researched in this chapter was associated with defining data marketplaces. Technically, data marketplaces offer the infrastructure for *data exchange* among different actors as their general functionality (Fricker & Maksimov, 2017). Moreover, to ensure the integrity of the exchange process, marketplaces should also provide mechanisms for data *storage, custody, and manipulation*. This could be regulated through established licensing models (Stahl et al., 2015). Lastly, the actors participating in data exchange were characterized as data providers, data consumers, third-party service providers, and marketplace owners. These actors should have appropriate access rights to their resources through identification, authentication, and authorization processes. This could be accomplished through *identity management* mechanisms. Therefore, concluding the second aspect related to defining data marketplaces, the following generic infrastructure components could be described: *data exchange; data storage, custody, manipulation; and identity management.*

The third aspect researched in this chapter corresponded to classifying data marketplaces and describing their data-trading frameworks. For classifying data marketplaces, the matching mechanisms of data supply and demand were analyzed. Data marketplaces were divided into four categories. These categories are one-to-one, one-to-many, many-to-one, and *many-to-many* marketplaces (Koutroumpis et al., 2017). The analysis showed that one-to-one marketplaces stand out with tight contracts ensuring safe and close monitoring of data exchange. However, this introduces high costs, long searching processes, and continuous relationship management. The costs can be reduced with standardized contracts in one-to-many marketplaces, but standardization influences data quality monitoring. For many-to-one marketplaces, in the absence of contacts, large amounts of data can be collected. However, providers could supply biased or irrelevant data, affecting data quality again. For multilateral (many-to-many) marketplaces, considering that multiple actors can exchange data, marketplaces should provide a regulatory environment for data usage and balance the supply and demand of data.

A clear distinction was analyzed between the centralized and decentralized approaches in the architecture of multilateral data marketplaces. The centralized multilateral marketplaces reflect the standard architecture of multisided platforms. In centralized marketplaces, the owners collect, host, and sell data (Spiekermann, 2019). Therefore, the marketplace owners are directly responsible for offering services to attract providers and consumers to their platforms. Thus, extensive efforts are required from the marketplace owners to grow the ecosystem of data providers through efficient services and competitive market prices. The alternative for the centralized architecture is the decentralized approach. In the decentralized architecture, data is exchanged without intermediaries as marketplaces owners. The transactions are transparently recorded on the public ledger that is distributed in the network enabling peer-to-peer exchange and eliminating the need for a central entity to collect and host data. However, decentralized data marketplaces could encounter similar challenges as centralized marketplaces regarding the commercialization of intangible products as data. In this research section, no new infrastructure components were subtracted. Although, according to the DSR methodology, this section aimed at strengthening the *relevance cycle* by introducing the opportunities with the decentralized architecture, namely the possibilities of the blockchain technology.

Concluding the first sub-research question, *What are the generic infrastructure components for data marketplaces?* With the determined challenges of data as economic assets and the data marketplace definition, the following generic infrastructure components could be described: *data process and analytics; data quality assessment; pricing mechanism; query & search; data exchange; data storage, custody, manipulation; and identity management.* These infrastructure components are further analyzed in the following chapter 3 according to the properties of the blockchain technology.

3

Designing a Decentralized Architecture for Data Marketplaces

This chapter marks the start of the design cycle, aiming to construct the decentralized HLA for data marketplaces. The HLA forms the design artifact, answering the second sub-research question: *How does a decentralized architecture for data marketplaces look like?* In this chapter, the infrastructure components from the conclusion of chapter 2 are analyzed according to the characteristics of blockchain technology and ultimately integrated into the HLA.

The remainder of the chapter is structured as follows. The literature search conducted on decentralized architecture for data marketplaces is described in section 3.1. Next, the characteristics of blockchain technology are researched in section 3.2. Thereafter, the infrastructure components for data marketplaces are analyzed in section 3.3. And lastly, the HLA is designed in section 3.4.

3.1. Literature Search and Selection Criteria

The literature search for this chapter was focused on identifying two categories of scientific articles. In the first category were the articles researching blockchain technology, and in the second category were the articles analyzing the application of blockchain technology in the context of data marketplaces. The two search processes were interconnected. This means that if blockchain concepts were weakly described in the articles analyzing the architecture of data marketplaces, then more research was assumed in the blockchain-based articles.

In searching for literature, similar to chapter 2, Scopus, Google Scholar, ResearchGate, and IEEE Xplore were the scientific databases used. The first search iteration performed on Scopus with the "blockchain technology" keyword resulted in a significant amount of articles. The two most cited articles describing a survey on blockchain technology were selected. The purpose of these two articles was to explore the general characteristics of blockchain technology. Next, considering the properties of infrastructure components described in section 2.6, Google Scholar, ResearchGate, and IEEE Xplore were queried to retrieve the technical papers researching the application of blockchain-specific functionalities. Different keyword combinations were used: ("extracting intelligence" AND "blockchain"), ("blockchain technology" AND " data use contract"), ("data indexing" AND "blockchain technology"), ("blockchain technology" AND "token economy"), etc. The need for querying the three scientific databases was related to the rights of downloading the articles. For instance, articles could be identified in Google Scholar, although only available to be downloaded in ResearchGate. With the literature search from this phase, additional four articles were selected for blockchain-specific functionalities that complied with the analysis of the infrastructure components for data marketplaces.

As it was previously explained in section 2.1, several articles from the chapter 2 literature search were allocated to complete the literature search for this chapter. Section 2.1 characterized the first literature search, which clearly revealed more articles than the research scope of chapter 2. Therefore once identified they were maintained for further analysis. Two articles were selected as research references for blockchain applications in the context of data marketplaces from the literature search of chapter 2. Lastly, another search iteration was performed on Scopus to find articles describing different conceptualizations of decentralized architectures. Keyword combinations as ("data marketplaces" AND "decentralized architecture") or ("data marketplaces" AND "blockchain architecture") were used. By analyzing the proposed architectures of scholars, the ones that provided well-detailed architectures were selected to represent the reference for architecture design. Thus, four more articles were selected for the design of the decentralized HLA.

In total, for this chapter, with the two blockchain technology surveys, the four articles describing the blockchain-specific functionalities, the two decentralized marketplaces articles, and the three articles that proposed a decentralized architecture for data marketplaces, eleven papers were selected as the primary reference for the analysis. These articles are briefly described as follows.

- First, the two surveys used for exploring blockchain technology are *Zheng et al. (2017)* and *Dinh et al. (2018)*. According to Scopus, these are the most cited blockchain surveys, presenting a comprehensive overview of blockchain technology, technical challenges, and smart contracts.
- Second, the four articles describing the blockchain-specific functionalities are *Lim et al. (2018)*, *Third & Domingue (2017)*, *Ramachandran et al. (2018)*, and *Hardjono & Smith (2016)*. These articles were used to research in more detail the functionalities of infrastructure components.
- Third, the two articles used in researching the application of blockchain technology in the context of data marketplaces are *Koutsos et al. (2020)* and *Hynes et al. (2018)*. These articles have a DSR approach demonstrating two practical implementations of blockchain-based data marketplaces.
- Fourth, the four articles used in designing the HLA are *Garcia-Font (2021)*, *Zhang et al. (2018)*, *Özyilmaz et al. (2018)*, *Ramachandran et al. (2018)*. These studies and their architectures have a domain-specific analysis (e.g., IoT, smart cities) for the data marketplaces. Although a more generic approach is considered in the current thesis, thus ideas were subtracted from these articles and adjusted according to the thesis scope.

3.2. Blockchain Technology

Since the first conceptualization of Bitcoin in 2008, blockchain technology has received significant attention (Zheng et al., 2017). Blockchain-based applications have been researched in various fields, such as financial services, the Internet of Things (IoT), supply chain management, and data marketplaces. Technically, blockchain could be regarded as a distributed ledger on which transactions between different actors are registered in a chain of information blocks. The chain of blocks expands when new transactions are recorded. Furthermore, cryptographic and distributed consensus algorithms have been implemented to ensure user security and consistency in the distributed ledger (Zheng et al., 2017). The cryptographically connected chain of information blocks recorded in a distributed ledger makes blockchain immutable and transparent. Businesses could leverage these characteristics to enable reliable peer-to-peer transactions. More specifically, users can trace the transaction history of other actors and estimate their honesty in various business use cases, making blockchain auditable.

The distributed ledger in blockchain excludes by design the need for intermediaries to host or manage transactions. Thus, with no intermediaries, applications developed on blockchain avoid the single point of failure. This could save costs in operations, enhance persistency in processing transactions, and strengthen system resilience towards malicious attacks (Dinh et al., 2018).

Decentralization, immutability, transparency, and security are a couple of blockchain characteristics ensured by design. However, the technology also encounters several technical challenges. The first challenge is related to scalability. Appending new transaction blocks to the distributed ledger requires computational power to validate the cryptographic connection of the blocks. Thus, blockchain needs a well-established network to support the technology. The second challenge is related to the size of blocks distributed in the network. Larger blocks could store more transaction information enabling higher frequency as the network would have fewer requests in appending blocks. However, this could affect the concept of decentralization as more transactions would be calculated by a single actor. Finally, the third challenge is associated with the consensus protocols responsible for validating transactions. Therefore, considering the two common protocols, *proof of work* wastes too much electricity, and *proof of stake* faces the problem of equivalent validations distribution (Zheng et al., 2017).

3.2.1. Blockchain Architecture

As previously described in the introduction of section 3.2, blockchain can be regarded as a distributed ledger. The distributed ledger is transparently revealed to the network, and participants can validate its content. Technically, the ledger records the complete sequence of blocks that contain information about the transactions. The connection between blocks is accomplished by registering the unique hash value of the previous block named the parent block. Thus, every block contains the hash value of the parent block constructing the blockchain (Zheng et al., 2017).

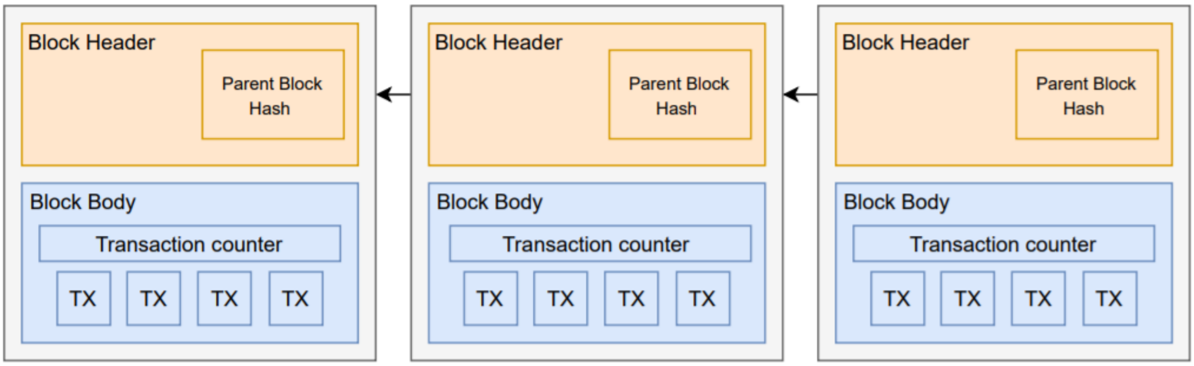

Figure 4: The sequence of blocks constructing the blockchain (Zheng et al., 2017)

In general terms, a block could be divided into two parts, the block body and its header. For the block body, the two basic components are the transactions (TX) and the counter of these transactions. The size of the block is pre-defined by design, and consequently, the number of transactions depends on the block size. For the block header, the basic component is the hash value of the parent block. The hash value is a fixed-size unique identifier for every block in the blockchain (Zheng et al., 2017). Besides the parent block hash, the block header also consists of several logical components that include:

- *Block version attribute:* that defines the block validation rules.
- *Merkle tree root hash:* describing the hash values of all transactions (TX) in the block.
- *Timestamp:* record of the current time defined in seconds as universal time since January 1, 1970.
- *nBits:* representing the target threshold of a valid bock hash.
- *Nonce:* a consensus attribute used in the validation process for adding new blocks to the chain.

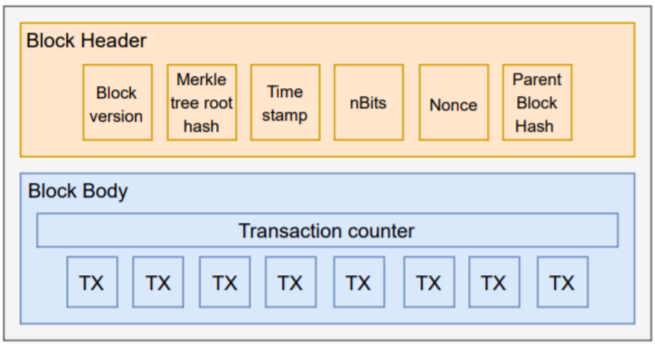

Figure 5: Block structure in the blockchain (Zheng et al., 2017)

3.2.2. Cryptographic Mechanisms

For validating the authentication of transactions between two parties, blockchain technically uses an asymmetric cryptographic mechanism based on the elliptic curve digital signature algorithm (ECDSA) (Zheng et al., 2018). Thus, considering the *"known actors"* in cryptography Alice and Bob, the representation of the algorithm is visualized in Figure 7.

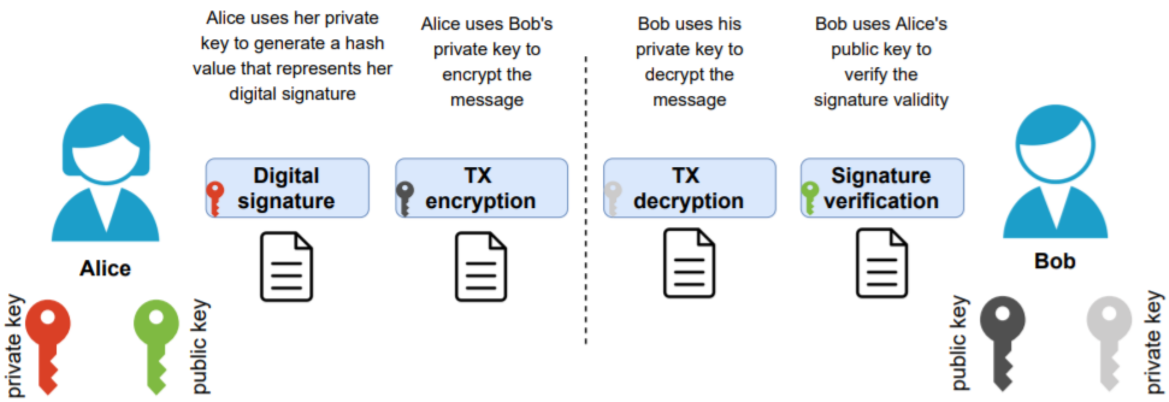

Figure 6: Cryptographic mechanisms and digital signature in blockchain

Both Alice and Bob hold a pair of private and public keys. These keys ensure digital authentication in sending and receiving transactions. In performing a transaction (TX) with Bob, Alice uses her private key to generate a hash value representing her digital signature appended to the transaction. The digital signature authenticates that the transaction was sent from Alice. Further, Alice uses Bob's public key to encrypt the content of the transaction. This process ensures that only Bob can decrypt the transaction using his private key. In processing the transaction, Bob uses his private key to decrypt the transaction. Moreover, Bob uses Alice's public key to decrypt the signature via the hash from which the signature was created. If Bob isn't able to decrypt the digital signature, this means that the signature isn't Alice's, or Alice changed her signature. In both cases, the transaction is considered invalid (Zheng et al., 2017).

Using cryptographic mechanisms, actors can securely perform various transactions. Although, these transactions should also be transparently and securely recorded in the distributed ledger. For this purpose, the network should verify its content and reach a consensus in the validation process. Moreover, considering that there is no central entity in the blockchain, the network should have by design operation protocols. As previously introduced, two common consensus protocols used in blockchain are *proof of work* and *proof of stake*, described as follows.

3.2.3. Consensus Protocols

The *Proof of Work (PoW)* concept originated as a computational method to deter service attacks and discourage spam abuses. Generally, "work" means the requirement of computational work from service requesters. Thus, this work requirement reduces the capabilities of attackers. An innovative approach was introduced with the conceptualization of Bitcoin that uses PoW to validate transactions and broadcast new blocks to the blockchain. Actors providing computational capacity to the network are called miners. These miners have to calculate the hash values of block headers using the *header nonce* to obtain a targeted condition. On the one hand, calculating the hash value and obtain a targeted condition requires computational effort, but on the other hand, the resulted calculations are easily verifiable. Miners compete in the network to calculate the targeted condition. The first miner reaching the targeted condition would broadcast the new block to the network and receive a reward in terms of a transaction fee. Other miners in the blockchain verify the validity of calculations. As a result, if network consensus is reached, the new block is added to the distributed ledger (Zheng et al., 2017). However, considering the computational effort required in PoW, the protocol may waste too much resources, mainly energy. Therefore, alternatives have been designed to mitigate the loss of resources. These alternatives imply special mathematical applications (e.g., through prime number chain to reduce the computational effort) or other consensus protocols. In this regard, an energy-saving alternative for PoW is *proof of stake*.

In *Proof of Stake (PoS)*, actors calculating the hash value to obtain the consensus in the network are called validators. Validators are selected randomly to calculate the targeted conditions and broadcast the new blocks to the network. Thus, delegating the computational power through random selection reduces the computational requirement from the entire network. However, random selection is vulnerable to attacks. In this regard, PoS protocol requires a "stake" (security deposit) from actors to be part of the validators network. Moreover, in the random selection of validators, PoS assign validators to computational work according to the lowest hash value in combinations with the size of the stake. In this context, validators with higher stakes have more chance of being selected to validate new blocks and fewer incentives to attack the network (Zheng et al., 2017). Thus, from one perspective, more blocks validated represent more perceived rewards, and from another perspective, if validators approve fraudulent transactions, their stakes would be seized. This could ensure the trust of validators in the network considering the financial motivator and the punishment for losing the stake. Although, the critics of this protocol refer to the unfair distribution of blocks validation, which means that validators with a higher stake are bound to be dominant in the network. This results in the trade-off between resources efficiency and equivalent distribution of validation work.

PoW and PoS are two common consensus protocols used in achieving decentralization in the blockchain. With the advancement of blockchain technology, more alternative protocols as PBFT (Practical byzantine fault tolerance) or Ripple are emerging. However, alternative protocols could be analyzed in separate research with a technology-based approach. Another essential aspect that is also analyzed in data marketplaces is smart contracts.

3.2.4. Smart Contracts

The technology functionality in the blockchain application with Bitcoin is limited to the state machine model, which ensures the transaction of coins from one address to another (Dinh et al., 2018). However, in recent years, blockchain technology has grown beyond the mechanisms of crypto coin transactions. In this regard, *smart contracts* are an innovative approach that supports the model of user-defined states (Dinh et al., 2018). This implies that users can define transaction conditions in programmable and self-executable contracts. Furthermore, these smart contracts are deployed in the blockchain network, making the user-defined conditions decentralized, immutable, transparent, secure, and network-verifiable. Thus, smart contracts represent a flexible solution to utilize blockchain by requiring certain conditions to be fulfilled in transactions with two or more participants. With the flexibility in defining transaction conditions, blockchain and the application of smart contracts gained research popularity in industries such as finance, the Internet of Things (IoT), or supply chain management. Thus, the characteristics of smart contracts can also be regarded in this context of data marketplaces. More specifically, considering that smart contracts support the model of the user-defined state, providers could define the constraints of data usage in smart contracts. Subsequently, with the deployed smart contracts on the blockchain ledger, consumers could purchase data without the implication of central structures. Lastly, as earlier stated, these smart contracts are self-executable; thereby, once the constraints in the defined contract are met, the financial resources are automatically transferred in a decentralized, transparent, secure, and network-verifiable process.

3.3. Data Marketplaces with Blockchain Architecture

With the overview of the blockchain characteristics from section 3.2, in this section, blockchain technology is analyzed in the context of data marketplaces according to the generic infrastructure components identified from the first sub-research question (see section 2.6). In the following, these infrastructure components are briefly summarized in table 2; thereafter, each infrastructure component is investigated from a technical point of view, describing the blockchain-specific functionalities.

Infrastructure components	Description
Identity Management	Actors participating in data exchange were characterized as data providers, data consumers, third-party service providers, and marketplace owners (Spiekermann, 2019). In the multilateral marketplaces, these actors should have appropriate access rights to their resources through *identification, authentication, and authorization* processes.
Data Storage, Custody, Manipulation	In multilateral data platforms were more than a single provider or consumer can exchange data, marketplaces should *ensure integrity* in the data exchange process. According to Stahl et al. (2015), this could be regulated through established *licensing models*.
Query & Search	Metadata represents an essential reference point both for data providers and consumers. Through metadata, providers can describe the supplied datasets, and consumers can trace these specifications. In this regard, metadata may enable the *query and search* functionalities for data marketplaces (Sharma et al., 2020).
Data Quality Assessment	As intermediate and intangible goods, scholars described the challenges arising in verifying data quality (Koutroumpis et al., 2017). Although in the commercialization process, data quality assessment is an essential benchmark for consumers.
Data Process and Analytics	Data are rarely valuable alone and are generally input into analytics to produce insights that can be expressed as content-based information goods (Koutroumpis et al., 2017). Thus, *data process and analytics* is an essential infrastructure component to enhance the usability of data assets
Pricing Mechanism	Considering the nonexistent property rights of intangible products like data, the *pricing models* are essential governance mechanisms in commercializing data assets (Teece, 2010).
Data Exchange	Technically, data marketplaces offer the infrastructure for *data exchange* among different actors as their general functionality (Fricker & Maksimov, 2017). Therefore, data exchange could be considered as the primary infrastructure component.

Table 2: Generic infrastructure components for data marketplaces

3.3.1. Identity Management

In traditional identification, authentication, and authorization processes, service providers need to develop and maintain their databases of user information. More specifically, service providers ensure that users have access securely to their data or resources. Generally, these solutions are costly and challenging to secure. Blockchain offers an alternative identity management solution that is closer to *self-sovereign identity*. The *self-sovereign identity* implies that each user controls and manages their digital identity (Lim et al., 2018). In technical terms, this solution includes two primary components, first the *digital identity* itself and second the associated *set of verifiable attributes* cryptographically signed by trusted issuers. In the context of blockchain-based marketplaces, these two components represented the distributed identifier and trusted third-parties signed attributes (e.g., signed by marketplace owner, third-party data quality validator, etc.). Both the distributed identifiers and signed attributes use the characteristics of asymmetric cryptographic mechanisms as previously described in the authentication of transactions between two parties (see section 3.2.2). The asymmetric cryptographic mechanisms with the appropriate use of private and public keys ensure that only the targeted party in the network receives the transaction information, and by signing the transaction, it has an evident origin. In the context of third-party signed attributes, the evident transaction origin could also identify additional trusted parties that validate certain characteristics of the transaction content. Thus, *self-sovereign identity* in the context of blockchain enables rightful users to access digital services or personal data by leveraging the proven security characteristics of cryptographic systems. Although, in *self-sovereign identity* management, users themselves have more responsibility in securing their private keys from malicious actors. Meaning that if malicious actors gain access to the user's private keys, it will be difficult to revoke the access rights in a distributed system.

3.3.2. Data Storage, Custody, Manipulation

In blockchain technology, transactions are communicated and validated in blocks recorded on a distributed ledger that is available to every network participant (see section 3.2.1). Thus, blockchain architecture eliminates by design the need for a central entity to store data or to manipulate the truncation flow. In this context, data providers decide on the storage, custody, and manipulation methods. This approach is commonly considered in articled analyzing blockchain-based data marketplaces as perceiving ownership, sovereignty, and control over data assets without relying on intermediaries during data exchange (Koutsos et al., 2020 and Hynes et al., 2018).

With the exclusion of intermediaries in data exchange, data providers could be eligible to decide on the *data storage* methods. For this purpose, in the decentralized architecture, data providers should have the possibility to store data locally on their devices, use web cloud storage (e.g., AWS, Microsoft OneDrive, Google Cloud, etc.), decentralized storage (e.g., filecoin), or public data oracles (e.g., Chainlink). Once data could be localized, providers may share their addresses through smart contracts that include the usage constraints (Hynes et al., 2018). Therefore, from a technical perspective, in the decentralized architecture, data providers could decide on the storage methods and share data through localization addresses implemented in smart contracts.

In maintaining sovereignty and control over personal digital assets, *data custody and manipulation* represent the ability of providers to govern the access rights for their data assets. The access rights in the blockchain architecture are secured through tokenization. In data security, tokenization represents replacing sensitive data with non-sensitive equivalents that reference the original data. More specifically, a token can be used as a utility providing digital access to an application or service based on blockchain (Kim & Chung, 2019). Technically holding a utility token to a specific data asset represents automatically having the license to use that data. Thus, providers could use utility tokens in data exchange, licensing the trustful actors to consume their data assets.

3.3.3. Query & Search

In multilateral data marketplaces, numerous providers could offer their data assets. Thus, with a large pool of data from various sources, consumers should be able to *query and search* through multilateral marketplaces to find suitable datasets. For this purpose, metadata represents an annotation identifier for data offered in the market. Technically, the annotation of data in the market may be accomplished through programming interfaces that index the metadata. These programming interfaces can implement different granularity levels of indexing. At a low level, it is essential to index the basic entities of the distributed ledger (e.g., blocks, transactions, and accounts) (Third & Domingue, 2017). This could provide direct and fast ability to locate and retrieve metadata in querying the distributed ledger. At a higher level, the smart contracts can be indexed according to the characteristics of the metadata. As a result, the indexed metadata can facilitate the segmentation of data pools, curating available datasets to create recommendations for consumers. Moreover, with the implementation of the indexing mechanism, the metadata in the blockchain can be efficiently queried with query and manipulation languages as GraphQL.

3.3.4. Data Quality Assessment

Data quality assessment as intangible assets is a complex task that may consider various aspects. In section 2.2.1, the objective and subjective criteria for quality assessment were explained. Therefore, on the one hand, for objective criteria, data provenance, collection methods, and data sources are essential properties of metadata. On the other hand, the quality of data could also be verified through subjective criteria according to consumer requirements. Technically in the blockchain architecture, the metadata, as well as subjective and objective criteria, could be implemented in the distributed ledger, ensuring transparency in quality assessment.

Besides the objective and subjective criteria, for digital marketplaces, participant-level quality verification by intermediaries or marketplace owners may also represent an efficient mechanism to secure the safety and trustworthiness of actors when there are high levels of moral hazards (Koutroumpis et al., 2020). Therefore, an additional method in ensuring data quality control could be accomplished through verification services of providers and consumers. In the context of blockchain technology, the complete history of all transactions and originated addresses could be subtracted from the distributed ledger. Moreover, untrustful providers or consumers who attempt to hide their identity using different blockchain addresses can be algorithmically grouped in clusters (Spagnuolo et al., 2014). This could help deduct intelligence from the distributed ledger and reveal the intentions of vicious actors. Moreover, the immutability and transparency of the blockchain ledger introduce opportunities for effective audit services. Therefore, monitoring data providers can take the form of tracing previous transactions and establishing reputations systems (Koutroumpis et al., 2020). Subsequently, monitoring data consumers can represent the mechanisms based on actor complaints with data usage agreements. A data provider or consumer found guilty of violating the contractual arrangements could be penalized by decreasing their trust value, making a provider or consumer less likely to be engaged in following transactions (Noorian et al., 2014). These governance mechanisms for quality assessment decentralize the authority of a single entity, leaving the judgment of trustfulness on the market participants.

3.3.5. Data Process and Analytics

As previously described in section 2.2, data are rarely valuable alone and are generally input into analytics to produce insights expressed as content-based information goods (Koutroumpis et al., 2017). To enhance data usability, marketplaces should support the integration of various *data processing and analytics* service*s*. In case data marketplaces do not have the necessary capacity for offering these services, they could appeal to the expertise of third-party providers (Spiekermann, 2019). For instance, third parties with expertise in Artificial Intelligence (AI) could provide consumers with the necessary tools to train Machine Learning (ML) algorithms for different use cases. Moreover, besides algorithms and applications, marketplaces could also establish collaborations with service providers as storage services, data aggregation services, data transformation services, or data quality validation services. As a result, with the integration of various third-party providers, marketplaces would offer the flexibility for consumers and providers to use common services and create an ecosystem that enhances trust in data exchange. In technical terms, the integration of third parties can be realized through smart contracts offered in the blockchain. For instance, in the case of aggregation or transformation algorithms, third parties may deploy smart contracts in the blockchain network, defining their services. Moreover, for data quality validation services, the signature of trusted third parties could be appended in the data exchange process, making the transactions approved by trusted or recognized entities.

3.3.6. Pricing Mechanism

For *pricing mechanisms*, besides flat prices, marketplaces could also offer enriched models as package pricing, flat fee tariff, two-part tariff, and freemium (see sections 2.2.2). Depending on the uses case, these pricing models could provide the flexibility in reaching most consumers. Although, in some cases, clear guidelines are required in estimating logical prices to favor both data providers and consumers. Technically, these pricing models can be programmable in smart contracts deployed in the blockchain. Moreover, the payment mechanisms of the smart contracts are self-executable, ensuring the automatic transfer of resources (Hynes et al., 2018).

3.3.7. Data Exchange

Data exchange could be characterized as the primary infrastructure component of data marketplaces. Technically, in blockchain-based marketplaces, data providers could specify the constraints in using data through smart contracts and define the access rights by utility tokenizing. Having a token as a utility to a specific dataset may represent the right to access the dataset or, more formally, holding a license to the dataset. Subsequently, consumers could acquire these tokens through smart contracts deployed on the blockchain ledger and use them to access the referring datasets.

As previously explained, transactions in blockchain are recorded in a distributed ledger publicly revealed to the network participants (see section 3.2). However, in the context of data exchange, this could impact the privacy perceiving aspect. Meaning that once data is purchased, the transaction will be registered in the distributed blockchain ledger, making data available to the whole network. Scholars describe that data could be exchanged off-chain through a Trusted Execution Environment (TEE) to ensure privacy in data transactions (Koutsos et al., 2020 and Hynes et al., 2018). A TEE may represent a federated network of known nodes operating under a restricted threat model. Technically, the network of known nodes must have installed security-related assets, codes, and underlying trusted operating systems to form a secure area for processing transactions without revealing any sensitive information to the whole network.

Therefore, considering the TEE, two environments are required to ensure security and privacy in data exchange. The first environment characterizes the on-chain distributed blockchain ledger, and the second environment represents the off-chain TEE. On the one hand, the terms of agreements specified in smart contracts stored on-chain establish the logic for buying and selling data. On the other hand, the actual transfer of data is accomplished off-chain through trusted executions environments. Furthermore, the two environments should communicate to guarantee a complete selling, exchanging, and payment process. More specifically, the off-chain environment should share a prove to the smart contacts stored on-chain that data was successfully transferred and payment method can be executed. This could be accomplished through cryptographic Zero-Knowledge proof (ZK proof) (Koutsos et al., 2020). The ZK proof is a cryptographic method by which the off-chain environment can prove to the on-chain environment that the transfer of data was completed without revealing any information apart. Additionally, considering the security characteristics of the TEE, the off-chain environment could also integrate mechanisms of Multi-Party Computation (MPC) or algorithms for data analysis. Finally, by summarizing the data exchange and payment process, the following workflow can be described:

1. A data provider can store data locally or upload data on storage services (e.g., AWS, Google Cloud, etc.). Once data is localized, the data provider publishes the smart contract containing the constraints in using data and the tokens to access data. The combination of data tokens and smart contract defines the access rights of the consumer.
2. A data consumer acquires data tokens and provides them to the on-chain smart contract. The self-executable smart contract signals the transfer of data through off-chain TEE.
3. Once data is transferred, a zero-knowledge proof is reported to the smart contract that is transparently stored and network-verifiable on-chain. With valid zero-knowledge proof, the funds are subtracted from the consumer and paid to the provider. Alternatively to the direct data exchange and depending on the particularities of the smart contract, multi-party computation or algorithms for data analysis can be executed in the TEE to receive insights for original data.

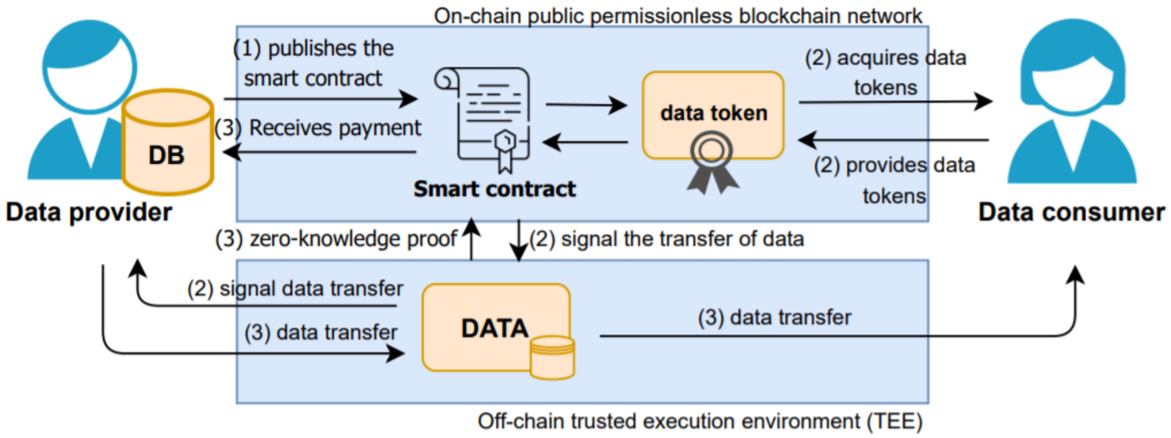

Figure 7: Data exchange with blockchain architecture

3.4. The Design of a Decentralized High-Level Architecture

In the following section, the decentralized HLA is designed, forming the artifact, which answers the second sub-research question, *How does a decentralized architecture for data marketplaces look like?* The HLA integrates the infrastructure components earlier analyzed in section 3.3 into an Enterprise Application Architecture (frequently referred to as Multitier Architecture). This architectural representation is a well-known approach used both by scholars as well as practitioners and with numerous theoretical and practical publications that have detailed the architectural patterns (Gamma et al., 1995; Buschmann et al., 2007; Fowler, 2012). The HLA is constructed through a domain engineering process that collects actions for obtaining and expressing data about systems with similar properties. Therefore, in the proposed blockchain architecture for data marketplaces, the infrastructure components were grouped according to their parametric functionalities applicable to the domain of eight layers. Even if the application of blockchain technology in data marketplaces is a relatively new topic, similar studies researching the architecture of blockchain-based data marketplaces can be identified (Garcia-Font, 2021; Zhang et al., 2018; Özyilmaz et al., 2018). However, most of these studies have a domain-specific analysis for the data marketplaces (e.g., IoT, smart cities). In this regard, the studies of previous scholars were used as primary references for a more generic approach, with infrastructure components that could be found in different industry applications. As a result, the HLA is presented in Figure 10, followed by the description of each layer.

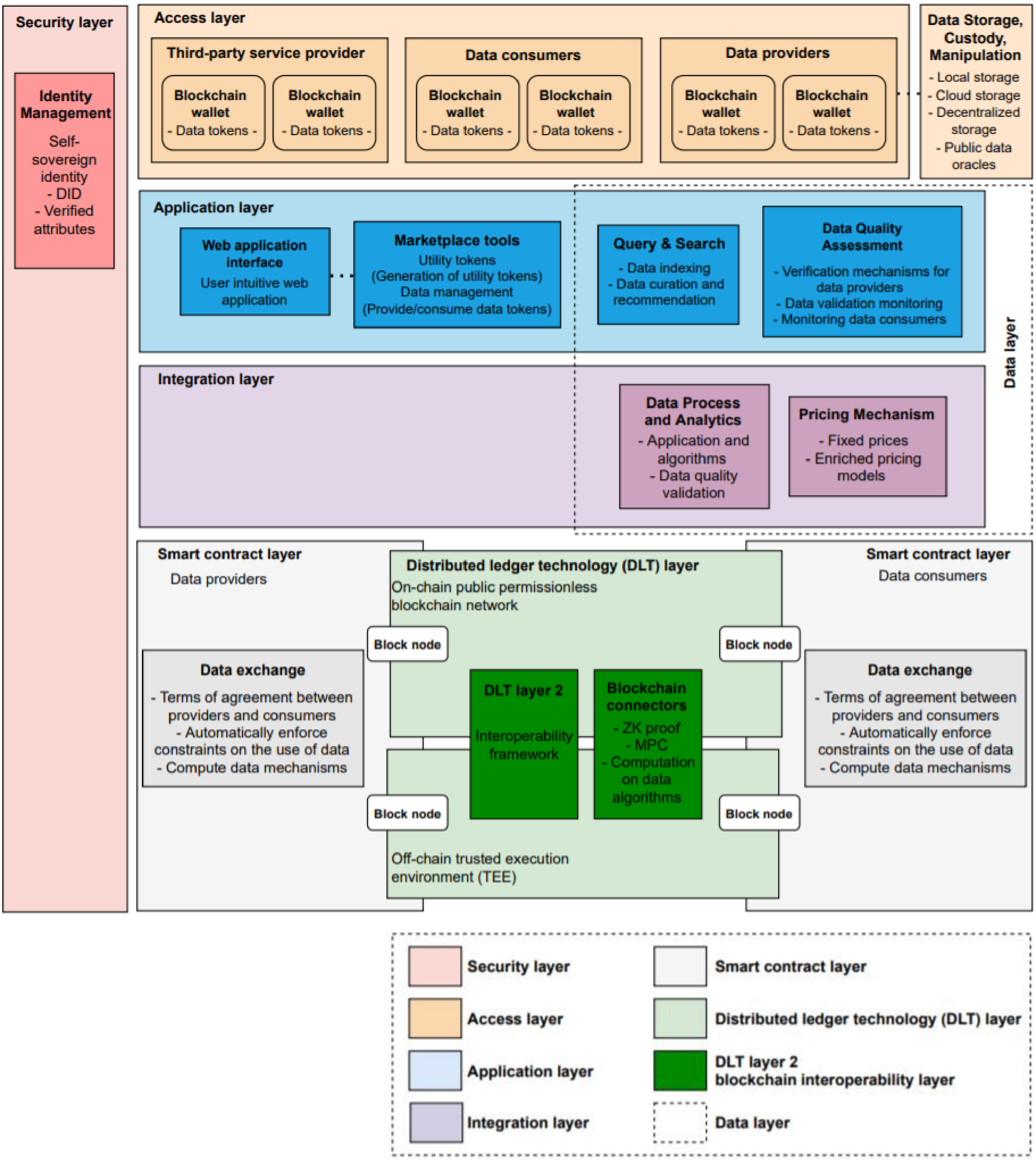

Figure 8: The High-Level Architecture for data marketplaces with blockchain architecture

Access Layer

The access layer describes the three main actors engaged in data exchange. Every actor keeps custody of data tokens used as utilities that provide digital access to their data assets or algorithms. In storing data tokens, actors use blockchain wallets. These blockchain wallets securely store and facilitate the management of data tokens in the exchange process. Moreover, actors can hold multiple blockchain wallets to manage their data adequately. The ownership, sovereignty, and control of data in the proposed architecture are organized according to the tokenization process previously analyzed in data storage, custody, and manipulation (see section 3.3.2). In the same context, the data storage component separated in the access layer represents the possibility of actors to decide on the storage method (e.g., local storage, cloud storage, decentralized storage, or public oracles).

Application Layer

The application layer gathers the infrastructure components with which actors directly interact in data marketplaces. From a user experience perspective, the infrastructure components should have similar functionalities as conventional digital marketplaces. However, technically the development and deployment of the infrastructure components are realized according to blockchain technology by using the information from the distributed ledger. Thus, these components differ from traditional backend systems, since besides managing data transactions, these systems also perform cryptographic operations to ensure the security of users interacting with a transparent ledger. In terms of transparency, blockchain provides a complete history of data transactions, which, combined with trust measures or reputation algorithms, can reveal the safety and trustworthiness of actors. Moreover, as intermediaries, marketplace owners have less technical abilities to block the flow of data in the market, which shifts their responsibility from being in control of the data market to providing the tools and instruments in using the market.

Integration Layer

It takes extensive effort for platforms to operate in isolation considering the characteristics of the multidisciplinary data industry. Thus, the integration of different technologies is an imperative practice in the context of distributed technologies. In this regard, the integration layer from the proposed architects offers flexibility in the exchange process, aiming to form the ecosystem for data services that enhance the usability of marketplaces. Technically, the architecture expands its capabilities through infrastructure components acting as middleware with external data services. These external data services are essential for marketplaces to incorporate the expertise of different service providers that could add value to the market through AI plugins, storage services, algorithms, applications, or pricing models to help providers and consumers gain intelligence from data.

Data Layer

Efficiently managing data assets requires different processing mechanisms, from simple searching for datasets to complex analytics and aggregation mechanisms. Moreover, actors should have an extensive access range to services valorizing their assets. In this regard, both application and integration layers offer the infrastructure components for data management. The data layer is composed of infrastructure components provided by marketplaces and incorporated from external services.

Smart Contract Layer

In the architecture representation, the smart contract layer interacts with the DLT layer. The interaction of the two layers represents the data flow in transactions between providers and consumers. Technically, smart contracts in blockchain technology are deployed in the distributed ledger, accountable for two main aspects. First, define the terms of agreements in using data through smart contracts. Second, automatically enforce the constraints in the data exchange process. As previously described in *data exchange* (see section 3.3.7), data providers can specify access rights to their assets through tokenization by applying similar principles as conventional licensing models. For instance, a data provider could use different degrees in accessing data (e.g., limited, partial, or complete) and offer data tokens according to the defined degrees in a smart contract. Subsequently, a data consumer can acquire the necessary access degree with data tokens and use them through the smart contract to receive data. Thus, tokens in combination with smart contracts enable the characteristics of licensing models in data consumption.

Distributed Ledger Technology (DLT) Layer

The DLT layer is divided into the on-chain blockchain network and the off-chain TEE. The division of the two environments is necessary to ensure transactions transparency and perceive privacy in data exchange (see section 3.3.7). According to the proposed architecture, both environments are based on blockchain technology. Thus, the two environments operate with information blocks that register transactions between providers and consumers. On the one hand, the on-chain environment stores the information blocks about the performed data exchange and payments. On the other hand, the actual exchange of data is accomplished through information blocks in the off-chain environment. Thus, with the two environments, the DLT ecosystem avoids having shared data in the distributed ledger and ensures that transaction processes are transparently registered in the blockchain network. This could enable a secure exchange of data and traceability of transactions to resolve possible conflicts.

DLT Layer 2

The DLT layer 2 represents the interoperability framework between the on-chain and off-chain environments. This interoperability is necessary to ensure the communication in two technical processes. The first process is the automatic trigger of the data exchange in the off-chain TEE. The second is the return of the network-verifiable transparent proof in the on-chain environment once data is exchanged. These two processes ensure that, on the one hand, consumers comply with their obligations to pay for data. On the other hand, providers fulfill their commitment to transferring paid data. Thus, considering the interoperability framework, the exchange process can be accomplished through the ZK proof protocol integrated with smart contracts. The ZK proof is a cryptographic protocol that can be explained as follows. *"Given a statement x, a prover wishes to prove to a verifier that there exists witness w such that $(x, w) \in R_L$ where R_L is the corresponding relation of the NP language L. A ZK proof achieves two security guarantees: (i) soundness, i.e., no cheating prover can convince the verifier for $x \notin L$, and (ii) zero-knowledge, i.e., the verifier learns nothing about the witness w (besides its existence)"* (Koutsos et al., 2020). The ZK proof protocol can be useful both for the direct exchange of data and for computation algorithms to subtract intelligence from data as MPC. The difference regarding the data exchange would be represented in smart contracts with defined constraints and computation processes. However, the ZK proof protocol in the case of MPC could apply similar principles in proving to the network that consumers received the necessary intelligence from data.

Security Layer

The two essential elements in the security layer are the unique identification of actors and the verifiable identity attributes contained in the identity management component.

In the context of the unique identification of actors, blockchain technology differs from conventional backend systems. Actors themselves manage their identity through asymmetric cryptographic mechanisms that use the combination of public and private keys. Thus, considering that there are no intermediaries to manage a database of user credentials, every infrastructure component has to be bound with cryptographic mechanisms to ensure self-sovereign identity. In this regard, the security layer has a vertical position in the architecture.

Further, regarding verifiable identity attributes, an example in this sense could be described by the digital identity of citizens in many European countries. For instance, citizens in the Netherlands have a governmental-managed digital identity used in accessing different public services. In this context, the governmental-managed digital identity is a trusted authenticator that proves the identity of citizens without revealing their actual credentials. Similar principles could be applicable for data marketplaces. Actors should be able to use the functionalities of trusted third parties to endorse their identity or transactions. For blockchain-based marketplaces, this can be achieved by digitally signing network-verifiable attributes. These attributes are used in communication between infrastructure components as authenticity poofs of trusted parties. The governmental-managed identity characterized an example of trusted third parties. However, it should be at the decision of owners operating marketplaces to offer the abilities for actors to use different services in the authentication process.

3.5. Conclusion of Chapter 3

In this chapter, the blockchain characteristics in the context of data marketplaces were analyzed. For this purpose, first, blockchain technology was studied. Thereafter, the infrastructure components of data marketplaces with a blockchain architecture were researched. Lastly, the infrastructure components were integrated into a decentralized HLA, answering the second sub-research question, *How does a decentralized architecture for data marketplaces look like?*

In designing the decentralized HLA, eight architecture layers were defined: *Access Layer, Application Layer, Integration Layer, Data Layer, Smart Contract Layer, DLT Layer, DLT Layer 2, and Security Layer*. The functionalities of these layers were described as follows. In the *Access Layer,* marketplace actors keep custody of data tokens used as utilities that provide digital access to their data assets or algorithms. In storing these utility tokens, actors may use blockchain wallets. Further, in the *Application Layer,* the infrastructure components with which actors directly interact were described. Considering the technical specification of blockchain technology, these components differ from traditional backend systems. More specifically, the infrastructure components in the *Application Layer* should also perform cryptographic operations to ensure the security of users interacting with the distributed ledger. Next, the integration of different technologies was described as an imperative practice in the context of distributed technologies. Thus, in the *Integration Layer*, the capabilities of the architecture were expanded with the infrastructure components acting as middleware with external data services. Moreover, actors should have an extensive access range to services valorizing their assets to enhance data usability. For this purpose, the *Data Layer* was composed of infrastructure components provided by data marketplaces and external services. Further, in the *Smart Contract Layer*, the user-defined states in data transactions between providers and consumers were presented. In particular, the automatic enforcement of data constraints was described. With the established data constraints, the *DLT Layer* defined the on-chain and the off-chain environments for data exchange. On the one hand, the on-chain environment was defined to store the information blocks about the performed transactions and payments. On the other hand, the off-chain environment was specified to handle the data transfer from providers to consumers. In the same context, for the interoperability between the two environments, the *DLT Layer 2* was defined. In this layer, the ZK-proof protocol was explained to accomplish trustful communication between the on-chain and off-chain environments without revealing the content of the data exchanged. Lastly, the *Security Layer* characterized the asymmetric cryptographic mechanisms for the unique identification of actors.

Further, the HLA is used in the discussion chapter to describe the technical specification of data marketplaces with blockchain architecture.

4

Conceptualization of a Taxonomy of Business Models for Decentralized Data Marketplaces

In the following chapter, the second design artifact is conceptualized. As it was described in the research approach, the second research artifact is the taxonomy of business models for decentralized data marketplaces. In constructing the taxonomy, the business models of data marketplaces are explored according to their compositional and subordinate elements. The compositional elements define what a business model is made of. Subsequently, the subordinate elements describe how decentralized data marketplaces create, deliver, and capture value. As a result, by subtracting these business model characteristics, the design artifact provides insights to answer the second sub-research question: *What are the characteristics elements in a taxonomy of business models for data marketplaces with a decentralized architecture?*

The chapter is structured in the following order. The literature search conducted on the business model taxonomies for data marketplaces and the selected data marketplaces for analyzes are described in section 4.1. Further, the theoretical background for the business models as well as the research framework is explained in section 4.2. Next, in section 4.3, the compositional elements in the business models of data marketplaces are researched. The subordinated elements are analyzed in section 4.4. Lastly, the chapter concludes with the conceptualization of the business model taxonomy in section 4.5.

4.1. Literature Search And Selection Criteria

Two aspects were examined in the literature search for this chapter. The first aspect was related to searching for scientific papers analyzing the business models of data marketplaces. Subsequently, the second aspect was associated with retrieving the technical whitepapers of data marketplaces with blockchain architecture. Therefore, starting with searching for scientific articles, databases as Scopus, Google Scholar, ResearchGate, and IEEE Xplore, were queried. Although, considering that the previous search iteration from section 2.1 revealed more articles than the research scope of chapter 2, no new reports that differ in the analyses of the business models could be found. Moreover, as was explained in section 2.1, several articles from the chapter 2 literature search were allocated to complete the literature search for this chapter. Therefore, the scientific articles used in this chapter, retrieved from previous search iteration, were as follows.

- The first article was *Fruhwirth et al. (2020)*, a research on the business models of data marketplaces that defined a general taxonomy for analyzing the business model dimensions and characteristics.
- The second paper is *Bergman (2020)*, a master thesis of a TU Delft student that researched the business model taxonomy for data marketplaces with industry-specific interpretation.
- The third article is *M. van de Ven et al. (2021)*, a study on the business models for data marketplaces from a multi-stakeholder perspective to view the data ecosystem as a whole.

With the identified scientific articles to be used in this chapter, the second aspect of the literature search was to retrieve the technical whitepapers of decentralized data marketplaces. A simple search iteration performed on Google revealed a pool of more than 40 marketplaces. Although, the majority of these projects were inactive. Therefore, to exclude the inactive projects, five *"should have"* criteria were introduced. In this context, data marketplaces should have: *a mission and vision, available documentation on the website and whitepapers, well-defined terms and conditions, development activity on Github,* and *should preferably be present on coinmarketcap*.

The reasoning behind the searching criteria was as follows. First, the *mission and vision* were considered to identify the specialization of marketplaces. Second, the *websites and whitepapers* were used to retrieve the necessary insights to be further applied in the research. Third, *well-defined terms and conditions* represented the criteria that defined the reliability of marketplaces in data exchange. Fourth, projects that are still active in the market are usually trying to improve or maintain their platforms. In this context, Github development represented an indicator for the activity of marketplaces. Fifth, as is later analyzed in this chapter, most blockchain marketplaces use cryptocurrency as payment currency. Similar to traditional assets, cryptocurrency requires liquidation through specialized exchange platforms (e.g., Coinmarketcap). Therefore, analyzing the presence of these marketplaces on *Coinmarketcap* represented a *bonus* criteria to reveal their financial status.

Several decentralized data marketplaces could be selected using different search engines and considering the selection criteria. However, the current thesis research has a specific time frame. In this regard, the research is limited to analyzing five marketplaces. The number five was chosen to have a good balance between quantity and granularity of analysis. More specifically, on the one hand, there are enough sources of comparisons, and on the other hand, in-depth investigation can be conducted. To distinguish clear business model particularities, the marketplaces with different data specializations were selected. Furthermore, to evaluate their decentralized architecture, these data marketplaces were compared with the HLA. It has been noted that few decentralized marketplaces could be identified within the EU. The ones that could be selected had closer access to their whitepapers or codebase (e.g., Dataeum), making the research difficult. Another observation was that more developed decentralized data marketplaces were located in Switzerland. As earlier mentioned, most blockchain marketplaces use cryptocurrency tokens as their payment currency. In this sense, the Swiss Financial Market Supervisory Authority has been publishing guidelines for treating cryptocurrency assets since 2018 under the Swiss supervisory law (finma, 2019). Although, The EU regulatory framework for crypto-asset was adopted in January 2021 (European Commission, 2021). Thus, it may still require time for the crypto-asset framework to be adjusted in the regulatory bodies of the EU Member States and for data marketplaces with a blockchain architecture to emerge.

Finally, Table 8 presents the selected marketplaces according to *applying the selection criteria*, *taking into account the time frame*, *filtering the marketplaces with different specializations*, and *searching through a more extensive geographic area*. Thus, the following selected cases could be considered a representative sample of established and active decentralized marketplaces, which have variations of theoretical interests on their business model dimensions suitable for analysis in this chapter.

	Marketplace name	Specialization	Headquarter	Source of analysis
1.	*Ocean Protocol*	Data owners & consumers use	Singapore	• Website/Blogs • Marketplace whitepapers • Terms and conditions • Coinmarketcap • Github
2.	*dHealth Foundation*	Healthcare data	Switzerland	
3.	*Streamr*	Industrial real-time data	Switzerland	
4.	*Vetri*	Personal data	Switzerland	
5.	*Enigma*	Data owners & consumers use	USA	

Table 3: Active data marketplaces with a decentralized architecture

4.2. Business Models

The concept of business models is widely researched in several disciplines, such as information systems (Al-Debei & Avison, 2010), strategic development (Wirtz et al., 2010), as well as technology and innovation (Chesbrough, 2010). However, considering the multiple subject matter lenses of different disciplines, scholars do not agree on a generally accepted definition for business models (Zott et al., 2011). Nevertheless, scholars and practitioners concise that business models are all about value (Teece, 2010; Santos et al., 2009; Zott et al., 2011). Therefore, for this thesis, business models are defined as the logic of how organizations *create* and *deliver* value to customers and then convert payments received to profits, describing the *captured* value (Teece, 2010).

As frequently expressed by various scholars, the conceptualization of business models engages a component-based perspective (Burkhart et al., 2011). Thus, the current research investigates two aspects of business models for data marketplaces: the *compositional* and *subordinate elements*. The *compositional elements* define what a business model is made of, and the *subordinate elements* represent the functional content of how independent organizations create, deliver, and capture value. As a result, the *compositional* and *subordinate elements* provide the characteristics for the taxonomy suitable to analyze independent organizations using business modeling as a viewing lens.

The thesis uses the business model rationale of value creation, delivery, and capture (Teece, 2010). This rationale represents the meta-characteristics of the business models taxonomy. In this chapter, the business model canvas is the framework used for studying the taxonomy meta-characteristics (Osterwalder & Pigneur, 2010). The original canvas defines nine interconnected components: *value proposition*, *customer segment*, *customer relationship*, *partners*, *key channels*, *key resources*, *key activities*, *costs*, and *revenues*. In the following, these components are aligned according to the business model rationale of value creation, delivery, and capture.

Value creation represents the processes that generate benefits for end consumers (Bouwman et al., 2008). Four compositional elements from the original canvas are applicable for value creation. These elements are *value proposition*, *customer segment*, *customer relationships,* and *partners.* The *value proposition* describes the products or services offered to customers aiming at solving their problems (Teece, 2010; Osterwalder & Pigneur, 2010; Chesbrough, 2010). According to Osterwalder & Pigneur (2010), companies, through their value proposition, target specific customer groups defined as the *customer segment*. Additionally, to develop the loyalty and trust of customers, the *customer relationship* is an essential component in the business models. Lastly, *partners* represent the entities that add or maintain value in the business process.

Value delivery describes the value chain of activities, resources, and abilities possed by companies to create and distribute the business proposition (Bouwman et al., 2008). Three main components may be identified in the business model canvas describing the value chain. These components are *key channels*, *key resources*, and *key activities* (Osterwalder & Pigneur, 2010). In value delivery, *key channels* represent the methods through which products or services are offered to end consumers. These methods can take different forms, ranging from selling products using digital platforms to direct personal services. Therefore, considering the different forms of the *key channels*, companies require technological, physical, or intellectual resources to offer products or services in the market, defined in the business model canvas as *key resources*. Lastly, to operate their business models, companies perform various *activities* characterizing their direct production efforts. These activities are defined in the business model canvas as *key activities*.

Value capture converts payments received to profits. The components from the business model canvas associated with value capture are *costs* and *revenues* (Osterwalder & Pigneur, 2010). In short, these two components describe the expenses to sustain the business models and the perceived value from commercial activities. However, the analysis of expenses to sustain the business model could be problematic in the current thesis. This is because the desk research is conducted only by retrieving available online documents, where financial activities are missing. Thus, the *costs* component is excluded from the taxonomy. Additionally, according to Teece, (2010), the implicit assumption that companies can capture value by simply selling output in established markets represents a deficient perception that disregards the essential business design issues. Furthermore, in the context of data marketplaces, intangible products as data introduce various challenges. Consumers want more than just products; they want solutions to their needs (Teece, 2010). However, the nonexisting property rights of intangible products cause difficulties in commercializing solutions and estimating their *pricing mechanism*. In this context, the pricing mechanism is also included in the value capture part of the taxonomy researched in the following section.

4.3. The Taxonomy Compositional Elements

In this section, the compositional elements for the taxonomy of business models for data marketplaces are researched. Different scholars studied the compositional elements for similar taxonomies. However, previous studies were focused on the generic aspect of data trading platforms, whereas the current thesis has a more specific application in the context of blockchain technology. Therefore, considering the scientific articles identified in the literature search (see section 4.1), *the compositional elements* suitable for studying data marketplaces are subtracted from the three selected papers and aligned with the characteristics of the blockchain architecture. In the following, the compositional elements are described according to the rationale of value *creation*, *delivery*, and *capture*.

4.3.1. Value Creation

In value creation, starting with Fruhwirth et al. (2020), the authors referred to three dimensions in analyzing value creation. These dimensions are *platform infrastructure*, *data origin*, and *review systems*. First, the *platform infrastructure* distinguishes the centralized and decentralized architectures. However, the centralized architecture is out of the scope in the current research. Moreover, the *platform infrastructure* itself has a closer association with the technology domain (M. van de Ven et al., 2021). Therefore, *platform infrastructure* has a more appropriate position in the following value delivery part of the taxonomy. The next dimension is *data origin*, which describes where data is coming from. Although, similar to *platform infrastructure*, data origin has a more appropriate position in value delivery. Lastly, the *review system* is an essential business dimension that characterizes the *quality of data*. In the taxonomy, the review system is used as part of the *data quality* dimension.

Next, pursuing the research of Bergman (2020), the author divided value creation between the business model canvas components of customer segment, value proposition, and customer relationship. For the customer segment component, the author distinguished the *data domain* and *participants* dimensions. Both dimensions could also be applied in researching data marketplaces with blockchain architecture. However, in Bergman (2020), the *domain dimension* has a specific application in the B2B automotive industry. Thus, to have a more generic *domain* for the current research, a similar dimension is subtracted from M. van de Ven et al. (2021), called the *industry domain*. Further, Bergman (2020) divided the value proposition into four dimensions: *data service*, *data output*, *data quality*, and *privacy*. These dimensions are also suitable for researching the value proposition of decentralized data marketplaces. However, the *data output* dimension describes the form of data delivery. Thus, this dimension is moved to the following *value delivery* part of the taxonomy. Besides the dimensions of Bergman (2020), *time relevancy* or *time frame* is another essential characteristic represented in both taxonomies of Fruhwirth et al. (2020) and M. van de Ven et al. (2021). The *time frame* dimension describes whether data needs updates to maintain relevancy. Therefore, the *time frame dimension* is also included in the value proposition of the taxonomy. The last compositional element according to the business model canvas for value creation is the customer relationship. In Bergman (2020), this element has a single *contract* dimension defined as negotiated, standard, or both. However, in blockchain-based data marketplaces, the self-executable *smart contracts* ensure the relationship between providers and consumers. Therefore, the *contract* dimension is replaced with the *smart contract* dimension that is also considered in M. van de Ven et al. (2021).

In addition to the dimension described by the three scientific articles, in the context of distributed technologies, the integration of different technologies is an imperative practice to enhance the usability of data exchange (see section 3.4). Well-established integration principles in data marketplaces may motivate the participation of third parties as storage services, data aggregation services, data transformation services, or data quality validation services. These third parties could provide mediating expertise in data exchange both for consumers and providers. For this purpose, a new dimension is introduced as *data mediating expertise* referring to the relationship with third-party providers or, more formally, the relationship with partners according to the business model canvas.

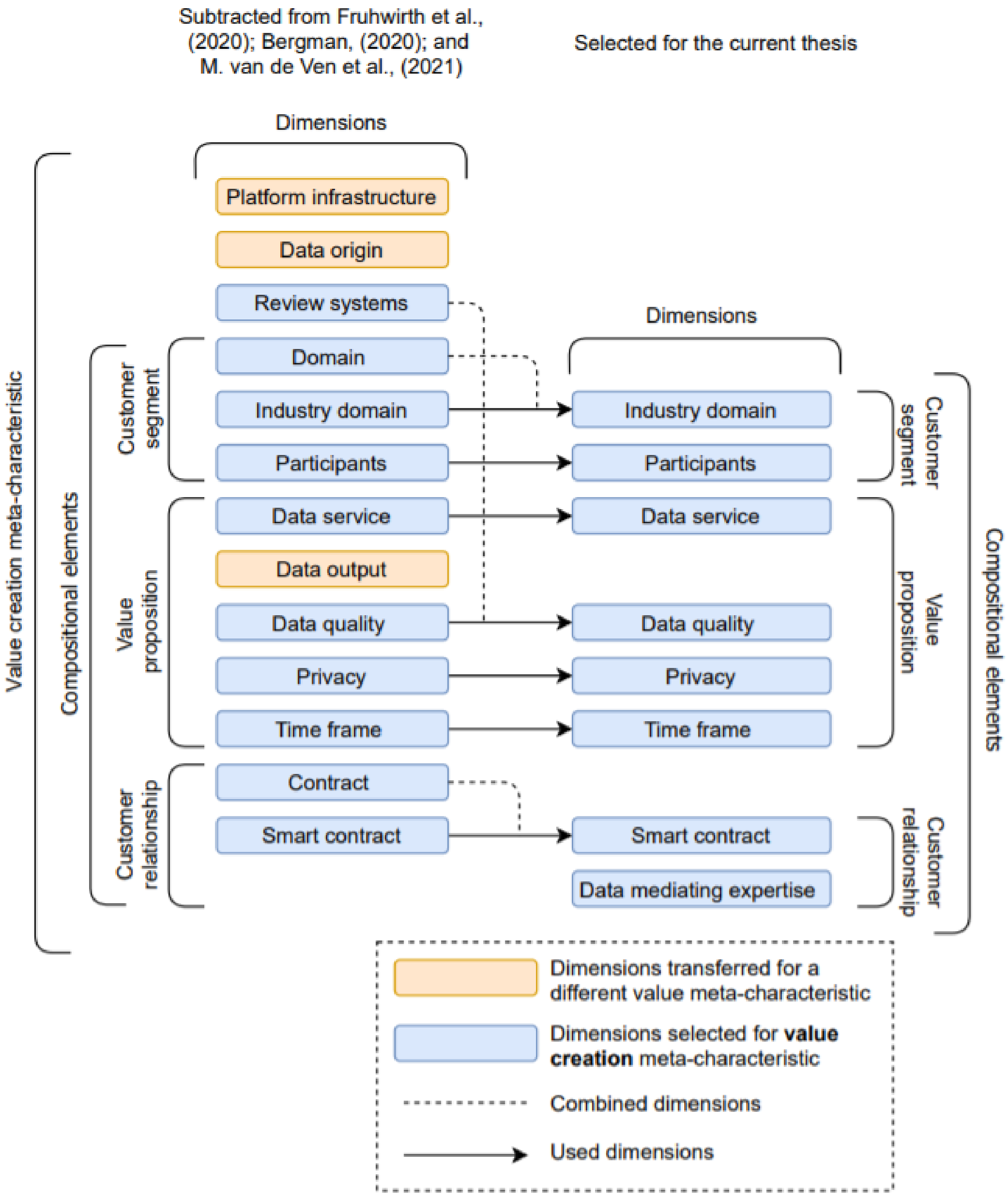

Figure 9: The alignment of compositional elements in the taxonomy for value creation

4.3.2. Value Delivery

In value delivery, considering the three articles selected in the literature search and starting with Fruhwirth et al. (2020), the authors distinguished the following business dimensions for value delivery: *data output*, *type of access*, *additional purchase support*, *domain*, *marketplaces participants*, and *smart contract with blockchain*. The first three dimensions: *data output*, *type of access*, and *additional purchase support,* could also be used in studying the decentralized data marketplaces. These dimensions are explained in the following section. However, the last three dimensions: *domain*, *marketplaces participants*, and *smart contract with blockchain*, were earlier analyzed as part of value creation in section 4.3.1. Therefore, these dimensions are disregarded in value delivery.

In compliance with the business model canvas, Bergman (2020), divided value delivery into key channels, key resources, and key activities. For key channels, according to Bergman (2020), the dimension describing the channels of accessing data is *platform access*. However, considering the characteristics of decentralization and transparency, data marketplaces with blockchain architecture primarily use only open platform access. Therefore, for key channels, more suitable is the business dimension of *data access* as described in Fruhwirth et al. (2020) and M. van de Ven et al. (2021), defining how consumers may access data. Additionally, *data origin,* which is similar to *data sources* describing where data is coming from, is also applicable in key channels (M. van de Ven et al., 2021). Moving to the business model canvas component of key resources, Bergman (2020) described a single business dimension: *platform infrastructure* as the resource of marketplace owners to deliver value to customers. Similar to Fruhwirth et al. (2020) and M. van de Ven et al. (2021), the platform infrastructure is represented as centralized or decentralized. However, considering the decentralized focus of the current thesis, the platform infrastructure in the taxonomy is used according to blockchain technology. Furthermore, besides *platform infrastructure*, a key resource that blockchain-based marketplaces govern is their *cryptocurrency*. In the three scientific articles Fruhwirth et al. (2020), Bergman (2020), and M. van de Ven et al. (2021), *cryptocurrency* is part of the payment method. However, in the context of blockchain technology, cryptocurrency tokens represent valuable resources that participants can use, stake, and exchange (Kim & Chung, 2019). This makes the *cryptocurrency* dimension also suitable for key resources part of the taxonomy. Lastly, in the key activities component of the business model canvas, Bergman (2020) described the dimension of the *data processing activity*. A similar dimension can be found in M. van de Ven et al. (2021) as the *data processing and analytical tools*. This dimension is used in the taxonomy as one of the marketplace activities to offer additional tools for data analytics. Moreover, considering that blockchain is an open source-based innovation technology, most blockchain-based projects adopt different forms of open-source governance (Savelyev, 2018). Therefore, an additional key activity is introduced in the taxonomy as *open-source governance* describing open-source project management and development activities.

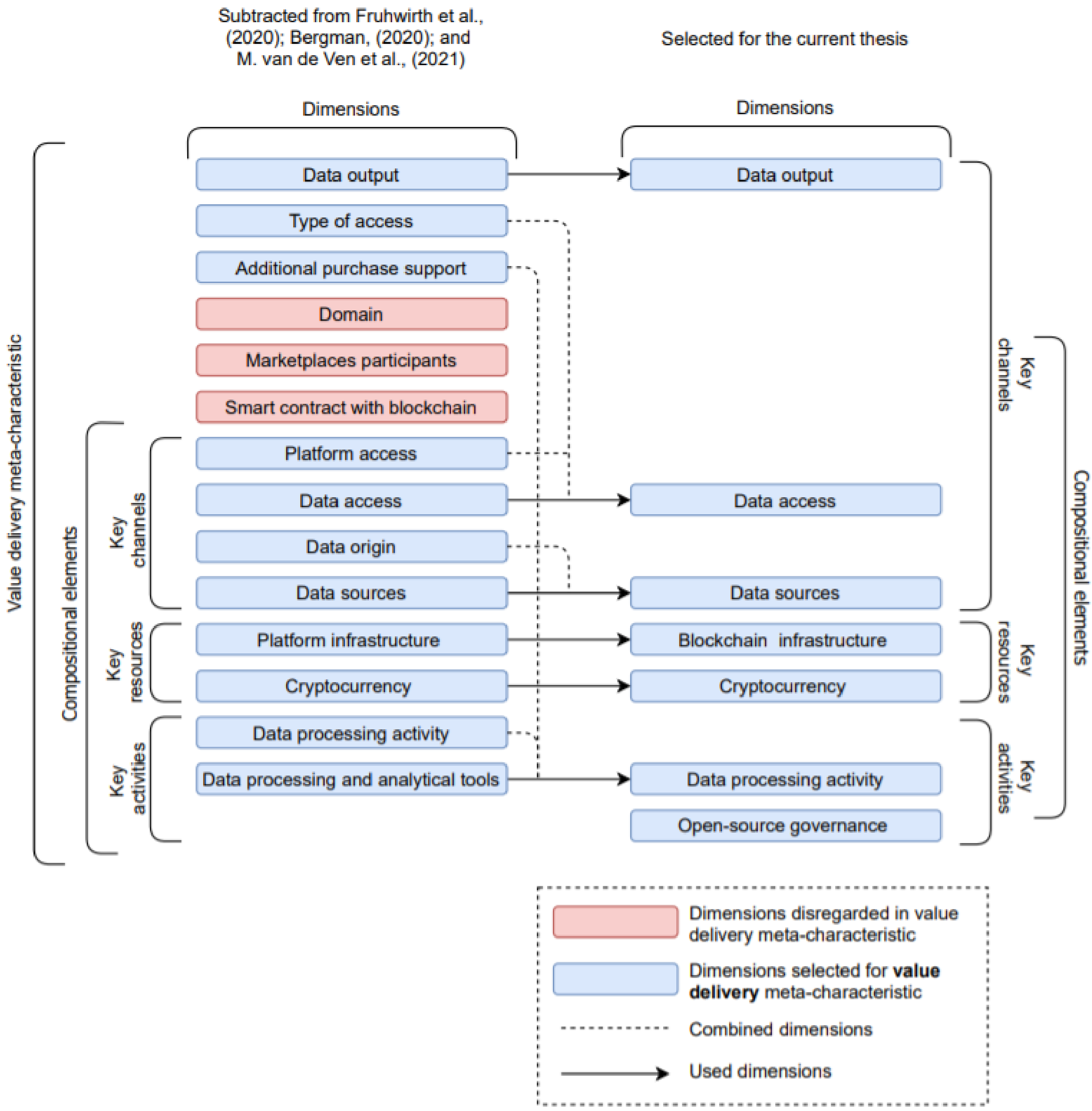

Figure 10: The alignment of compositional elements in the taxonomy for value delivery

4.3.3. Value Capture

The value capture meta-characteristic represents the value collected by data marketplaces themselves as business entities and the value received by network service providers. Similar to the earlier analyzed value creation in section 4.3.1 and value delivery in section 4.3.2, the compositional elements of value capture are investigated according to the selected articles in the literature search in section 4.1.

In compliance with the business model canvas (see section 4.2), Bergman (2020) distinguished in the taxonomy of business models for data marketplaces the revenue streams and the pricing models as the compositional elements for studying the value capture. For the *revenue streams*, Bergman (2020) described the fees charged by marketplace owners according to the free model, usage-based price, package price, flat fee tariff, and the freemium model. However, as is was analyzed in section 2.2.2, these models are used as *data pricing mechanisms,* not as *revenue streams* for data marketplaces. Thus, in this section, the approach as described in M. van de Ven et al. (2021) is used for the *revenue stream*. According to M. van de Ven et al. (2021), data marketplaces may perceive revenue through different models as commissions, subscriptions, usage base, and assets sales. Furthermore, besides the traditional *revenue model*, blockchain-based projects raise funding by selling cryptocurrency (Kim

& Chung, 2019). Therefore for value capture, this could describe a revenue stream different from the traditional models. For this purpose, a new revenue dimension is introduced in the taxonomy, called *cryptocurrency governance*. Returning to the pricing model, Bergman (2020) distinguished two business dimensions: *data pricing mechanisms* and *payment currency*. Although, regarding the *payment currency*, most marketplaces with blockchain architecture use only cryptocurrency as the payment method. Thus, *payment currency* is removed from the taxonomy. The last dimension introduced in the taxonomy, according to M. van de Ven et al. (2021), is *price discovery*. This dimension describes the possibility of data providers and consumers to determine a transaction price they agree on.

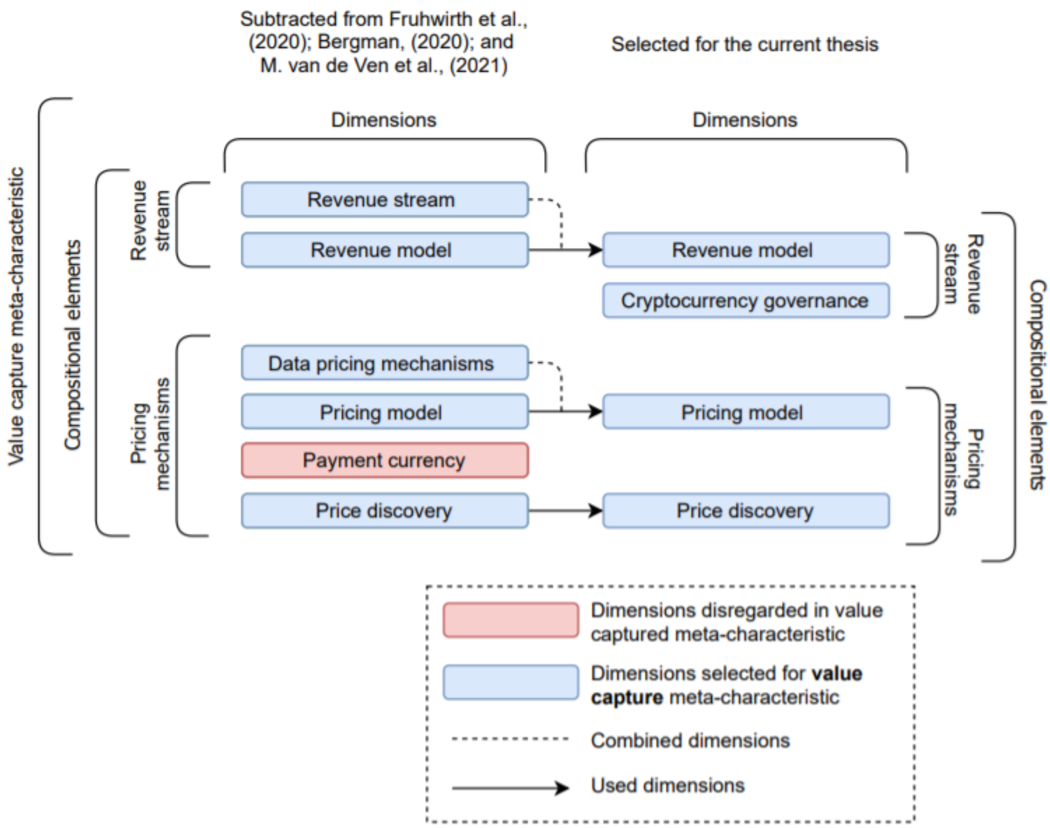

Figure 11: The alignment of compositional elements in the taxonomy for value capture

Concluding the research on value creation in section 4.3.1, value delivery in section 4.3.2, and value capture in section 4.3.3, the analyzed business dimensions are combined in figure 14, presenting the compositional elements for the business model taxonomy suitable for studying the data marketplaces with blockchain architecture. These business dimensions are used in the following section to analyze the documents of the selected decentralized data marketplaces active in the data industry.

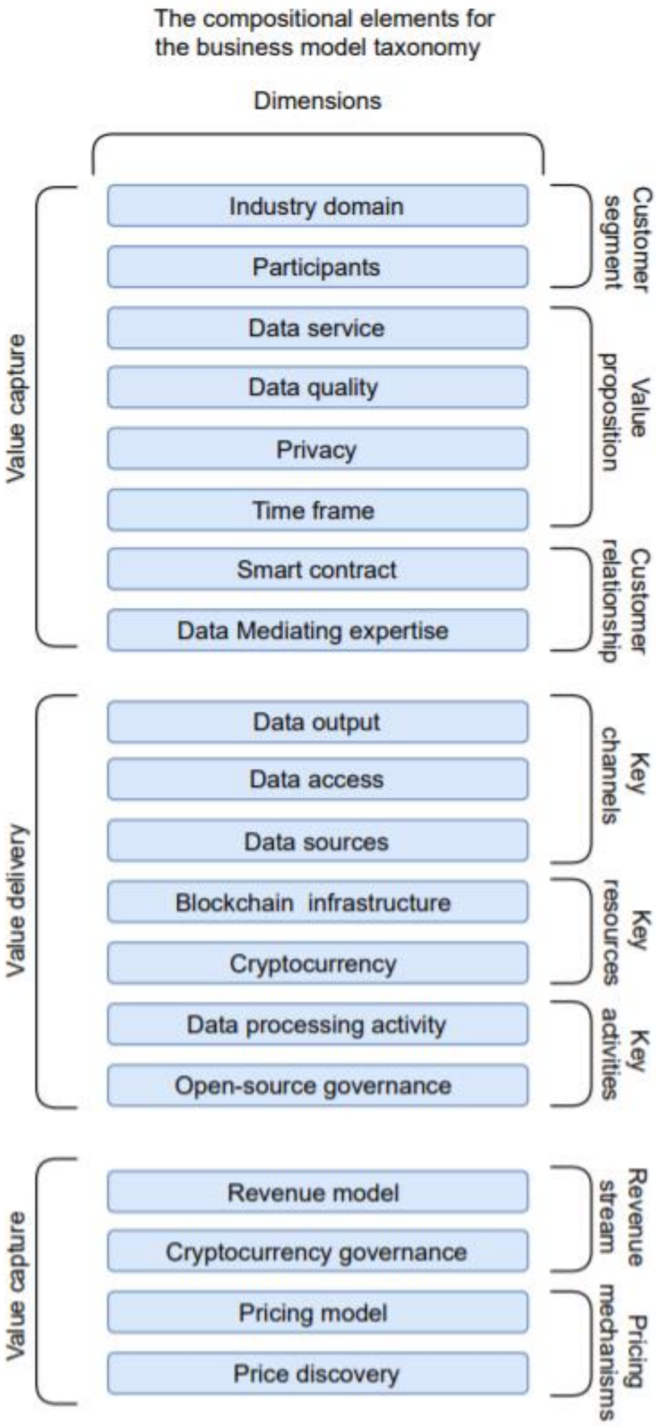

Figure 12: The compositional elements for the business model taxonomy

4.4. The Taxonomy Subordinate Elements

According to the compositional elements from section 4.3, in the following, the subordinate elements in the business models of five active decentralized data marketplaces are analyzed. These marketplaces are Ocean Protocol (OP), dHealth Foundation (dHF), Streamr (STR), Vetri (VET), and Enigma (EN).

The subordinate elements are analyzed according to marketplace websites, whitepapers, platform blogs, terms and conditions. Numerous insights could be subtracted from these documents. Although considering the multiple sources of qualitative data, the contents of the documents were codified to have a systematic and structured analysis. More specifically, coding is the process of labeling and organizing qualitative data to identify various themes or patterns. For example, by examining smaller samples from the documents, similar coding characteristics could be assigned to Web Applications or UI Engines. Both Web Applications or UI Engines represent the *standardized software* of data marketplaces in the data access dimension described in section 4.3.2. Thus, the final coding for this example was *standardized software* as one of the business characteristics for data access. The complete content analysis of the five marketplaces is attached in Appendix A.

The documents analysis and the results from the coding process are combined into the taxonomy of business models for decentralized data marketplace. In some dimensions, more subordinate elements could be identified if taken a larger sample of marketplaces. For example, according to M. van de Ven et al. (2021), the industry domain may also have the elements of geodata, financial & alternative data, audience data, sensor & mobility data. However, only two subordinate elements were observed in the five selected marketplaces: any data and health & personal data. In the same context, the data source dimension was excluded from the taxonomy because the five analyzed marketplaces have only customer-provided data, offering the infrastructure for data exchange, and they do not engage in acquiring or self-generating data.

This section is structured as follows. In the first half, the observations taken from marketplace documents are used to analyze each business dimension separately. Therefore, distinct characteristics of business models are identified. These business characteristics represent the subordinate elements that are aligned in the taxonomy from the second half of the section. As a result, the business characteristics aligned in the taxonomy form the second design artifact, which provides the insights necessary to answer the third sub-research question, *What are the characteristics elements in a taxonomy of business models for data marketplaces with a decentralized architecture?*

Industry Domain

The industry domain describes the categories of data offered by marketplaces. In the content analysis of the selected marketplaces, two categories of data domains could be distinguished: *any data* and *health & personal data*. Although, as earlier mentioned in the introduction of this section 4.4, more data domains could be identified by taking a larger sample of marketplaces. Regardless of the number of domains, an important observation in the analyzed documents of the five data marketplaces is that two major groups for the industry domain of data could be defined. On the one hand, marketplaces provide technical solutions for *domain-specific data* (e.g., healthcare). For this purpose, the business models of dHF and VET mainly focus on delivering an explicit technical platform for industry-specific stakeholders. On the other hand, marketplaces offer flexibility with *any data* domain to establish an ecosystem of data providers. Thus, the OP, STR, and EN business models are focused more on classifying their marketplaces as frameworks for different data providers to expand the data ecosystem.

Participants

Evident participants as *data consumers*, *data providers*, or *internal developers* are present in the business models for the selected marketplaces. Although, a difference in participants is noted in regards to *external developers*. Marketplaces that are longer active in the data market, like OP, STR, and EN, also encourage the participation of *external developers* by offering different grants through project governed funds. Furthermore, considering the technical characteristics of blockchain technology, a new category of participants is observed for all analyzed marketplaces. The new category of participants provides computational power or storage space to support the decentralized infrastructure. These participants are defined in the taxonomy as *node operators*.

Data Service

The primary observed service provided by marketplaces is *data brokering*. This service describes the fundamental characteristics of marketplaces to offer the infrastructure for data exchange. Additionally, marketplaces also adopt *customized map services* through various infrastructure properties or specialized applications. For instance, STR provides technical properties through the functionality of streaming real-time data. Thus, providers offer their data in stream topics to which consumers subscribe instead of publishing simple data packages. Another example is the specialized application offered by VET, where data providers connect their social accounts to share personal data.

Data Quality

A common perception for data marketplaces is the need to ensure data quality through different reviewing mechanisms. In the analyzed data marketplaces, three categories of reviewing mechanisms could be observed: *providers & consumers reviews*, *intermediaries reviews*, and *automated technical processes*. For *providers & consumers reviews*, the reviewing mechanisms are based on participants assigning trust values to trading parties. These trust values take various forms, as a security deposit, quality voting, or reputation systems. For instance, OP participants add liquidity, aka security deposit, as a proxy to dataset quality; in dHF, the supernode operators possess special voting rights regarding quality assurance; and in EN, data providers have reputation systems in offering qualitative data. Next, for *intermediaries reviews*, the reviewing mechanisms are based on the expertise of trusted intermediaries to audit the activity of providers and consumers. The example in this sense is the identity verification of VET that relies on third-party attribute verification. Lastly, marketplaces may review data quality themselves through *automated technical processes*. For instance, in the STR marketplace, providers gain "karma" when published data are delivered to subscribers; subsequently, subscribers earn "karma" by receiving data, making the reviewing mechanism technically automated.

Privacy

It is of primary importance to perceive privacy in data exchange processes. According to the technical specification of blockchain architecture, privacy in decentralized marketplaces is ensured by design through cryptographic mechanisms (see section 3.2.2). Thus, most of the analyzed marketplaces as OP, dHF, STR, EN, implement *encryption* methods for data provided in the market. Although in some cases, when personal data is aggregated and shared for different types of research, *anonymization* technics are also applicable. An example in this regard is VET, where providers are monetizing their personal data, and consumers gain direct access to reliable and anonymized data.

Time Frame

The time frame dimension describes whether marketplaces require frequent data updates to maintain relevancy (M. van de Ven et al., 2021). From the observations, the time frame of the marketplaces distinguishes between *up-to-date data*, *(near) real-time data*, or *multiple* classes of data. This indicates that the time frame is essential for data relevancy and that business models based only on static data are commonly disregarded. In the case of OP and dHF, the business model relying on *multiple* classes of data can be associated with the purpose of establishing the ecosystem of data providers. In this case, providers have the flexibility in data sharing. Contrary, VET, EN, and STR have a stronger conviction in the need for *up-to-date* or *(near) real-time* data to ensure relevancy for data categories offered in their data marketplaces.

Smart Contract

The relations between providers and consumers in the context of data marketplaces with blockchain architecture are safeguarded through programable, transparent, and self-executable smart contracts. Most analyzed marketplaces use the Etherium framework of smart contracts. Although, a difference is observed in smart contracts implementation associated with enabling *standard* or *configurable* properties. Therefore, on the one hand, smart contracts have a *standard* implementation with simple data exchange logic. The examples in this regard are dHF are VET, where *standard* smart contracts ensure transparency and self-execution of transactions once data is purchased. On the other hand, smart contracts have *configurable* properties where the term of agreements can be adjusted. For instance, OP, STR, and EN support smart contract upgrades and configurations according to the needs of data providers. However, these configurable properties are limited (e.g., setting the price per day/month or adjusting the minimum amount of constraints)

Data Mediating Expertise

The data mediating expertise represents the services like storage, aggregation, transformation, or quality validation, provided to consumers to enhance the usability of data exchanged in the market. Two characteristics could be distinguished in the analyzed marketplaces regarding the data mediating expertise: *provided by intermediaries* or *provided by marketplaces*. For the data mediating expertise provided by intermediaries, OP, dHF, and VET, encourage various third parties to offer their tools and value-added service in the market. Subsequently, contrary to the expertise provided by intermediaries, STR and EN implement their own usability layer with data-driven tools.

Data Output

The data output business dimension describes the technical delivery of data to end consumers. In the analyzed marketplaces, three observed data outputs are *standardized*, *aggregated*, or *both*. Technically, marketplaces choose the delivery methods based on the business use cases applicable for their data domain. For instance, STR and EN have a *standardized* output of any categories of data where consumers subscribe to existing data sets. Thus, STR and EN have a general specialization where any data sets are available in the market, applicable in various use cases. On the contrary, the use case of VET is specialized in offering personal data where larger samples of *aggregated* data are required to subtract more accurate mean values for different research purposes. Lastly, OP and dHF provide *both* standardized and aggregated forms of data output. For OP, the *standardized* output can be explained similarly to STR and EN with any data domain. Although considering their firm conviction in establishing the ecosystem of data providers, OP also delivers data in *aggregated* common pools. In a different context, for dHF, besides *aggregated* data useful for research purposes, healthcare practitioners also require *standardized* histories of patient data. Therefore, OP and dHF deliver data to the end consumers in *both* standardized and aggregated forms**.**

Data Access

Marketplaces can offer various technical possibilities for accessing data. Three data access options were observed in the analyzed data marketplaces: *standardized software*, *API*, and *multiple options*. The primary observed difference in data access between the selected marketplaces relies on the flexibility of data providers to decide on the access methods for their datasets. Thus, starting with STR, VET, and EN, these marketplaces offer *standardized software* for accessing data aiming to save effort in building, deploying, and monitoring smart contracts in the process of data commercialization. In this context, data providers are constrained in using the marketplace standardized software. Contrary, OP and dHF provide more flexibility in accessing data through *APIs* or *multiple options* (i.e., URLs). In this context, dHF and OP offer the exchange functionalities as an add-on to the existing ecosystem of data providers.

Blockchain Infrastructure

The blockchain infrastructure used by the analyzed marketplaces can be divided into *public (permissionless)*, *private (permissioned)*, and *hybrid* infrastructures. The differences between these infrastructures are related to the node operator rights in managing the distributed ledger (Zheng et al., 2017). Thus, starting with *public infrastructure*, note operators have complete rights to take part in managing the public ledger that registers transactions. Although considering the transparent registration of transactions in the public ledger, data providers clearly do not have any incentives to share data publicly without receiving any rewards. In this sense, OP and STR heavily rely on cryptography to encrypt data location in transactions between parties. In contrast to the *public infrastructure*, dHF implements a *private infrastructure*, where not all node operators have equal rights to manage the distributed ledger. Thus, only pre-selected nodes that are direct partners of dHF ensure validation and consensus in the data exchange process. The *private infrastructure* is designed to establish a trusted network where the motivation behind participants may not necessarily be monetary. Lastly, the combination of the two infrastructures can be observed with the *hybrid* approach of VET and EN. More specifically, the *hybrid* approach contains both a *public blockchain* and a *private blockchain*. On the one hand, the *public blockchain* registers the purchase of data, and on the other hand, the actual content of data is transferred through the *private blockchain*. This approach aims at ensuring both transparency and privacy in the complete data exchange process.

Cryptocurrency

According to the Swiss Financial Market Supervisory Authority, cryptocurrency can be classified into three categories of tokens: *payment tokens*, *utility tokens*, and *asset tokens*. The *payment tokens* are used as means of financial transfers or payments; the *utility tokens* provide digital access to applications or services based on blockchain; and *asset tokens* are designed as stocks, which are used as entitlements for dividends (Kim & Chung, 2019). In the analyzed marketplace documents, all five data marketplaces use their own *payment tokens* as means of currency in data transactions. Furthermore, besides *payment tokens*, OP and dHF also offer their *utility tokens* as digital access to datasets. Meaning that the holders of *utility tokens* can access specific datasets to which the tokens refer. The *asset tokens* were not observed as resources in the selected marketplaces. However, this can be related to the particular application for *asset tokens* as stocks that differ from the data exchange use case.

Data Processing Activity

The data processing activity describes whether marketplaces offer additional tooling on top of data exchange (M. van de Ven et al., 2021). Two characteristics of the data processing activity were observed in the analyzed data marketplaces: *all* and *limited* tools offered by marketplaces. The difference between *all* and *limited* tools relates to the dimension of data access, where marketplaces provide specialized software. More specifically, for VET and STR, *all* tools in managing data are offered through their specialized platform software. Contrary, OP and dHF do not bind participants to specialized platform software. Thus, OP and dHF provide *limited* tooling, enabling flexibility in data management. This approach has a stronger emphasis on establishing the ecosystem of data and service providers that have the incentives in a competitive market to maximize the utility of data through their own tools. Lastly, at the time of conducting the thesis research, the EN marketplace was still in the development phase, meaning that practical tools for data processing activity still needed to emerge. Therefore, EN is classified in *limited* tooling.

Open-Source Governance

The open-source governance dimension describes the management activity of data marketplaces in the open-source development process. According to open-source guidance, three governance structures are associated with open source development: *Benevolent Dictator*, *Meritocracy*, and *Liberal Contribution* (opensource.guide, 2021). The *Benevolent Dictator* or *BDFL* model stands for *"Benevolent*

Dictator for Life", which implies the leadership of a single person or organization in the decision-making process. This central decision-making authority differs in the *meritocracy* model, where participants also have formal decision-making power, describing consensus-based community projects. Lastly, the *Liberal Contribution* model recognizes the influence of participants according to their contribution; in other words, people that do the most work are recognized as most influential (opensource.guide, 2021). Only *Benevolent Dictator* and *Meritocracy* models could be observed in the analyzed data marketplaces. The absence of the *Liberal Contribution* can be explained that marketplaces are business entities sustained by formal organizations that still require some governance control mechanism. For the *Benevolent Dictator* model, STR and EN owners directly regulate the contribution to the project codebase. Both STR and EN have development programs, where marketplace owners analyze proposals from external parties and give final funding decisions. Contrary, the OP marketplace considers a different open-source governance approach with the *meritocracy* model. The OP marketplace also implements a development program. However, all proposals from external developers are publicly registered, and the community votes their fundings. Lastly, for dHF and VET, *no info* about open-source governance could be identified. dHF and VET have a more closed development community and less codebase publicly available. Although, this also could be related to the still inception project phase.

Revenue Model

The revenue model dimension describes the chosen approach of data marketplaces in the monetizion of value delivery. In the analyzed data marketplaces, different revenue models were observed. These differences could be associated with the key channels in value delivery and the appropriate monetization of these channels. In this context, starting with STR, the marketplace offers specialized software for outputting standardized data. For the monetization of this process, STR uses *commissions* to collect rewards from standard data exchange performed through their software. In the same context, VET also provides specialized software, but data is outputted into an aggregated form. Thus, data consumers subtract necessary insights from the market by *subscribing* to the aggregated data. A different revenue model approach is observed for OP and dHF. These marketplaces offer flexibility in outputting and accessing data, characterizing the marketplace infrastructure more as a framework for data exchange. Therefore, OP and dHF perceive *usage fees* from data providers that apply their data exchange framework. Lastly, with no reference to key channels, EN disregards the collection of commissions or fees in data exchange. However, to finance the marketplace development, EN *sells digital assets* (cryptocurrency) and raises funding from the community.

Cryptocurrency Governance

As earlier described in the cryptocurrency dimension, the analyzed marketplaces use their own cryptocurrency tokens as means of payment or utility resources. Although besides being evaluated as resources, cryptocurrency tokens also attract public attention, offering an alternative revenue stream for blockchain projects. More specifically, in the context of data marketplaces, the revenue stream originates from external investors acquiring cryptocurrency through specialized platforms, thus providing funds to marketplaces. Furthermore, considering that all transactions are performed using cryptocurrency, data providers and consumers are also required to invest in this type of digital currency. In this context, by establishing various strategies to raise the cryptocurrency value, marketplaces could benefit from a desirable revenue model utilizing crypto-economic systems. Several strategies as *sale, investment*, *incentivization*, *stake*, and *burning* were observed in the analyzed marketplaces. These strategies are described as follows.

Cryptocurrency Sale Strategy

The *cryptocurrency sale strategy* represents the process of offering cryptocurrency on specialized platforms in return for fiat currency. This strategy supplies fiat currency for marketplaces and provides liquidation abilities for cryptocurrency.

Crypto-Investment Strategy

The *crypto-investment strategy* describes the allocation of funds for development activities by substituting the costs with cryptocurrency. An essential aspect of this strategy is keeping the distribution of cryptocurrency inside marketplaces; thereby, reducing the withdrawal of fiat currency.

Incentivization Strategy

The *incentivization strategy* aims at programing the behavior of participants through incentive systems (Kim & Chung, 2019). These incentives can be designed to raise the cryptocurrency value and enhance system efficiency. For instance, EN offers cryptocurrency transaction fees for node operators that provide computational power in data exchange. Therefore, EN provides incentives for different actors to support the data exchange infrastructure and be rewarded for their contribution.

Crypto-Staking Strategy

Another strategy in the crypto-economy is *crypto-staking*. This strategy aims to decrease the velocity of cryptocurrency by restricting the withdrawal of fiat currency from the market. For example, data providers have to stake cryptocurrency tokens as a security deposit for data quality in the OP marketplace. Thus, the staked amount of cryptocurrency reduces the ability of fiat currency withdrawal.

Cryptocurrency Burning

The value of cryptocurrency is determined by the supply and demand of crypto tokens in the market. Therefore, the last strategy observed was *burning (destroying) cryptocurrency* tokens to reduce the supply and increase their value. For instance, as earlier described in open-source governance, OP implements a development program that makes available funds for a certain period. In this context, if allocated funds for the specific period are not consumed, these funds are burned, thus reducing the cryptocurrency tokens supply aiming at increasing their value.

The described cryptocurrency strategies are subtracted from the observations of the analyzed data marketplaces. For example, OP, STR, and EN implement all the described strategies as data marketplaces with a longer presence in the cryptocurrency market. Regarding VET, the investment strategy is absent, and for dHF, the cryptocurrency burning strategy could not be observed. Although, these data marketplaces are still in the inception phase, where most of the attention is focused on raising funds through cryptocurrency selling.

Pricing Model

The pricing model represents how final prices for data assets or services are composed (M. van de Ven et al., 2021). In the analyzed data marketplaces, three pricing models were observed. For OP and dHF, consumers have to pay fixed prices for the datasets offered in the market. Thus, OP and dHF implement the *fixed pricing model*, which is related to the observation that these marketplaces sustain an extensive industry domain of datasets where fixed pricing represents a straightforward model. A different pricing approach can be observed in VET with the *pay-per-use pricing model*. This pricing model enables consumers to pay only for the part of data they require. VET offers personal data in an aggregated form. Thus the need to consume only parts of the aggregated data for specific use cases suits the *pay-per-use pricing model*. Another pricing model is the *flat fee tariff* implemented by STR and EN. This pricing model is based on time as the attribute for price calculations. In this regard, providers specify prices according to periods (i.e., days, weeks, months) for subscription to their data. This approach could be related to the use cases of STR and EN in providing up-to-date or real-time data.

Price Discovery

The price discovery dimension describes the method of marketplaces to directly determine the prices for different categories of datasets or data providers to set their own prices. In the analyzed marketplaces, OP, dHF, STR, and EN enable data providers to have the flexibility to set prices according to their own evaluations. A different approach can be observed in VET, where the marketplace has specific categories of personal data for which prices are predetermined; thereby, providers offer data for already priced categories.

4.5. The Conceptualization of a Taxonomy of Business Models for Decentralized Data Marketplaces

With the compositional elements studied in section 4.3 and the analysis of the marketplace documents from section 4, the taxonomy of business models for decentralized data marketplaces is conceptualized in the following table 4. The taxonomy is designed according to four main components that were investigated in the business models across the entire chapter. These four components are the business model meta-characteristics, compositional elements, dimensions, and subordinate elements.

	Compositional element	Dimension	Characteristics (Subordinate Elements)			
Value creation	Customer segment	Industry domain	Any data (OP, STR, EN)	Domain-specific data (dHF, VET)		
		Participants	Data providers, data consumers, internal and *external developers*, node operators (OP, STR, EN)	Data providers, data consumers, internal developers, node operators (dHF, VET)		
	Value proposition	Data service	Data brokering service (OP, dHF, EN)	Customized map service (STR, VET)		
		Data quality	Automated technical processes (STR)	Providers & consumers review (OP, dHF, EN)	Intermediaries reviews (VET)	
		Privacy	Encrypted (OP, dHF, STR, EN)	Anonymized (VET)		
		Time frame	Up-to-date (VET, EN)	(Near) real-time (STR)	Multiple (OP, dHF)	
	Customer relationship	Smart contract	Standardized (dHF, VET)	Configurable (OP, STR, EN)		
		Data mediating expertise	Provided by marketplace (STR, EN)	Provided by intermediaries (OP, dHF, VET)		
Value delivery	Key channels	Data output	Standardized (STR, EN)	Aggregated (VET)	Both (OP, dHF)	
		Data access	Specialized software (STR, VET, EN)	API (dHF)	Multiple options (OP)	
	Key resources	Blockchain infrastructure	Public (permissionless) (OP, STR)	Private (permissioned) (dHF)	Hybrid (VET, EN)	
		Cryptocurrency	Payment tokens (STR, VET, EN)	Payment tokens and utility tokens (OP, dHF)		
	Key activities	Data processing activity	All (VET, STR)	Limited (OP, dHF, EN)		
		Open-source governance	Benevolent Dictator (STR, EN)	Meritocracy (OP)	No info (VET, dHF)	
Value capture	Revenue stream	Revenue model	Commissions (STR)	Subscriptions (VET)	Usage fees (OP, dHF)	Asset sale (EN)
		Cryptocurrency governance	Sale, investment, incentivization, stake, burning (OP, STR, EN)	Sale, investment, incentivization, stake (dHF)	Sale, incentivization, burning (VET)	
	Pricing mechanisms	Pricing model	Fixed pricing (OP, dHF)	Pay-per-use (VET)	Flat fee tariff (STR, EN)	
		Price discovery	Set by the marketplace provider (OP, dHF, STR, EN)	Set by the marketplace (VET)		

Table 4: The Taxonomy of Business Models for Decentralized Data Marketplaces

4.6. Conclusion of Chapter 4

In this chapter, a component-based approach was used to study the business models of data marketplaces with blockchain architecture. For this purpose, the compositional elements defining what a business model is made of were investigated. Subsequently, the subordinate elements describing how decentralized data marketplaces create, deliver, and capture value were analyzed. Lastly, the findings from compositional and subordinate elements were combined into a taxonomy of business models for decentralized data marketplaces. The taxonomy formed the design artifact that answered the third sub-research question: *What are the characteristics elements in a taxonomy of business models for data marketplaces with a decentralized architecture?*

The final taxonomy comprised 3 meta-characteristics according to the rationale of how data marketplaces *create* and *deliver* value to customers and then convert payments received to profits, describing the *captured* value (Teece, 2010). These meta-characteristics were studied using the business model canvas as the leading framework (Osterwalder & Pigneur, 2010). For this purpose, 8 compositional elements were distinguished from the business model canvas. To accomplish a focused analysis of data marketplaces, the compositional elements were divided into 18 business dimensions subtracted from previous studies (Bergman, 2020, M. van de Ven et al., 2021, and Fruhwirth et al., 2020). Furthermore, considering the thesis scope, these dimensions were aligned according to the specification of blockchain technology. Lastly, with the identified business dimensions, the documents of five decentralized data marketplaces with blockchain architecture active in the data industry were analyzed. These analyses revealed 46 business model characteristics.

Further, the taxonomy is used in the discussion chapter to describe how decentralized data marketplaces sustain comprehensive business models and create value for actors engaging in data exchange.

5
Formative Evaluation

According to the research approach from section 1.4, the first part of the thesis resembled the design cycle of the DSR methodology as described by Hevner (2007). This methodology supports the so-called construction problems, evident in the objectives of previous chapters with the construction of the artifacts. Another key activity in DSR is the evaluation of the designed artifacts, as it provides input for further development and research rigor. Therefore this chapter marks the start of the first evaluation phase, defined as formative evaluation. Numerous guidance on the evaluation of DSR projects can be found in the literature (Verschuren & Hartog, 2005, Peffers et al., 2012, Venable et al., 2016). This chapter uses the evaluation framework as described in Venable et al. (2016). The Framework for Evaluation in Design Since (FEDS) is well-cited and fits the current chapter's goals and objectives for several reasons. First, the framework clearly describes the purpose of formative evaluation as the process for improving the characteristics of the evaluand. This purpose fits the evaluation objectives of this chapter to improve the designed artifacts. Second, DSR evaluation has different problems like Type I or Type II erros, known as false positive or false negative errors. For this purpose, FEDS provides distinct guidance and solutions for these evaluation problems with a more practical approach and less philosophical. Lastly, the framework consists of an explicit design process of why, when, how, and what to evaluate. This design process is efficiently used in the chapter to evaluate the designed artifacts.

5.1. Formative Evaluation Design

The evaluation design is first explained in this section; thereafter, the evaluation results are discussed in section 5.2. Therefore, the formative evaluation is defined in section 5.1.1. The problems like Type I or Type II errors with formative evaluation are described in section 5.1.2. Lastly, in section 5.1.3, the evaluation process is presented according to the steps of the FEDS framework.

5.1.1 Defining the Formative Evaluation

According to Venable et al. (2016), formative evaluation *"is used to produce empirically-based interpretations that provide a basis for successful action in improving the characteristics or performance of the evaluand"*. In this chapter, the evaluands are the *decentralized HLA* and the *Taxonomy of Business Models for Decentralized Data Marketplaces*. These artifacts were consulted with researchers and industry experts in the subject areas of data marketplaces. The consultation produced various empirically-based interpretations, characterized as recommendations for improving the artifacts. As a result, consulting the experts and gathering their recommendations resembles the research task associated with this chapter to evaluate the artifacts obtained from the technical and business requirements specifications.

5.1.2 Problems with Formative Evaluation

As previously introduced, evaluation methods have different problems like Type I or Type II errors (Venable et al., 2016). In DSR, the Type I and Type II errors are known as false-positive or false-negative errors. For instance, a false-positive result in the HLA design could indicate that the architecture works when in fact, the architecture is not logically structured. In this sense, FEDS provides specific guidance for the evaluation process. The framework distinguishes between *artificial evaluation* and *naturalistic evaluation*. The *artificial evaluation* uses interpretative technics to understand why the artifacts work. Thus, by adding the interpretation of different experts, their recommendations reduce possible errors, clarify misunderstandings, and strengthen the generalized interference of the designed artifacts. Further, in the *naturalistic evaluation*, the goal is to explore the performance of a technical solution in its real environment. For instance, the designed taxonomy or the HLA could be implemented in market-active data marketplaces to evaluate their properties. However, the thesis has a limited time frame. Thus, the real environment application represents a limitation for this chapter. As a result, considering that the naturalistic evaluation is a limitation, only the *artificial evaluation* is applied**.**

5.1.3 Evaluation Process

On the basis of FEDS, the evaluation process is divided into four steps: (1) explicate the goal of the evaluation, (2) choose the evaluation strategy, (3) determine the properties to evaluate, and (4) design the individual evaluation episode. These steps are used in this chapter and are described as follows.

Step 1: Evaluation Goals

In developing evaluation components for DSR, there are at least four distinct competing goals. The goals are *(i) rigour*, *(ii) uncertainty and risk reduction*, *(iii) ethics*, and *(iv) efficiency* (Venable et al., 2016). These goals are also relevant for evaluating the HLA and the Taxonomy of BMs.

(i) Rigour goal has two senses: one is to ensure that only the artifact causes the observation and that no independent variables are included. The other sense is to ensure that the artifact works in a real environment. For the former sense, *completeness* is evaluated for both researched artifacts. The *completeness* ensures that various perspectives of the architecture and the taxonomy are included in the final results. The second sense represents a limitation for this chapter. Real environmental evaluation requires more time for implementation. Although to be mentioned, the taxonomy analyses were conducted on active data platforms.

(ii) Uncertainty and risk reduction goals describe the process of reducing risks due to design uncertainties. These goals in the evaluation of the two artifacts are of particular importance, considering the complexity of blockchain technology and the business models. Two categories of risks could be identified. First, for the HLA, *technical risks* could the recognized. These risks may indicate that blockchain technology can not be made functional in the designed architecture. For this purpose, the HLA is evaluated based on the *logical structure*. Second, for the taxonomy, the risks of use can be identified. The use risks emphasize that the taxonomy could not fit the analysis of BMs for data marketplaces and therefore could not be applicable in studying these digital platforms. For this purpose, the Taxonomy of BMs for Decentralized Data Marketplaces is evaluated according to the *detailedness*.

(iii) Ethics goals address the risks of organizations and their users. However, these types of risks are more relevant for critical system artifacts or technologies. The current chapter evaluates only conceptual artifacts. In this sense, the ethics evaluation is disregarded for the HLA and the Taxonomy of BMs. Furthermore, the evaluation activity itself should not put the consulted experts at risk. For this purpose, each consulted expert was approached individually through a personalized email. They were given background information about the thesis as well as the evaluation process and questions to be asked. In addition, experts were able to ask questions about the thesis research. Lastly, the evaluation process was completely voluntary; thereby, experts could withdraw from the evaluation at any time without having to give reasons.

(iv) Efficiency goal balances the above *ethics, rigour, uncertainty and risk reduction* against available resources. In the case of the current thesis, time is the only consumed resource. Thus, in choosing the evaluation goals, the time frame represents the constraints that reduce possible extra evaluation activities. The activities limited by time are as follows. First, only artificial evaluation is applied. This non-empirical evaluation is less time-consuming compared to naturalistic evaluation. Second, in regard to the rigour goal, only the completeness criteria are evaluated for both researched artifacts. Other possible criteria for the rigour goal as real environment usability, testability, or reliability

are not considered in the evaluation with experts. Third, for *uncertainty and risk reduction* goals, only the logical structure for technical risks and the detailedness for the use risks are evaluated. Other evaluations criteria for uncertainty and risk reduction as security or correctness are limited in this evaluation chapter.

Step 2: Evaluation Strategy

Considering the evaluation goals from the previous step and in compliance with FEDS, *Technical Risk and Effectiveness* is the strategy used for evaluation. This strategy is chosen for several reasons. First, the strategy underlines formative evaluation early in the process. This characteristic fits the purpose of evaluating the two conceptual artifacts that are still in the initial research phase, considering that a real environment implementation still represents a limitation. Second, the design risks of the two artifacts are technical-oriented. Both the architecture and the business models are researched in the context of blockchain technology. Third, for the two artifacts and the use of blockchain technology, the strategy is appropriate for the evaluation of significant technical and usage uncertainties with a strong requirement to establish long-term effectiveness.

The evaluation method for the *Technical Risk & Effectiveness* strategy used in this chapter is semi-structured interviews. Six semi-structured interviews were conducted with researchers and industry experts in the subject areas of data marketplaces. Two aspects describe the semi-structured approach. The first "structured" aspect relies on pre-defined questions. These questions are directly related to the evaluation goals of the designed artifacts. The second "semi-"structured aspect represents the follow-up questions that appeal to an open discussion to clarify the responses of researchers and industry experts. As a result, the semi-structured interviews helped to gather the different empirically-based interpretations to improve the researched artifacts.

Step 3: Evaluation Properties

The evaluation properties in the FEDS framework represent the formulation of what to evaluate. Therefore, on the basis of the previously explained evaluation goals, the evaluation properties are described as follows. First, considering the *rigour* goal to ensure that only the artifact causes the observation and that no independent variables are included, the evaluation property for both artifacts is *completeness*. Second, for the *uncertainty and risk reduction* goals, two distinct evaluation properties were identified. On the one hand, the *logical structure* and the *comprehensibility* of components are introduced to reduce the technical risks associated with the HLA. On the other hand, the *detailedness* is evaluated to enhance the usability of the Taxonomy of BMs. As a result, the evaluation of the described properties aims at answering the fourth and fifth sub-research questions.

Is the decentralized architecture for data marketplaces evaluated as comprehensible, logically structured, and complete?

Is the taxonomy of business models for data marketplaces with a decentralized architecture evaluated as detailed and complete?

The designed artifacts, examples of interview questions, and properties subject to evaluation are presented in the following table 5.

Research artifacts	Evaluation properties	Interview questions
Evaluate the High-Level Architecture	Comprehensible	Are the infrastructure components comprehensive in naming, definition, and functional requirements?
	Logical structure	Is the High-Level Decentralized Architecture for Data Marketplaces logically structured?
	Completeness	Are infrastructure components missing in the High-Level Architecture to describe individual functional requirements?
		Are architecture layers missing in the High-Level Architecture to describe the technical characteristics of decentralized data marketplaces?
Evaluate the Taxonomy of BMs for Decentralized Data Marketplaces	Detailedness	Are the Business Model Characteristics for Decentralized Data Marketplaces explicitly detailed?
	Completeness	Are business dimensions missing in the taxonomy to describe the business models for the decentralized data marketplace?
		Are characteristics missing in the taxonomy to describe a particular business dimension?
Recommendations	-	What are the expert's recommendations of change to improve the designed artifacts?

Table 5: The formative evaluation properties for the design artifacts

Step 4: Individual Evaluation Episodes

Having chosen the evaluation goal, strategy, and properties, six researchers and industry experts were interviewed. These researchers and industry experts are actively involved in projects with data management characteristics and with affiliation to data marketplaces. Background information about the thesis, a structured agenda, and interview questions were sent three days in advance. The interview process for the evaluation of artifacts contained three distinct parts. First, the designed artifacts were explained. This concerned a ten minutes presentation with a detailed explanation of the researched concepts. At the end of the presentation, experts were encouraged to address questions if the presentation or any of the described artifacts were not clear. The second interview part marked the start of the artifacts evaluation. Particularly for the second part, the HLA was evaluated. Thus, considering the earlier described properties in step 3 regarding comprehensibility, logical structure, or completeness, several questions were asked in the interview. It has been noted that questions about the logical structure were hard to be answered. This is due to the topic complexity of data marketplace architecture in general and additionally in the context of blockchain technology. Moreover, interviews were limited to one hour; thereby, time was also a constraining resource. In this regard, if experts encountered difficulties in answering the questions related to the logical structure of the HLA, then more inclusive questions associated with the architecture design were discussed. For instance, experts were asked if HLA follows a common architecture pattern they could encounter in practice; if they could identify explicit distinctions between the architecture layers; and if any of the infrastructure components could be modified or removed. As a result, if experts could not identify any inconsistencies, this would be perceived as an indicator of the HLA logical structure. The third and final part of the interviews involved the evaluation of the Taxonomy of BMs for Decentralized Data Marketplaces. According to the evaluation properties of detailedness and completeness, several questions were addressed. Moreover, one example of the taxonomy application was presented with the information from the content analysis. Experts were asked if they could specify if the taxonomy is explicitly detailed; if more business dimensions for a certain component could be included; or if any of the business dimensions could be removed. Therefore, if experts could not identify additional business dimensions or distinguish business dimensions to be removed, this would indicate the detailedness and completeness of the taxonomy. More details about the interviewed researchers and industry experts as well as two examples of the interview transcripts are attached in Appendix B.

5.2. Formative Evaluation Results

In this section, the results from the interviews with researchers and industry experts are presented. As explained in step 4 from the individual evaluation episode, six experts in the data management field were interviewed. In the following, their recommendations for improvements were integrated into the HLA and the Taxonomy of BMs. The evaluation results of this section are structured as follows. First, the evaluation results from the HLA are described. In compliance with step 3 from the evaluation properties, the HLA is evaluated according to its *comprehensibility*, *completeness*, and *logical structure*. Second, the evaluation results from the taxonomy of BMs for decentralized data marketplaces are described. In the same context, according to step 3 from the evaluation properties, the taxonomy is evaluated on *detailedness* and *completeness*.

5.2.1. Evaluation of the High-Level Architecture

This section describes the evaluation of the HLA and the recommendations for improvements collected from researchers and industry experts. According to step 3 from evaluation properties, the HLA is evaluated according to its *(i) comprehensibility*, *(ii) completeness*, and *(iii) logical structure*. The results from the evaluation are described as follows.

(i) comprehensibility: researchers and experts evaluated that most of the components are intuitive. The things that remain for the text of the thesis are the well-structured arguments describing these components. Moreover, the architecture does not strike with anything obscure; it follows common sense. As a result, experts could not identify any comprehensibility issues. This was perceived as an indicator that the naming, definition, and functional requirements are insightful for the HLA.

(ii) completeness: several recommendations were suggested. First, the HLA is missing the third parties. Besides the data providers and consumers, they are still the parties who, for example, verify the attributes in identity management. Second, the HLA defines a relatively static representation. To see if the HLA works, it would require putting dependencies between the components.

(iii) logical structure: experts introduced a couple of improvement recommendations. Starting with having a layered architecture between components typically does not contain broken (interacting) layers. The smart contract layer in the HLA suggests that there are two different smart contracts, which should not be the case. Another recommendation is regarding the horizontal security layer; it could be represented as an infrastructure security aspect. Better to add a horizontal layer of other components, such as authentication, assets discovery, and access negotiation within architecture.

The recommendations of experts from this section were integrated into the HLA. These recommendations are summarized in the following table 6.

Evaluation properties	Recommendations for improvement	Thesis integration
Comprehensibility	-	Experts could not identify any comprehensibility issues. This is perceived as an indicator that the naming, definition, and functional requirements are insightful for the HLA
Completeness	Missing the third parties	This recommendation was given at the early phase of researching the HLA architecture. Thus, in the access layer of the HLA, the third parties were already included (see section 3.4)
	Putting dependencies between the components	In the layered architecture, the components are self-independent, meaning that they are interconnected but not dependent on each other. Requests should go through all layers from top to bottom; in this sense, layers are closed. Therefore, no dependencies were added within the architecture.
Logical structure	Broken (interacting) layers	This recommendation represented an insightful observation because, in the case of a layered architecture, layers should be self-independent. For this purpose, the HLA was restructured in the following figure 13.
	Infrastructure security aspect	With the recommendations from experts, the security layer was renamed into the infrastructure security layer, and the "authentication" suffix was added to integrity management. Additionally, the cryptographic mechanisms are included in this layer as the component to which every infrastructure component should be bound to ensure self-sovereign identity.

Table 6: The integration of the evaluation properties in the High-Level Architecture

Based on the recommendations gathered from researchers and industry experts, the HLA was restructured in the following figure 13. According to the remarks regarding the broken layers, the most significant modification was the separation of the smart contract, DLT, and DLT 2 layers. Furthermore, a particular change in this sense was the exclusion of the DLT2 layer. Technically, the DLT 2 layer described the interoperability between the on-chain and off-chain environments. This interoperability framework may also be achieved through the functional properties of the two environments. More specifically, the functional properties of the DLT2 layer were separated as follows. The interoperability framework is included in the on-chain environment; the network-verifiable proof is included in the smart contract layer as part of data exchange; and the blockchain connector is included in the off-chain environment.

Further, several new components were added to describe the technical functionalities of the on-chain and the off-chain environments. First, for the on-chain environment, the smart contract deployment component is introduced as the functionality of deploying the self-executable programs of the smart contracts. Second, for the same on-chain environment, the distributed ledger is included to describe the sequence of blocks in the blockchain environment. Third, considering that there are no central blockchain entities to manage transactions, consensus protocols between network participants are required (see section 3.2.3). For this purpose, the consensus protocol component is also included in the on-chain environment.

Moving on to the off-chain environment, the federated network of known nodes is incorporated to represent the main functionality of securely exchanging data into a decentralized network. This component represents the network of participants with security-related assets, codes, and underlying trusted operating systems to form the secure area for processing transactions. The participants in the off-chain environment also require consensus protocols to manage data transactions. Therefore, the consensus protocol is part of the off-chain environment. Lastly, as described in section 3.3.7 of the data exchange, the off-chain environment could integrate mechanisms of multi-party computation (MPC) and algorithms for data analysis. For this purpose, the computation data algorithms component is introduced in the HLA.

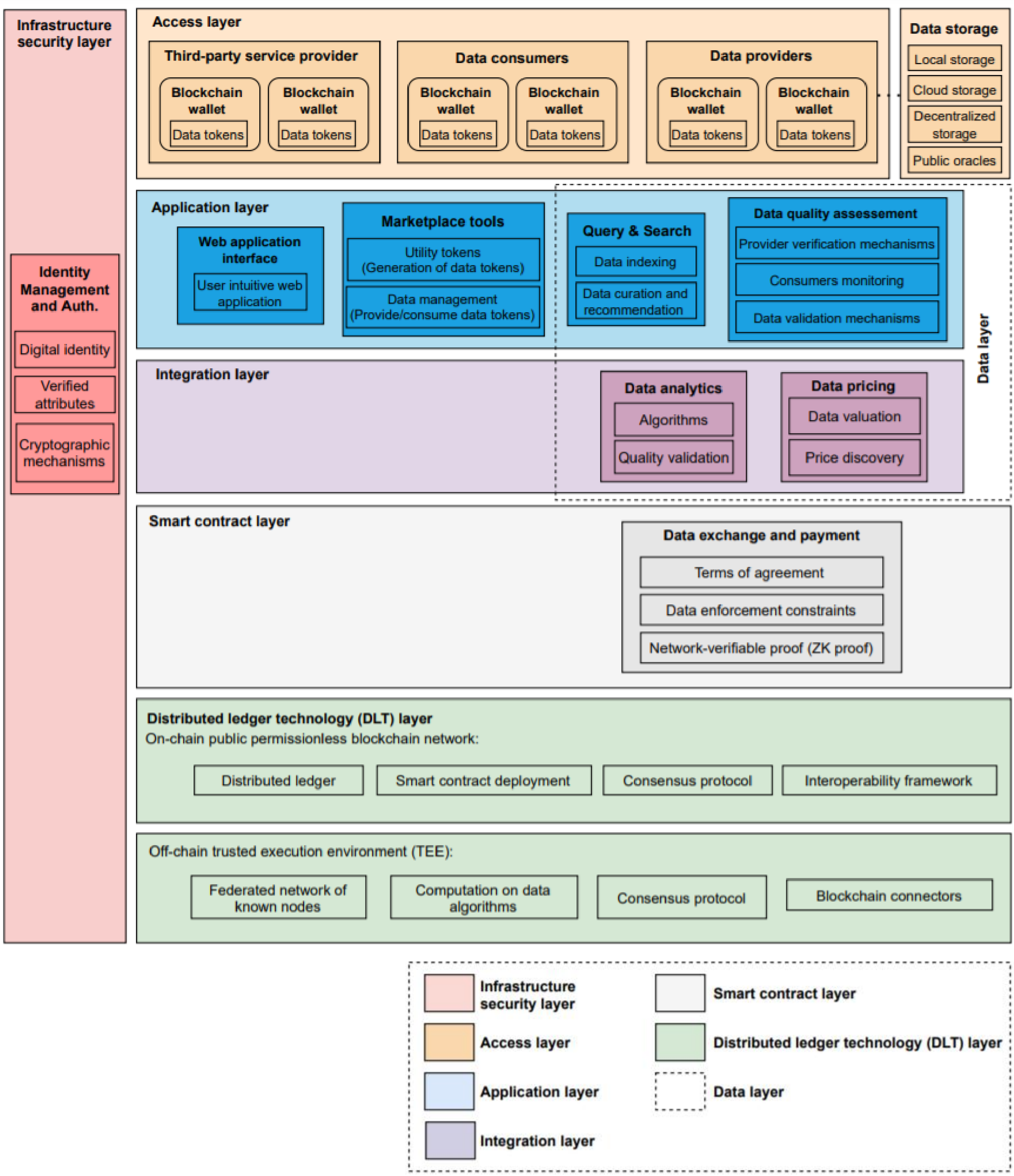

Figure 13: The Evaluated High-Level Architecture for data marketplaces with blockchain architecture

5.2.2. Evaluation of the Taxonomy of Business Models

With the evaluated HLA from previous section 5.2.1, the second design artifact that was subjected for evaluation was the Taxonomy of Business Models for Decentralized Data Marketplaces. According to step 3, the evaluation properties of the taxonomy are *(i) detailedness* and *(ii) completeness*. These properties were evaluated with the same researchers and experts from section 5.2.1 to gather their recommendations for improving the taxonomy. Their recommendations are described as follows.

(i) detailedness: two recommendations for this evaluation property were proposed. The first recommendation for improvement was to identify if business components can be divided into more than one dimension. Subsequently, the second recommendation suggested by experts was to avoid labeling the business component and the dimension with the same naming.

(ii) *completeness:* experts evaluated that the taxonomy of BMs fit-for-purpose of analysis the decentralized data marketplaces on an aggregated level. Most of the experts evaluated the taxonomy as completed. This was perceived as an indicator of completeness for the business model taxonomy.

<table>
<tr><th>Evaluation properties</th><th>Recommendations for improvement</th><th>Thesis integration</th></tr>
<tr><td rowspan="2">Detailedness</td><td>To introduce more than one dimension</td><td rowspan="2">According to the literature search from section 4.1, one of the primary articles used in analyzing the taxonomy of BMs was Bergman (2020). The revenue stream initially was deducted from that paper. As a result, one compositional element was labeled as "revenue stream" with a single business dimension labeled the same name "revenue stream". In implementing the remark of avoiding labeling the BM components and dimensions with the same naming, the BM dimension was changed to "revenue model" which is also present in M. van de Ven et al. (2021). Additionally, after more in-depth analysis, a new revenue stream was identified for blockchain data marketplaces as cryptocurrency governance.

Both of these recommendations were given at the early phase of researching the taxonomy of BMs. Therefore, they were already included in the taxonomy from section 4.5.</td></tr>
<tr><td>To avoid labeling the business component and the dimension with the same naming</td></tr>
<tr><td>Completeness</td><td>-</td><td>Experts evaluated that the taxonomy fit-for-purpose to research the decentralized data marketplaces. Therefore, this was perceived as an indicator of completeness for the business model taxonomy.</td></tr>
</table>

Table 7: The integration of the evaluation properties in the Taxonomy of BMs

In the evaluation of the taxonomy of BMs, the consulted researchers and industry experts could not identify numerous inconsistencies regarding the detailedness and completeness of the taxonomy. The two recommendations for improving the detailedness and completeness were already implemented during the design of the artifact in section 4.5. The few recommendations may also be explained due to the fact that the taxonomy was constructed according to similar studies. Thus, the taxonomy had an essential reference to already studied characteristics. Although besides the recommendations from researchers and experts, the taxonomy could also be evaluated according to a comprehensive set of objective conditions. In the following, these objective conditions are subtracted from Nickerson et al. (2013), which offers a structured and well-cited approach in the field of information systems to develop and evaluate taxonomies. According to Nickerson et al., (2013), these objective conditions are tested at the end of each phase of the taxonomy development to assist the taxonomy researcher in determining when the process may stop. Four significant phases were required to meet all the objective conditions, described as follows. In the first phase, the taxonomy compositional elements were aligned according to the business model canvas framework. During the second phase, the business dimensions were derived from previous studies. In the third phase, the business dimensions and characteristics were researched in the documents of five active data marketplaces. Lastly in the fourth phase, the taxonomy was revised by comparing the business model characteristics with the case documents. The last phase ensures that by mapping individual marketplaces through the taxonomy characteristics, the business dimensions of the marketplaces correspond to the case documents. For example, what is the value proposition of STR? In this context, mapping the characteristics of STR in the value proposition of the taxonomy, the following can be defined: STR provides customized map service through specialized software with data quality ensured by the platform. For STR, privacy is perceived through data encryption, and mostly real-time data is offered to maintain relevancy in data exchange. According to the case documents, this definition could describe the value proposition of STR.

With the defined phases, Table 8 summarizes the objective conditions subtracted from Nickerson et al. (2013) and the evaluation reasoning for the taxonomy of BMs.

Taxonomy research phases overview: (1) Alignment of compositional elements; (2) Derived business model dimensions; (3) Researched business model characteristics; (4) Mapped individual marketplaces through the taxonomy characteristics

	Objective conditions	Research phases				Evaluation reasoning
		(1)	(2)	(3)	(4)	
1	All objects or a representative sample of objects have been examined			x	x	After researching the business model characteristics in the case documents from phase 3, the selected decentralized marketplaces were analyzed, fulfilling the first objective condition.
2	No object was merged with a similar object or split into multiple objects in the last iteration	x	x	x	x	The business models of independent decentralized data marketplaces were analyzed; therefore, no conflicting objects could be encountered.
3	At least one object is classified under every characteristic of every dimension			x	x	With the marketplace documents analysis from phase 3, the business dimensions and characteristics contain at least one selected data marketplace.
4	No new dimensions or characteristics were added in the last iteration	x		x	x	In the context of blockchain technology, in phase 2, new dimensions as data mediating expertise for value creation, open source governance for value delivery, or cryptocurrency governance for value capture were added (see sections 4.3.1, 4.3.2, and 4.3.3).
5	No dimensions or characteristics were merged or split in the last iteration	x		x	x	Similar to the previous objective condition, in phase 2, considering the similarities between business characteristics, a couple of dimensions were merged as follows. For value creation, the review system was merged to data quality, the contract was merged into the smart contract, and the domain into the industry domain. For value delivery, platform access was merged into data access, data origin into data sources, and data processing activity into data processing and analytical tools (see sections 4.3.1, 4.3.2, and 4.3.3).
6	Every dimension is unique and not repeated (no dimension duplication)	x	x		x	In phase 3, the data source dimension was excluded from the taxonomy because the five analyzed marketplaces have only customer-provided data, offering the infrastructure for data exchange, and they do not engage in acquiring or self-generating data.
7	Every characteristic is unique within its dimension (no characteristic duplication within a dimension)			x	x	After analyzing the marketplace documents, unique characteristics could be subtracted and aligned according to the business model dimension.
8	Each cell (combination of characteristics) is unique and is not repeated (no cell duplication)			x	x	With the recommendation from experts for avoiding labeling the business component and the dimension with the same naming, the revenue stream dimension was adjusted to the revenue model.

Table 8: The evaluation of the Taxonomy of BMs according to the objective conditions of Nickerson et al. (2013)

5.3. Conclusion of Chapter 5

In this chapter, the FEDS framework of Venable et al. (2016) was used to evaluate the designed artifacts. More specifically, the HLA and the Taxonomy of BMs were evaluated. According to FEDS, the evaluation process included: the *evaluation goals*, the *evaluation strategy*, the *evaluation properties*, and the *design of individual evaluation episodes*.

There are at least four distinct competing evaluation goals in DSR: *rigour*, *uncertainty & risk reduction*, *ethics*, *and efficiency*. These evaluation goals were used for the two designed artifacts as follows. The objective of the *rigour* goal was to ensure that only the artifacts caused the observation and that no independent variables were included. For this purpose, the two artifacts were evaluated on *completeness*. Next, two categories of risks were identified for *uncertainty & risk reduction.* First, the *technical risks* were associated with the HLA, and second, the *risks of use* were linked to the taxonomy. To reduce these evaluation risks, the HLA was evaluated according to its *logical structure*, and the taxonomy was evaluated on *detailedness*. Further, in regards to *ethics*, the evaluation activity itself should not put the consulted experts at risk. Thus, background information about the thesis was given to experts in advance, as well as the questions to be asked. Furthermore, the evaluation process was completely voluntary. Thereby, experts could withdraw from the evaluation at any time. Lastly, for the *efficiency goal*, time was described as the only consumed resource. Therefore, according to the available time, only the *artificial evaluation* was applied in the evaluation process.

Considering that the design risks of the two artifacts were technical-oriented, the evaluation strategy chosen in the chapter was *Technical Risk & Effectiveness*. The evaluation method used for the chosen strategy was *semi-structured interviews*. These interviews were conducted with researchers and industry experts in the subject areas of data marketplaces. As a result, the semi-structured interviews helped gather the different empirically-based interpretations to improve the designed artifacts.

With the evaluated, designed artifacts, this chapter marked the final phase of artifacts development. The recommendations for improving the artifacts were integrated into the thesis research in sections 5.2.1 and 5.2.2. Moreover, the evaluation of the designed artifacts introduced new insights to be approached in the following summative evaluation chapter.

6
Summative Evaluation

In DSR, formative and summative evaluation characterizes the two main distinctions for why to evaluate (Venable et al., 2016). This distinction inhabits the functional purpose of evaluation. More specifically, the functional purpose (why to evaluate) of the previous formative evaluation chapter was to improve the researched artifacts. In a different perspective, this summative evaluation chapter uses the refined artifacts to derive empirical interpretations. The chapter follows the FEDS as described in Venable et al. (2016). FEDS fits the goals and objectives of this chapter for several reasons. First, the framework describes the purpose of summative evaluation as the process of producing empirical interpretation for creating shared meanings about the evaluand. The evaluands in this chapter are still the two designed artifacts. However, the research task in this chapter is to derive empirical interpretations of value creation for actors in data exchange with the decentralized architecture. Second, both critical and supportive insights about the decentralized architecture in data exchange should be considered. In this regard, FEDS provides guidelines in the positivism or interpretivism paradigm that facilitate the structure of critical and supportive insights. Third, the framework establishes a clear evaluation process that was applied in the previous formative evaluation chapter and is also effectively followed in this chapter.

6.1. Summative Evaluation Design

In the following section, the summative evaluation design for the two artifacts is explained. For this purpose, in section 6.1.1, the summative evaluation is defined. The problems of summative evaluation like positivism or interpretivism are discussed in section 6.1.2. Lastly, the evaluation process is presented according to the steps of the FEDS framework in section 6.1.3.

6.1.1. Defining the Summative Evaluation

According to Venable et al. (2016), summative evaluation *"is used to produce empirically based interpretations that provide a basis for creating shared meanings about the evaluands"*. In this chapter, the application of decentralized architecture in data exchange is evaluated according to the two researched artifacts. Similar to the previous formative evaluation chapter, semi-structured interviews with researchers and industry experts in the subject areas of data marketplaces are used. Although the purpose of these interviews is to produce empirically based interpretations. As a result, shared meanings about the application of decentralized architecture in data exchange can be established to answer the sixth sub-research question: *What value is created for actors in data exchange with the decentralized architecture?*

6.1.2. Problems with Summative Evaluation

The summative evaluation can introduce possible errors that often originate from the qualitative analysis of a single author. These errors follow paradigms like *positivism* or *interpretivism* (Venable et al., 2016). For instance, by analyzing numerous blockchain articles, one could adopt the *positivism* approach in the arguments regarding this technology. However, a *critical* viewpoint is also important in producing empirical interpretations to create shared meanings about blockchain technology. For this purpose, as described earlier, researchers and industry experts are consulted to derive both *critical* and *supportive* insights about the application of a decentralized architecture in data exchange.

6.3. Evaluation Process

In compliance with FEDS, the evaluation process is divided into four steps: *evaluation goals*, *evaluation strategy*, *evaluation properties*, and *individual evaluation episodes*. Similar steps were used in the previous formative evaluation chapter. Although considering the functional purpose of the summative evaluation described in section 6.1.1, these steps have a different application in this chapter.

Step 1: Evaluation Goals

Two summative evaluation goals are associated with the application of blockchain technology in data exchange. The first goal of the evaluation is to *rigorously* establish the possible benefits or utilities for actors in data exchange with the decentralized architecture. The second goal is to evaluate the *uncertainty and risk reduction* in data exchange with the decentralized architecture. These goals are evaluated with researchers and industry experts, as well as aligned according to the two designed artifacts and currently available literature.

Step 2: Evaluation Strategy

According to FEDS, *"the formative and summative functional purposes of evaluations can be characterized as the ends of a continuum along which any evaluand may be located"* (Venable et al., 2016). For a more explicit understanding, this could be visualized on the x-axis in the following figure 9. Therefore, considering the described functional purpose, the summative evaluation could be viewed as an evaluation that typically follows the formative evaluation. In this regard, the *Technical Risk and Efficacy* strategy adopted in the formative evaluation is further applicable in this chapter with similar semi-structured interviews as the evaluation method. Although, in this chapter, the interviews mainly had an open discussion with pre-defined topics on decentralized architecture and data exchange to attain a more fluent evaluation process.

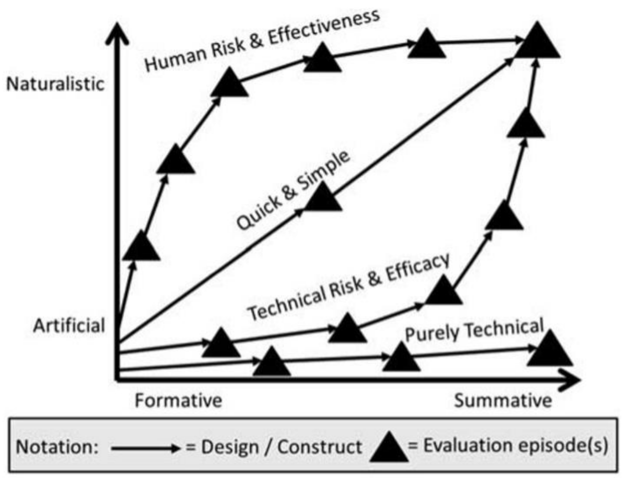

Figure 14: The Framework for Evaluation in Design Science (Venable et al., 2016)

Step 3: Evaluation Properties

The third step in FEDS represents what to evaluate. The evaluation properties for the two defined goals are presented in the following table 9. Certainly, more properties may be relevant for the evaluation. For instance, whether decentralized architecture could mitigate some of the concerns (i.e., transparency, privacy, ownership) in exchanging data; or whether decentralized architecture from an operational viewpoint is scalable and flexible. However, time in semi-structured interviews is the primary constraint. Thus, for the summative evaluation, only the following evaluation properties are considered.

To evaluate the possible benefits or utilities for actors in data exchange with the decentralized architecture	*To evaluate the uncertainty and risk reduction in data exchange with the decentralized architecture*
- Whether the decentralized architecture provides more security for actors in data exchange - Whether the decentralized architecture provides more trust for actors in data exchange	- Whether decentralized architecture provides more means of control in data exchange

Table 9: The summative evaluation properties

Step 4: Individual Evaluation Episodes

On the basis of the chosen evaluation goals, strategy, and properties, two parts characterize the complete summative evaluation. In the first part, the same six experts from the formative evaluation were interviewed (see Appendix B). These experts are actively involved in different industry projects with data management characteristics (e.g., public mobility services, decentralized data platforms). The interview process contained three distinct parts. First, the thesis research concepts were explained. Second, experts were encouraged to address questions if the scope of the research was not clear. And third, questions according to the evaluation properties were asked. It has been noted that questions about whether decentralized architecture provides trust or whether decentralized architecture provides more means of control could lead to only responses with positivism characteristics. Therefore, complementary questions with a critical perspective were also addressed. For example, whether smart contracts introduce an extra layer of complexity or whether the hype and trend of cryptocurrency cover the risks associated with decentralized architecture (blockchain). Further, in the second part of the summative evaluation, the viewpoints gathered from interviews with experts are analyzed according to the insights that could be subtracted from the two researched artifacts and the current scientific literature. Thereby, the discussion of the summative evaluation is introduced.

6.2. Summative Evaluation Results

In this section, the results from interviews with experts are presented. According to the evaluation properties from step 3, experts were questioned about: *(i) whether decentralized architecture provides more security for actors in data exchange*, *(ii) whether decentralized architecture provides more trust for actors in data exchange*, *(iii) whether decentralized architecture provides more means of control in data exchange*. In the following, both *critical* and *supportive* viewpoints of experts are presented. Every evaluation property starts with *critical* perspectives followed by *supportive* perspectives.

(i) whether decentralized architecture provides more security for actors in data exchange: experts stated that security in data exchange has long been solved as far as encryption is concerned. The real problem resolves in the systems, storage, and interfaces of data sources and in the problem of data sovereignty. When looking at decentralized architecture and what is often being done, is that it just pushes the problem to the periphery. This means that marketplaces shift the risk from themselves to peer-to-peer exchange to the providers and consumers managing data. From a risk and cost perspective – just thinking centralized storage and security – that is perfect. And from the user's perspective, this removes another point of failure from the equation, hoping that they use secure interfaces and connectors. Emerging decentralized data marketplaces are interesting in studying how they mitigate some of the concerns, although data exchange also works in the existing structures.

A different perspective commonly shared between the interviewed experts is that blockchain technology uses strong encryption and smart contracts where it is difficult to perform fraud actions. In decentralized architecture, we need to think about security and trust problems by design, analyzing the architecture itself. This could be considered a step forwards in preventing potential risks that organizations formally have hidden behind a trusted central entity. There are many data brokers and centralized data marketplaces, which are doing reasonably well. Although a lot of what centralized marketplaces have to deal with is overhead. The detailed agreements between providers and consumers take a lot of resources, time, and effort to monitor, sometimes slowing things considerably.

(ii) whether decentralized architecture provides more trust for actors in data exchange: according to the interviewed experts, the problem of data markets is not one of lacking trust fueled by incidents along the lines. Centralization has not led to a lack of trust. Data marketplaces still need central components like identity management, brokerage, execution, facilitation. It seems like blockchain technology platforms are trying to solve a problem that either does not need to have a solution, or these platforms are trying to find a solution to applying decentralization – which indeed has benefits in many domains – and novel technology solutions while we still do not have an emerging understanding of what it could be. Moreover, regarding smart contracts, it gets interesting that even with legal contracts, they get complex and normal people do not understand them intuitively. "*So now you get this half-understandable agreement to a smart contract, where agreements are coded and hard to audit? Does this create an additional layer of trust between the business and users on both ends? I don't even know what is going on there; I have to blindly trust more parties involved*". In the context of data marketplaces, smart contracts are de facto just being used to auto-execute the most simple contracts. They are great conceptually and possibly for other use-cases in the blockchain community, but it doesn't have to be the complexity of the context behind them.

The second perspective regarding the trust for actors in data exchange with a decentralized architecture is that sometimes there are different levels of decentralization that are useful. In the case of data sovereignty and data monetization, blockchain offers an interesting solution. At least from the ideological sense, blockchain places access control, rewards, and trust into individuals. In the centralized architecture, you have to trust a single entity to secure your data custody. Furthermore, this single point of trust should act in your interest so that it would restrict any transaction without your consent or exclude any fraud actions. Moreover, regarding the complexity of smart contracts, it depends on how it is implemented. More specifically, when smart contracts are within an ecosystem of blockchain technologies that facilitate the understanding for users and they may provide means to protect them if they are implemented correctly. Although if you have a smart contract without any other complementary technologies, this is problematic.

(iii) whether decentralized architecture provides more means of control in data exchange: The first perspective is that what is believed to be achieved with decentralization often is a projection of human nature because we all like to feel the illusion of being in control. Decentralization still requires deeper reflection and a very clear weighing of all the pros and cons. Organizations want to keep relatively tight control. This requires tight contracts with trusted parties. So, more often than not, we see one-to-one use cases of data exchange that emerge by using a search functionality on a platform.

Another perspective is that blockchain technology is transparent and traceable. Moreover, it is controllable in terms of *"your keys, your bitcoin; your keys your data"*. Therefore, an individual or organization has the necessary control and ability to determine how data is being used or been used. So, if organizations have offered data for usage in a blockchain environment, they can trace how it has been used, when, and by whom, giving more control to organizations in terms of sharing data. In this context, you do not have to deal with any uncertainty. Thus, you know your data is encrypted, you know what keys have been used to encrypt your data, and you do not put this responsibility in the hands of other entities. In this regard, this might increase the acceptance of data marketplaces if companies that are commercializing their data understand what they are using. However, another concern for decentralized architecture is the complexity of such systems. Transparency may be more understandable for someone who studies these species of marketplaces.

6.3. Summative Evaluation Discussion

In the following, the viewpoints deducted from interviews with experts from the previous section 6.2 are analyzed according to the current scientific literature and in compliance with the designed artifacts.

(i) whether decentralized architecture provides more security for actors in data exchange: the security problem can be investigated from two perspectives, mainly related to where data leakage could occur? Or, who may be targeted in the cyber-attacks? Starting with the perspective of data marketplaces with blockchain technology being targeted as an organization, this would be widely challenging, at least for two reasons. The first reason is that blockchain heavily relies on cryptographic mechanisms and consensus protocols (see sections 3.2.2 and 3.2.3). For new blocks to be added to the blockchain, the network participants compete to calculate a targeted condition. The first participant reaching the targeted condition would broadcast the new block so that other network participants in the blockchain verify the validity of calculations; if network consensus is reached, the new block is added to the distributed ledger (Zheng et al., 2017). Why are these blocks important? In the blockchain, the information about the transactions is stored in blocks connected to each other (see section 3.2.1). Therefore for cyber-attacks, these would be the forging targets; modifying the contents of blocks, introducing new blocks, or removing blocks from the chain. This would imply that attackers have to compete with the honest participants. Technically, there are a large number of nodes and computing power in the blockchain network, so the attacker needs more computation power than the entire network (Yue et al., 2017). Therefore, this aspect significantly reduces the capacity of the attackers. The second reason for the challenging aspect of targeting data marketplaces as organizations is the combination of blockchain technology, TEE, and smart contracts. The TEE represents a federated network of known nodes operating under a restricted threat model (see section 3.3.7). Technically, the network of known nodes must have installed security-related assets, codes, and underlying trusted operating systems to form the secure area for processing transactions (Hynes et al., 2018). Additionally, smart contracts on blockchain enable automatic payments and rely solely on the security properties of the decentralized architecture; there is no need for a trusted third party to process payments in marketplaces built on such a decentralized architecture (Koutsos et al., 2020). As a result, this also describes the viewpoints of experts that *"in decentralized architecture, we need to think about security and trust problems by design, analyzing the architecture itself. This could be considered a step forwards in preventing potential risks that organizations formally have hidden behind a trusted central entity"*.

Another perspective related to where data leakage could occur is that actors participating in data exchange may be targeted in cyber-attacks. When using blockchain, the private keys of actors are regarded as their identification and security credentials. Actors manage their identity instead of relying on third-party organizations. Therefore, once malicious actors gain access to the user's private keys, it will be difficult to revoke the access rights in a distributed system (Li et al., 2020). This introduces a higher risk to the actors themselves as being targeted in attacks. As experts stated, *"your keys, your bitcoin; your keys your data"*. Thus blockchain technology may introduce more control in data governance, although it also comes with more responsibility for actors to secure the integrity of their data assets, approaching the argument that *"when looking at decentralized architecture and what is often being done, is that it just pushes the problem to the periphery"*. This statement is also on the pulse of something true about the responsibility of actors to secure their data. Thus returning to *whether decentralized architecture provides more security for actors in data exchange*, from an organizational viewpoint, data marketplaces using blockchain technology could enhance the security aspect through their cryptographic mechanism, consensus protocols, and architecture properties as TEE. Cyber-attacks on these organizations are more challenging to be accomplished, so removing the single point of failure. However, from the perspective of actors, there is significant accountability regarding their identity; once lost, it appears to be difficult to return their security credentials.

(ii) *whether decentralized architecture provides more trust for actors in data exchange:* trust is a complex evaluation property in data exchange; it requires multiple aspects to be addressed. Scholars described that actors could be hesitant to exchange data due to a lack of trust between trading parties (Schomm et al., 2013; Koutroumpis et al., 2017; Fruhwirth et al., 2020). Therefore, both centralized or decentralized data marketplaces require trust management tools to monitor the data providers and consumers (Koutroumpis et al., 2020). On the one hand, monitoring data providers can take the form of tracing previous transactions and establishing reputations systems. On the other hand, monitoring data consumers can represent the mechanisms based on actor complaints with data usage agreements. Data providers or consumers found guilty of violating the contractual agreements could be penalized by decreasing their trust value, making a consumer or provider less trustworthy to be engaged in data exchange with other parties (Noorian et al., 2014). As stated by experts in the centralized architecture, *"the single point of trust should act in your interest so that it would restrict any transaction without your consent or exclude any fraud actions"*. The primary difference from a central structure is that the information about the trustworthiness of participants in the decentralized architecture may now be transparently registered and not executed by a central structure (Koutroumpis et al., 2017). Moreover, due to the immutability properties of blockchain, the complete history of all transactions and originated addresses could be subtracted from the distributed ledger. This introduces opportunities for effective audit services without trusting a central entity.

According to experts, *"centralization has not led to a lack of trust"*, and *"data marketplaces still need central components like identity management, brokerage, execution, or facilitation"*. In the HLA, it was mentioned that the integration of different technologies is an imperative practice in the context of distributed technologies (see section 3.4). Thus, it is correct that decentralized architecture shares many primary attributes with central components. However, the most important in the decentralized architecture is that these central components may be directly verifiable by marketplace participants. For instance, a third-party data brokerage may use the decentralized architecture to provide aggregated data to different consumers. However, assuming data quality is not as anticipated, consumers could evaluate the data brokerage as a central component with low data quality. Moreover, as earlier described, this evaluation can not be technically changed or influenced due to the immutability and transparency characteristics of blockchain technology. This makes the quality or trustworthiness assessment transparent and immutable.

In the same trust context, smart contracts were also discussed with experts. Therefore, starting with the argument that *"smart contracts are de facto just being used to auto-execute the most simple contracts"*, from a practical viewpoint, this was observed in the taxonomy of BMs for decentralized data marketplaces. So, even though smart contracts in a couple of analyzed data marketplaces have configurable properties, these properties are limited in application, for example, only setting the price of data assets per day or adjusting the minimum number of constraints in using data (see section 4.4). Although from a theoretical viewpoint, Hynes et al., (2018) demonstrated an application of configurable smart contracts where both data providers and consumers have to deploy their smart contracts into the blockchain. More specifically, *(1) a data provider can deploy a smart contract (C_d) with their data constraints in blockchain; (2) a data consumer desiring to use the provided data writes a similar smart contract (C_c) that satisfy the same constraints; (3) C_d automatically verifies that C_c satisfies the constraints, and data can be securely exchanged*. Technically, this could represent a feasible solution ensuring precision and certainty. However, this would require both data providers and consumers to have technical capabilities, which corresponds to the argument that smart contracts are *"great conceptually and possibly for other use-cases in the blockchain community, but it doesn't have to be the complexity of the context behind them"*. The ecosystem of blockchain technologies could facilitate the understanding of smart contracts for regular actors. This was observed in the taxonomy of BMs with the data mediating expertise and open-source governance. More often than not, decentralized data marketplaces are trying to establish an ecosystem of data service providers with the technical capabilities to use smart contracts or blockchain technologies. Thus this ecosystem could offer complementary technologies or necessary knowledge in facilitating the operation of smart contracts. However, smart contracts are simply programs stored on the blockchain. Therefore, it is essential to remember that smart contracts can guarantee perfect performance only if the code is perfect: it correctly reflects the business agreement between data providers and consumers, has no code errors, and has no security breaches. Needless to say, such assumptions cannot be made (Mik, 2017).

(iii) whether decentralized architecture provides more means of control in data exchange: in the role model of traditional data marketplace ecosystems (see section 2.5), the marketplace owners collect and host data from data providers and consequently sell the data in the form of queries to data buyers (Spiekermann, 2019). Thus, the marketplace owners are directly in control of the data flow. A different approach is considered in the decentralized architecture where data providers or consumers directly control data storage, custody, and manipulation (see section 3.3.2). According to the requirements specification of the HLA, data providers have the flexibility to decide on the *data storage* methods. In this regard, data providers can store data locally on their devices, use web cloud storage (e.g., AWS, Google Cloud, etc.), decentralized storage (e.g., filecoin), or public data oracles (e.g., Chainlink) (see section 3.4). Furthermore, in maintaining sovereignty over personal digital assets, blockchain architecture enables data tokenization. In data security, tokenization represents replacing sensitive data with non-sensitive equivalents that reference the original data (see section 3.3.2). More specifically, a token can be used as a utility providing digital access to the data or service based on blockchain (Kim & Chung, 2019). This meets the arguments of experts that *"you know your data is encrypted, you know what keys have been used to encrypt your data, and you do not put this responsibility in the hands of other entities. In this regard, this might increase the acceptance of data marketplaces if companies that are commercializing their data understand what they are using"*. However, as experts stated, *"another concern for decentralized architecture is the complexity of such systems"*. For data marketplaces with blockchain architecture to succeed, it takes an ecosystem to be established, and they are very hard and take so long to build. Providing data exchange functionality is so much easier as an add-on to an existing ecosystem. Therefore, as observed in scientific literature and as stated by experts, *"more often than not, we see one-to-one use cases of data exchange that emerge by using a search functionality on a platform"* (Stahl et al., 2015).

6.4. Conclusion of Chapter 6

Similar to the previous formative evaluation chapter, the FEDS framework of Venable et al. (2016) was subsequently applied in this chapter; although, with a different functional purpose of evaluation. More specifically, the functional purpose of this chapter was to derive empirical interpretations for creating shared meanings about the refined artifacts. For this purpose, two evaluation goals were defined as follows. First, to *rigorously* establish the possible benefits or utilities for actors in data exchange with the decentralized architecture. And second, to evaluate the *uncertainty and risk reduction* in data exchange with the decentralized architecture.

Based on the established goals three evaluation properties were conceptualized: *(i) whether decentralized architecture provides more security for actors in data exchange*, *(ii) whether decentralized architecture provides more trust for actors in data exchange*, and *(iii) whether decentralized architecture provides more means of control in data exchange*. Certainly, more evaluation properties may be relevant. However, considering that in the evaluation process semi-structured interviews were conducted, time was the primary constraint in addressing more questions.

These evaluation properties were consulted with researchers and industry experts in the subject areas of data marketplaces to gather their *critical* viewpoints as well as their *supportive* perspectives. Finally, the results from interviews with experts were analyzed according to the current scientific literature and in compliance with the designed artifacts. Concluding the summative evaluation and considering the earlier describe properties the following could be mentioned. The decentralized architecture provides a platform for securely exchanging data, through cryptographic mechanisms, consensus protocols, and architecture properties as TEE. However, with no intermediaries in the exchange process, actors are directly accountable for their identity; once lost, it appears to be difficult to return their security credentials. Further, the decentralized architecture provides immutable records of prior transaction activities that could be applied in gathering intelligence about the trustworthiness of actors through monitoring tools that are not executed by a central structure. These could enhance the trust in the transparent assessment itself. However, solely relying on immutability and transparency without complementary technology provided by knowledgeable actors, the decentralized architecture could offer more uncertainties than trust. Technically, a decentralized architecture could offer control to actors through storage, custody, and management enabled by tokenization. However, to have utility from this control, a decentralized architecture requires an established ecosystem to accomplish data exchange without intermediaries as marketplace owners.

7
Answering the Sub-Research Questions

Data Marketplaces are emerging platforms in the data industry that offer data exchange services as their business activity. These marketplaces are intermediary platforms that connect data providers, consumers, and other complementary technology providers. Studies on data marketplaces as digital platforms have already been conducted. From the technical perspective, Stahl et al. (2015), Koutroumpis et al. (2017), and Spiekermann M. (2019) classified the data markets and described the centralized and decentralized data-trading frameworks. From the business perspective, Bergman (2020), M. van de Ven et al. (2021), and Fruhwirth et al. (2020) researched the business models for data marketplaces. In most of these examples, data marketplaces are studied as a general phenomenon. Although, within data management research, there is considerable interest in decentralized technologies. Moreover, decentralized data marketplaces are emerging on the market (i.e., Ocean Protocol, Streamr, etc.). Therefore, in this thesis, a different approach with a specific focus on decentralization, namely blockchain technology, was considered. Furthermore, the research followed the firm belief that both the technical and business specifications are essential in understanding the data marketplaces with blockchain technology. For this purpose, the thesis was designed according to DSR methodology to develop two research artifacts. As already noted, the two artifacts were the HLA and the Taxonomy of BMs for decentralized data marketplaces. In the following, the findings from the designed artifacts are analyzed according to each sub-research question divided between technical background, technical specifications, business specifications, formative and summative evaluations.

7.1. Technical Background

Before researching the decentralized architecture and business models, data marketplaces had to be explicitly understood. Moreover, for the design of the HLA, the infrastructure components had to be studied according to current scientific literature. Although, with the advancement of data management and related technologies, data marketplaces may encounter technical and business changes. Therefore, the research in the thesis considered the generic infrastructure components according to the primary functionalities of data trading platforms. For this purpose, the first sub-research question was:

What are the generic infrastructure components for data marketplaces?

Technically, data marketplaces enable *data exchange* by providing the infrastructure for finding, buying, and selling datasets. In principle, data exchange describes the general functionality of data marketplaces, although it does not include the regulatory aspect for *data storage*, *custody*, or *manipulation*. In this context, the custody and manipulation of data could be regulated through different licensing models. Theoretically, every data set should fall under a specific license describing the terms and conditions that explicitly regulate the use of data.

In the commercialization process, data can be viewed as intangible products that *need to be further processed and analyzed* with complementary technologies for consumers to gain utility. However, the commercialization of intangible products with nonexisting property rights like data enables challenges in assessing the *quality* and *pricing mechanisms*. For *pricing mechanisms*, data providers should cautiously determine the pricing models that reach the most consumers. Subsequently, for *data quality*, metadata is an essential reference point both for providers and consumers. Through *metadata*, providers can describe the supplied datasets, and consumers can trace these specifications; thereby enabling the *query & search* functionalities for data trading platforms. Lastly, marketplaces should ensure that actors have appropriate access to their data resources. This could be accomplished through *identity management* mechanisms for identification, authentication, and authorization processes. Therefore, considering the primary described functionalities the following generic infrastructure components were defined for data marketplaces: *data exchange*; *data storage, custody, and manipulation*; *data process and analytics*; *data quality assessment*; *pricing mechanism*; *query & search*; and *identity management*.

7.2. Technical Specifications

With the established infrastructure components, the HLA, which represented the first research artifact, was designed to explore the technical specifications of data marketplaces with blockchain technology. Thus, the second sub-research question associated with this purpose was:

How does a decentralized architecture for data marketplaces look like?

According to the HLA presented in section 3.4, in the following, several findings could be discussed regarding the technical specifications of data marketplaces with blockchain architecture. Starting with the technology itself, blockchain can be regarded as a distributed ledger on which transactions between different actors are recorded in a chain of information blocks. The distributed ledger is transparently revealed to the network, and participants are actively validating its contents. Therefore, blockchain excludes by design the need for intermediaries to host or manage transactions. In this context, the role in the data ecosystem of marketplaces with blockchain technology changes from data intermediaries to communication structures that facilitate the operation of the decentralized market. The shift from the intermediary role technically influences the functional requirements of the infrastructure components earlier described. From a high-level perspective, it can be noted that actors in the decentralized architecture could be characterized as independent entities directly responsible for their data assets. Therefore, the core activities of decentralized marketplaces as business entities change from collecting, hosting, and selling data to offering operation tools in the data market. Moreover, since actors are directly responsible for their data assets, the primary focus of a marketplace using decentralized architecture may shift more towards establishing the operation base for development and maintenance; increasing the overall platform and tools quality; and ultimately growing the user base to develop a functional ecosystem for data exchange.

7.3. Business Specifications

Decentralized data marketplaces also require sustaining comprehensive business models to maintain and create value for actors engaging in data exchange. Therefore, the second researched artifact in the thesis was the taxonomy of BMs for data marketplaces with a decentralized architecture. The obtained taxonomy aligned the business model characteristics of five market-active decentralized data marketplaces and provided the insights necessary to answer the third sub-research question:

What are the characteristics elements in a taxonomy of business models for data marketplaces with a decentralized architecture?

With the established taxonomy of BMs from section 4.5, a number of observations could be discussed to describe the business specification of decentralized data marketplaces. In most of the observed business model characteristics, decentralized data marketplaces share the same properties as centralized marketplaces. However, in the context of decentralization, data marketplaces also use blockchain technology to sustain the data exchange infrastructure, automate the relationship with actors, and explore new sources of revenue. For sustaining the data exchange infrastructure, node operators are incentivized to provide computational power or storage space in return for rewards calculated as transaction fees. On the one hand, this describes the decentralization of transactions storage. On the other hand, data marketplaces may also reduce the resources spent on maintaining a centralized repository of data transactions. Further, self-executable smart contracts are commonly offered in decentralized markets for automated relationships with actors. This enables peer-to-peer transactions and excludes the need for tight monitoring of contractual compliance. Lastly, considering that cryptocurrency is attracting worldwide attention, this digital currency may also substitute the initial investment costs and raise funding. More specifically, strategies as selling, investing, staking, incentivization, or burning were observed in the analyzed marketplaces that can provide a revenue model utilizing the crypto-economic trend.

7.4. Formative Evaluation

With the established HLA and the taxonomy of BMs for decentralized data marketplaces, these artifacts were evaluated according to the FEDS framework (Venable et al., 2016). For this purpose, six experts in data management with affiliation to data trading platforms were interviewed. The HLA was evaluated on *comprehensibility, logical structure, and completeness*. Subsequently, the taxonomy was evaluated on *completeness* and *detailedness*. As a result, these evaluation properties aimed at answering the fourth and fifth sub-research questions:

Is the decentralized architecture for data marketplaces
evaluated as comprehensible, logically structured, and complete?

Is the taxonomy of business models for data marketplaces
with a decentralized architecture evaluated as detailed and complete?

Several results from the evaluation activities with experts can be discussed. In the following, these results are structured in two parts: first, the evaluation results from the decentralized architecture, and second, the evaluation results from the taxonomy of BMs for decentralized data marketplaces.

The first evaluation activity in the interviews was related to the HLA. According to evaluation properties, the HLA was evaluated on *comprehensibility*, *completeness*, and *logical structure*. For *comprehensibility,* experts stated that most of the infrastructure components are intuitive. Thus, this was perceived as an indicator that the naming, definition, and functional requirements are insightful for the HLA. For completeness, experts had remarks regarding the missing third parties and the static representation of the architecture. These remarks were given at the early phase of the research; thus, they were already implemented at the moment of writing the evaluation chapter. Lastly, the most significant remark that required restructuring the HLA was the broken architectural layers for logical structure. In this context, the layers were separated, and the HLA was redesigned in section 5.2.1.

For the evaluation of the taxonomy of BMs for decentralized data marketplaces, experts could not identify inconsistencies regarding the *detailedness* or *completeness*. Only two recommendations were given. First, to have more than one business dimension, and second, avoid labeling the business components and dimensions with the same naming. These recommendations were given at the early phase of researching the taxonomy. Therefore they were already included in the taxonomy at the moment of writing the evaluation chapter. Lastly, considering that they were few recommendations for improvements, the taxonomy was also evaluated according to the objective conditions based on the author's judgment. The objective conditions were subtracted from Nickerson et al., (2013). These objective conditions were tested at the end of each phase of the taxonomy development to assist the taxonomy researcher in determining when the process may stop characterizing that the taxonomy is detailed and complete.

The recommendations for improving the researched artifacts marked the final phase of artifacts development with their evaluation. Moreover, the evaluation of the research artifacts introduced new insights to be approached in the following summative evaluation.

7.5. Summative Evaluation

Similar to previous sub-research questions, the FEDS framework of Venable et al. (2016) was further used with two additional evaluation goals. First, to rigorously establish the possible benefits or utilities for actors in data exchange with the decentralized architecture. And second, to evaluate the uncertainty and risk reduction in data exchange with the decentralized architecture. These evaluation goals were consulted with experts in semi-structured interviews to derive empirical interpretations and answer the sixth sub-research question:

What value is created for actors in data exchange with the decentralized architecture?

In the interviews with experts, three evaluation properties were debated: first, *whether the decentralized architecture provides more security for actors in data exchange*, second, *whether the decentralized architecture provides more trust for actors in data exchange*, and third, *whether decentralized architecture provides more means of control in data exchange*. Therefore, based on the chosen evaluation properties and according to the results subtracted from interviews, the following empirical interpretations can be discussed to answer the sixth sub-research question.

The decentralized architecture heavily relies on cryptographic mechanisms and consensus protocols. Therefore, the properties of the decentralized architecture provide a platform for securely exchanging data. However, when using blockchain, the private keys of actors are regarded as their identification and security credentials. Actors manage these private keys themselves instead of relying on third-party organizations. Once lost, actors will not be able to recover their credentials. Therefore, regarding whether the decentralized architecture provides more security for actors in data exchange, it provides a platform for securely exchanging data; however, actors are directly responsible for their identity and credentials integrity.

The decentralized architecture offers an immutable record of prior transaction activities that could be applied in gathering intelligence about the trustworthiness of actors through monitoring tools that are not executed by a central structure. These could enhance the trust in the transparent assessment itself. However, solely relying on immutability and transparency without complementary technology provided by knowledgeable actors, the decentralized architecture could offer more uncertainties than trust. Therefore, regarding whether the decentralized architecture provides more trust for actors in data exchange, it provides the transparent infrastructure on which effective auditing mechanisms could be implemented. Although, these mechanisms require technical tools for monitoring the activities of actors in the decentralized market.

Regarding whether decentralized architecture provides more means of control in data exchange, technically, a decentralized architecture could offer more control to actors through storage, custody, and manipulation of datasets enabled by tokenization. Thus, it maintains the authority of actors to decide on the method of storage, secure sharing, and data manipulation. Although, for data marketplaces with blockchain architecture to succeed, it takes an ecosystem to be established. The control in data exchange is much easier as an add-on functionality to an existing ecosystem.

8 Conclusion

Digitalization could be considered the central innovation process for every economic area, whether it is the healthcare sector, agriculture, or automotive industry. Nowadays, digital technologies play a prominent role in shaping the traditional economy. Moreover, with data being created at various industrial levels, the center of transformation in the emerging digital economy relies on data and data-driven innovations. Therefore, it comes as no surprise that data is regarded as the new gold for organizations to attain a competitive advantage. In particular, data harvesting could be employed to release better business services, innovate in science, and address different global issues like climate change or pandemics. The need for data harvesting led to the inception of specialized data marketplaces where data is exchanged as a commodity. Conceptually, data marketplaces are intermediary platforms that enable the infrastructure for data exchange as their core activity, offering the opportunity for data providers to commercialize their datasets and save time for data consumers in identifying appropriate datasets. Although exchanging data through intermediaries may introduce multiple technical, legal, and ethical concerns. Different processes should be investigated to anticipate the healthy process of data exchange. For this purpose, numerous studies are focusing on leveraging the properties of decentralization, namely blockchain technology, smart contracts, and cryptographic techniques, on addressing the challenges and requirements around security, privacy, transparency, and ownership in data exchange. Thus, in understanding the potential benefits of decentralization in the subject areas of data marketplaces, this thesis researched the technical and business specifications of blockchain technology in the context of data exchange with the overarching research question: *What is the impact of a decentralized architecture on the business models of data marketplaces?*

From a technical perspective, data marketplaces with blockchain architecture could be viewed as communication structures that facilitate the operation of the decentralized market. In the blockchain environment, data exchange is accomplished through a distributed ledger on which transactions between different actors are recorded in a chain of information blocks. Subsequently, the distributed ledger is transparently revealed to the network, and participants actively validate its contents. Therefore, blockchain excludes by design the need for intermediaries as data marketplaces to host or manage data transactions. In this context, providers and consumers may perceive control and maintain ownership over the exchange process. However, more control and ownership comes with more responsibility in ensuring the integrity of the data assets. In the case of a security breach, once malicious actors gain access to the provider's or consumer's identity, it will be difficult to revoke the access rights in a distributed system where there are no central managing authorities. Thus, more often than not, actors participating in data exchange may also require technical capabilities to securely benefit from blockchain technology or utilize the properties of smart contracts. Well-established ecosystems of data services may facilitate the understanding of blockchain or smart contracts for regular actors by offering complementary technologies or necessary knowledge. Therefore, returning to the research question, from a technical perspective, several characteristics could be noted regarding the impact of a decentralized architecture on the business models of data marketplaces. First, a primary role of data marketplaces in the decentralized environment is offering technical tools for data providers and consumers to operate themselves in the decentralized market. Second, since actors are directly responsible for their data assets, the primary focus of decentralized data marketplaces may also consider establishing the operation base for the development and maintenance to increase the overall platform and tools quality. Lastly, growing the user base to develop a functional ecosystem could be imperative in the inception phase of blockchain application in data exchange.

From a business perspective, marketplaces with a decentralized architecture share the same properties as centralized marketplaces. Although, in the context of decentralization, marketplaces also leverage several blockchain-related characteristics, which could be described as follows. In the business model characteristics of the decentralized marketplaces, a new category of marketplace participants could be observed. These participants were defined as node operators responsible for providing computational power or storage space in return for cryptocurrency rewards calculated as transaction fees. The cryptocurrency rewards determine the incentivization design of human behavior that lies at the basis of the decentralization mechanisms. More specifically, rewards given to note operators for computational power or storage space represent a substitute for a central structure that would maintain, manage, and expand the system. Further, the relations between marketplace participants are ensured through programmable user-defined conditions in smart contracts. This enables peer-to-peer transactions avoiding the potential extra effort in managing relationships between providers and consumers. Technically, smart contracts could represent a feasible solution ensuring precision and certainty in data exchange. Although, it is also essential to remember that smart contracts are simply programs stored on the blockchain which could guarantee perfect performance only if the code is perfect. Needless to say, such assumptions cannot be made. Next, a distinct resource that has been attracting significant attention over the last period is cryptocurrency. Most of the time, data marketplaces with blockchain architecture enable their participants to use these resources as payment or utility tokens. Furthermore, this digital currency may also substitute the initial investment or maintain the development process. For instance, it was observed in the analysis that decentralized data marketplaces are founding through their own cryptocurrency tokens various development programs based on open-source principles. Furthermore, besides being evaluated as resources, cryptocurrency tokens offer an alternative revenue stream for blockchain projects. The revenue stream originates from external investors acquiring cryptocurrency through specialized platforms, thus providing funds to marketplaces. Therefore, returning to the research question, from a business perspective, several characteristics could be noted regarding the impact of a decentralized architecture on the business models of data marketplaces. First, data marketplaces with blockchain architecture use reward mechanisms through cryptocurrency tokens to program human behavior through incentive design, thereby sustaining the infrastructure for data exchange. Second, self-executable smart contracts are commonly offered in decentralized markets to automate the relations between data providers and consumers. This enables peer-to-peer transactions and excludes the need for tight monitoring of contractual compliance. Third, decentralized data marketplaces leverage their cryptocurrency tokens as resources to substitute the initial investment costs or maintain the development process. Moreover, selling cryptocurrency tokens on specialized platforms may represent an additional revenue stream which is boosted by the significant attention over cryptocurrency in the last period.

With the main question answered from both technical and business perspectives, this study may conclude with the following closing comment. The decentralized data marketplaces, as well as the application of blockchain architecture, are still in the emerging phase. It is on the pulse of something genuine that decentralization may introduce numerous benefits on access to, sharing of, and use of data. However, a critical viewpoint should always be present to enable a competitive, secure, inclusive, and ethical digital economy. Solely relying on blockchain properties without complementary technology or knowledgeable actors, the decentralized architecture could offer more uncertainties than trust.

9
Reflection

In this section, a reflection on the research process and study results is discussed. For this purpose, the research relevance in the technology management domain of data marketplaces and decentralized technologies is presented in section 9.1. Further, the scientific contribution and the practical contribution of this master thesis according to the designed artifacts are introduced in sections 9.2 and 9.3. Next, in section 9.4, the research limitations associated with the research process and results are addressed. Lastly, the recommendations for future research in the subject area of decentralized technologies and data trading platforms are introduced through several questions in section 9.5.

9.1. Research Relevance

With the growth of the digital economy and the increase of value perceived from data, the EU announced several strategies and governmental initiatives to establish the framework in ensuring attractive policies for businesses, researchers, and public administrations to cooperate in data sharing (see section 1.1.1). These initiatives aim at creating a European Data Market where both industrial data and personal data are securely regulated and used. In this context, the concept of data marketplaces has been widely investigated in the recent period. Multisided platforms that offer reliable mechanisms for an efficient match between data supply and demand are the primary researched characteristics.

Data marketplaces are emerging platforms that are still gaining momentum. Thus, numerous aspects lack theoretical and practical background. Platforms that provide data exchange could be considered early business adopters in the data market. The business models for these platforms are in the exploration phase. Additionally, the architecture for data marketplaces is an important research aspect. A plausible architecture for data management could shape the business models of marketplaces and create the necessary value for actors to engage in data exchange. In this context, the cross-level analysis regarding the architecture and business models for data marketplaces highlights the specific relevance of research in the technology management domain.

Since the inception of Bitcoin, blockchain technology has grown beyond the mechanisms of crypto coin transactions. Moreover, with the application of user-defined states through smart contracts, blockchain gained research popularity in industries such as finance, the Internet of Things (IoT), or supply chain management. In the more recent period, numerous studies are also researching the properties of blockchain technology, smart contracts, and cryptographic techniques, aiming at addressing the challenges and requirements around security, privacy, transparency, and ownership in data exchange. Although, most of these studies have a domain-specific analysis for the data marketplaces (e.g., IoT, smart cities). In this regard, this master thesis was engaged with a more generic approach, viewing the data marketplaces with blockchain architecture from both technical and business perspectives.

As described in section 1.1.2, the current thesis research was carried out as part of the TRUSTS project. Therefore, the research results could be applied in the several planned deliverables known as follows. The requirements analysis of architecture design and technical specifications. The platform implementation for TRUSTS infrastructure and smart contracts. The exploitation & innovation for the business model for TRUSTS data marketplace.

9.2. Scientific Contribution

On the basis of the two designed artifacts, in the following, the HLA and the Taxonomy of BMs are discussed to describe their theoretical contributions. Starting with the first researched artifact, the HLA provides a generic framework, which could help further research the application of blockchain architecture in data trading platforms. Moreover, the analyzed infrastructure components may offer the fundamental knowledge required in studying the technical specification of decentralized marketplaces as well as the domain-specific application of blockchain technology in different use cases for data exchange. Data marketplaces with blockchain architecture are a relatively new research topic. Most of the studies have a single industry analysis (e.g., IoT, smart cities). From a different perspective, this thesis considered a more generic approach. Thus, the HLA could be generalized in the subject area of various scientific research where the blockchain architecture and data marketplaces are investigated.

For the second designed artifact, the Taxonomy of BMs contributes to the inclusive understanding of the business models for data marketplaces, which distinguishes from previous studies with the focused approach to cover the decentralization aspect. Additionally, the taxonomy provides an overview of the business model characteristics and introduces several new dimensions that either had not been researched yet or either had just a basic research application. For instance, the taxonomy integrated the emerging business model dimensions such as *smart contracts* in customer relationship management, *data mediating expertise* with third-party providers, the *blockchain infrastructure* for a decentralized architecture, *open-source governance* as a different key activity, or *cryptocurrency governance* as a new revenue stream utilizing crypto-economic systems. Especially in the context of decentralization, these business dimensions make the classification of data marketplaces with blockchain architecture more reliable and exhaustive. For example, the dimension of *cryptocurrency governance* plays a significant role in the business activities of data marketplaces. Through their cryptocurrency resources, marketplace owners may sustain the infrastructure for data exchange, substitute the initial investment costs or maintain the development process. However, in previous studies on the Taxonomy of BMs for data marketplaces, cryptocurrency is only considered a payment resource. In this sense, the elementary classification as payment resources may introduce confusion regarding the actual operation of the decentralized data platforms. More specifically, what are the sources of revenue for data marketplaces that only provide tools to operate in the decentralized market? How do they sustain their business models if cryptocurrency is only evaluated as a payment resource? How to liquidate these resources? The answers to these questions could be discovered in the new business dimension of *cryptocurrency governance* researched in this thesis. The cryptocurrency governance represented an example to indicate the multiple aspects that should be considered for a single business dimension and its importance. Although, as earlier described the Taxonomy of BMs also integrates the new dimensions of *smart contracts*, *data mediating expertise*, *blockchain infrastructure*, or *open-source governance*.

9.3. Practical Contribution

Similar to section 9.2, in the following, the practical contribution of the two artifacts is discussed. Starting with the first researched artifact, the HLA offers a structural framework for the emerging decentralized data marketplaces to define the architectural layers and their infrastructure components. In this context, a potential development team may design easily understandable components without much thought or explanation. Moreover, the proposed layered architecture could be extendable, adding infrastructure components or modifying the existing ones. Thus, writing new features would not require much effort to grasp how everything is tied together from a high-level perspective.

For the second designed artifact, the practical contribution of the Taxonomy of BMs may function as a map for different business dimensions to analyze the position of data trading platforms in the market as commercial organizations. Therefore, this mapping tool could represent an instrument for marketplace owners to establish a comprehensive business overview for their platform, supporting their business decisions. Furthermore, for data marketplaces in the inception phase, the taxonomy may also be applicable in defining the development strategies to comply with the taxonomy business dimensions.

For instance, on a more generic note, it has been noted that few decentralized marketplaces could be identified within the EU. Thus, it seems that these decentralized marketplaces are still emerging in the market. In this sense, marketplace owners may use the HLA to examine the possibility of establishing the operation base for increasing the overall platform and tools quality. Subsequently, they could employ the Taxonomy of BMs to analyze their market positions with respect to their competitors.

9.4. Research Limitations

The research process and results in this study could be subjected to several limitations. In the following, these limitations are discussed to clearly acknowledge the awareness of their presence and how they may affect the conclusions that could be drawn from the research.

A common limitation for a master thesis concerned with qualitative data is the interpretations of a single researcher. Therefore, there is a chance that essential knowledge has been overlooked or the acquired information was misinterpreted. Moreover, the described results could also be subjected to the researcher's understanding and views about the subject area. In minimizing the effect of this limitation, interviews were conducted with researchers and industry experts to reduce possible errors, clarify misunderstandings, and strengthen the generalized interference of the research.

Regarding the evaluation process performed with researchers and industry experts, an evident limitation is the time constraints and the interpretative technics of artificial evaluation. The individual interviews were scheduled for a time limit of one hour, with different questions being addressed. It has been noted that, occasionally, more time was required to clarify the thesis concepts or expand on the recommendations for improving the research. Therefore, to reduce the impact of not clear research objectives, interviewees were provided with the background information about the thesis, a structured agenda, and interview questions three days in advance. Another limitation in the same evaluation context is the use of artificial evaluation only. According to section 5.1.2, the following step of artificial evaluation would be the naturalistic evaluation to explore the performance of a technical solution in its real environment. However, a comprehensive naturalistic evaluation requires more time than the predetermined thesis time frame. Thus, the real environment application represents a limitation. In seeking to reduce the impact of this limitation, the documents of active data marketplaces were also used in the thesis analysis.

For the designed artifacts, several limitations could be addressed. Therefore, starting with the HLA, only the generic infrastructure components were explored. Clearly, the architecture could be extended with more in-depth analysis and infrastructure components. Although, as earlier stated, the thesis has a specific time frame. Thus, the scope of the study was limited, balancing between delivering a comprehensive understanding of decentralized data marketplaces and available resources, namely time. Furthermore, according to section 3.1, the research on decentralized architecture was guided by specific scientific articles. Thus, the designed HLA could have been influenced by existing blockchain technology applications or architectural concepts. To reduce this limitation, besides the interviews with researchers and industry experts, the HLA was compared to the technical whitepapers of active data marketplaces. Further, regarding the Taxonomy of BMs, according to section 4.1, an evident limitation could be highlighted concerning the derivation of business dimensions from the taxonomies of previous scholars. As a result, the already specified business dimensions may reduce the exploratory assessment of other characteristics in the business models of data marketplaces. Additionally, considering the same time frame constraint, the analyses for the business models were limited to five marketplaces. The number five was chosen to have a good balance between quantity and granularity of analysis. More specifically, on the one hand, there are enough sources of comparisons, and on the other hand, in-depth investigation can be conducted.

9.5. Recommendations for Future Research

The research process and study results from this thesis could suggest a number of potential further research directions. Therefore, in the following section, several questions are described which may define the objectives for future research.

Similar to centralized marketplaces in the decentralized environment, challenges in the commercialization of intangible products could be encountered. Consumers want more than just products; they want solutions to their needs. Aspects as supporting the long tail of data sales to increase the data market share or curating available datasets to create problem solutions in different use cases were shortly described in sections 2.4 and 2.5. Although, creating problem solutions with data assets through reasonable strategies could represent a complex and essential research direction with questions like *What mechanism can be used to match actors with available data to increase data usability? Or how can we shift from data as a product to data as a problem solution?*

During the research process, it has been often mentioned that a functional ecosystem could represent an imperative practice in the context of distributed technologies. The ecosystem may enhance data exchange usability through different technologies and the expertise of third-party providers. However, motivating service providers that offer hard-to-find data to contribute to the data platform could represent a challenging task. For this purpose, *how to build ecosystems for value creation* could define a meaningful question for future research.

Blockchain also encounters technical and business challenges. A known technical challenge that is actively discussed in the development community is the scalability aspect. Blockchain technology requires a well-established network to support its infrastructure with considerable technical assets, such as computational power, storage space, or energy resources. Next, from a business perspective, it has been noted that marketplaces using blockchain technology most commonly place significant attention on crypto tokens governance. Although sustaining business models with cryptocurrency could be uncertain. Thus, several questions arise for further research: *Is the decentralized architecture from an operational viewpoint scalable? How to sustain comprehensive business models in the decentralized environment? What strategies can be applied in managing cryptocurrency assets? Or are these tokenization processes essential for the operation of data marketplace business models?*

References

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

Anderson, C. (2006). *The long tail: Why the future of business is selling less of more*. Hachette Books.

Bergman, R. (2020). *A Business Model Taxonomy for Data Marketplaces: Data Trade in Various Trading Structures*.

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

Burkhart, T., Krumeich, J., Werth, D., & Loos, P. (2011). *Analyzing the business model concept—a comprehensive classification of literature*.

Buschmann, F., Henney, K., & Schmidt, D. C. (2007). *Pattern-oriented software architecture, on patterns and pattern languages* (Vol. 5). John wiley & sons.

Carnelley, P., Schwenk, H., Cattaneo, G., Micheletti, G., & Osimo, D. (2013). Europe's data marketplaces—current status and future perspectives,'. *European Data Market SMART*, *63*.

Chesbrough, H. (2010). Business model innovation: opportunities and barriers. *Long Range Planning*, *43*(2–3), 354–363.

Dinh, T. T. A., Liu, R., Zhang, M., Chen, G., Ooi, B. C., & Wang, J. (2018). Untangling blockchain: A data processing view of blockchain systems. *IEEE Transactions on Knowledge and Data Engineering*, *30*(7), 1366–1385.

European Commission. (2020a). A European strategy for data. In *ec.europa.eu*. https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf

European Commission. (2020b). European data strategy. In *europa.eu*. https://ec.europa.eu/commission/presscorner/detail/en/fs_20_283

European Commission. (2021). *Financial services – EU regulatory framework for crypto-assets*. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12089-Financial-services-EU-regulatory-framework-for-crypto-assets_en

European Council. (2019). Building a data economy in the european union. Data. Innovation. Growth. In *ec.europa.eu*. https://ec.europa.eu/digital-single-market/en/news/building-data-economy-brochure?fbclid=IwAR1UkTQTMS4e1sQvh4MwOywItMpGtJOMLX4qzTi6AoVHY3NZ_gFn-tGzDEU

European Parliament. (2018). *Framework for the free flow of non-personal data in the European Union*.

European Parliament. (2019). *DIRECTIVE (EU) on open data and the re-use of public sector information*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2019.172.01.0056.01.ENG

European Parliament. (2020). *European data governance (Data Governance Act)*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020PC0767

finma. (2019). *FINMA publishes 'stable coin' guidelines*. https://www.finma.ch/en/news/2019/09/20190911-mm-stable-coins/

Fowler, M. (2012). *Patterns of Enterprise Application Architecture: Pattern Enterpr Applica Arch*. Addison-Wesley.

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

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

Gamma, E., Helm, R., Johnson, R., Vlissides, J., & Patterns, D. (1995). *Elements of reusable object-oriented software* (Vol. 99). Addison-Wesley Reading, Massachusetts.

Garcia-Font, V. (2021). Conceptual Technological Framework for Smart Cities to Move towards Decentralized and User-Centric Architectures Using DLT. *Smart Cities*, *4*(2), 728–745.

Hardjono, T., & Smith, N. (2016). Cloud-based commissioning of constrained devices using permissioned blockchains. *Proceedings of the 2nd ACM International Workshop on IoT Privacy, Trust, and Security*, 29–36.

Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian Journal of Information Systems*, *19*(2), 4.

Hynes, N., Dao, D., Yan, D., Cheng, R., & Song, D. (2018). A demonstration of sterling: A privacy-preserving data marketplace. *Proceedings of the VLDB Endowment*, *11*(12), 2086–2089.

Kim, M. S., & Chung, J. Y. (2019). Sustainable growth and token economy design: The case of steemit. *Sustainability*, *11*(1), 167.

Koutris, P., Upadhyaya, P., Balazinska, M., Howe, B., & Suciu, D. (2012). Querymarket demonstration: Pricing for online data markets. *Proceedings of the VLDB Endowment*, *5*(12), 1962–1965.

Koutroumpis, P., Leiponen, A., & Thomas, L. D. W. (2017). *The (unfulfilled) potential of data marketplaces*. ETLA Working Papers.

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

Koutsos, V., Papadopoulos, D., Chatzopoulos, D., Tarkoma, S., & Hui, P. (2020). Agora: A privacy-aware data marketplace. *2020 IEEE 40th International Conference on Distributed Computing Systems (ICDCS)*, 1211–1212.

Li, X., Jiang, P., Chen, T., Luo, X., & Wen, Q. (2020). A survey on the security of blockchain systems. *Future Generation Computer Systems*, *107*, 841–853.

Lim, S. Y., Fotsing, P. T., Almasri, A., Musa, O., Kiah, M. L. M., Ang, T. F., & Ismail, R. (2018). Blockchain technology the identity management and authentication service disruptor: a survey. *International Journal on Advanced Science, Engineering and Information Technology*, *8*(4–2), 1735–1745.

Mik, E. (2017). Smart contracts: terminology, technical limitations and real world complexity. *Law, Innovation and Technology*, *9*(2), 269–300.

Muschalle, A., Stahl, F., Löser, A., & Vossen, G. (2012). Pricing approaches for data markets. *International Workshop on Business Intelligence for the Real-Time Enterprise*, 129–144.

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

Noorian, Z., Iyilade, J., Mohkami, M., & Vassileva, J. (2014). Trust mechanism for enforcing compliance to secondary data use contracts. *2014 IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications*, 519–526.

opensource.guide. (2021). *Leadership and Governance. Growing open source projects can benefit from formal rules for making decisions.* https://opensource.guide/leadership-and-governance/

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

Özyilmaz, K. R., Doğan, M., & Yurdakul, A. (2018). IDMoB: IoT data marketplace on blockchain. *2018 Crypto Valley Conference on Blockchain Technology (CVCBT)*, 11–19.

Peffers, K., Rothenberger, M., Tuunanen, T., & Vaezi, R. (2012). Design science research evaluation. *International Conference on Design Science Research in Information Systems*, 398–410.

Ramachandran, G. S., Radhakrishnan, R., & Krishnamachari, B. (2018). Towards a decentralized data marketplace for smart cities. *2018 IEEE International Smart Cities Conference (ISC2)*, 1–8.

Santos, J., Spector, B., & Van der Heyden, L. (2009). Toward a theory of business model innovation within incumbent firms. *INSEAD, Fontainebleau, France*, 1–53.

Savelyev, A. (2018). Copyright in the blockchain era: Promises and challenges. *Computer Law & Security Review*, *34*(3), 550–561.

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

Sharma, P., Lawrenz, S., & Rausch, A. (2020). Towards Trustworthy and Independent Data Marketplaces. *Proceedings of the 2020 The 2nd International Conference on Blockchain Technology*, 39–45.

Spagnuolo, M., Maggi, F., & Zanero, S. (2014). Bitiodine: Extracting intelligence from the bitcoin network. *International Conference on Financial Cryptography and Data Security*, 457–468.

Spiekermann, M. (2019). Data marketplaces: Trends and monetisation of data goods. *Intereconomics*, *54*(4), 208–216.

Stahl, F., Schomm, F., Vomfell, L., & Vossen, G. (2015). *Marketplaces for digital data: Quo vadis?* Working Papers, ERCIS-European Research Center for Information Systems.

Stahl, F., Schomm, F., & Vossen, G. (2014a). *The data marketplace survey revisited*. ERCIS Working Paper.

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

Teece, D. J. (2010). Business models, business strategy and innovation. *Long Range Planning*, *43*(2–3), 172–194.

Third, A., & Domingue, J. (2017). Linked data indexing of distributed ledgers. *Proceedings of the 26th International Conference on World Wide Web Companion*, 1431–1436.

trusts-data. (2020). *About the project*. https://www.trusts-data.eu/about-the-project/

van de Ven, M., Antragama Ewa, A., Zenlin, K., & Mark, de R. (2021). *Creating a Taxonomy of Business Models for Data Marketplaces*. https://research.tue.nl/en/publications/creating-a-taxonomy-of-business-models-for-data-marketplaces

van de Ven, M. R. (2020). *Creating a Taxonomy of Business Models for Data Marketplaces*.

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: a framework for evaluation in design science research. *European Journal of Information Systems*, *25*(1), 77–89.

Verschuren, P., & Hartog, R. (2005). Evaluation in design-oriented research. *Quality and Quantity*, *39*(6), 733–762.

Wirtz, B. W., Schilke, O., & Ullrich, S. (2010). Strategic development of business models: implications of the Web 2.0 for creating value on the internet. *Long Range Planning*, *43*(2–3), 272–290.

Yue, L., Junqin, H., Shengzhi, Q., & Ruijin, W. (2017). Big data model of security sharing based on blockchain. *2017 3rd International Conference on Big Data Computing and Communications (BIGCOM)*, 117–121.

Zhang, W., Yuan, Y., Hu, Y., Nandakumar, K., Chopra, A., Sim, S., & De Caro, A. (2018). Blockchain-based distributed compliance in multinational corporations' cross-border intercompany transactions. *Future of Information and Communication Conference*, 304–320.

Zheng, Z., Xie, S., Dai, H.-N., Chen, X., & Wang, H. (2018). Blockchain challenges and opportunities: A survey. *International Journal of Web and Grid Services*, *14*(4), 352–375.

Zheng, Z., Xie, S., Dai, H., Chen, X., & Wang, H. (2017). An overview of blockchain technology: Architecture, consensus, and future trends. *2017 IEEE International Congress on Big Data (BigData Congress)*, 557–564.

Zott, C., Amit, R., & Massa, L. (2011). The business model: recent developments and future research. *Journal of Management*, *37*(4), 1019–1042.

Appendices

A. Content analysis

This appendix presents the analyzed documents of the five selected decentralized data marketplaces. The primary sources used in the analysis were the *official websites, technical or business whitepapers*, the *terms & conditions*, as well as other *external sources* relevant for the case.

Table 10: Data sources

Name	Source type	Source of document	Reference code
Ocean Protocol	Website	https://oceanprotocol.com/	OP-1
	Whitepaper	Ocean Protocol Foundation with BigchainDB GmbH (2020, December 09). Ocean Protocol: Tools for the Web3 Data Economy	OP-2a
		Ocean Protocol Foundation (2017, October 19). A decentralized data exchange protocol, powered by blockchain technology and a crypto token	OP-2b
	Terms and conditions	https://oceanprotocol.com/privacy	OP-3
	Platform blog	https://blog.oceanprotocol.com/	OP-4
dHealth Foundation	Website	https://dhealth.network/	dHF-1
	Whitepaper	dHealth Foundation (2021, June 16). dHealth Network. Blockchain Infrastructure for the Healthcare Market	dHF-2
	Terms and conditions	https://dhealth.network/terms-conditions/	dHF-3
	Platform blog	https://dhealth.network/news/	dHF-4
Streamr	Website	https://streamr.network/	STR-1
	Whitepaper	Streamr (2017, July 25). Unstoppable Data for Unstoppable Apps: DATAcoin by Streamr - Whitepaper	STR-2a
		Risto Karjalainen (2020, May 21). Governance in decentralized networks	SRT-2b
	Terms and conditions	Streamr Network and Marketplace (2018, May 30). TERMS OF USE	STR-3
	Platform blog	https://blog.streamr.network/	STR-4
Vetri	Website	https://vetri.global/	VET-1
	Whitepaper	Vetri (2021, May 8). Vetri – value your data.	VET-2
	Terms and conditions	vetri.globa - privacy policy	VET-3
	Platform blog	https://vetri.global/	VET-4
Enigma	Website	https://www.enigma.co/	EN-1
	Whitepaper	https://www.enigma.co/marketplace/	EN-2a
		Zyskind, G., Nathan, O., & Pentland, A. (2015). Enigma: Decentralized computation platform with guaranteed privacy	EC-2b
	Terms and conditions	https://www.enigma.co/privacy-policy/	EN-3
	Platform blog	https://scrt.network/	EN-4

A.1. Ocean Protocol (OP) Codes

The following table presents the codes that are subtracted from content analysis on the case documents of Ocean Protocol. The codes are aligned according to the business dimensions for the taxonomy of business models researched in chapter 4. In the coding process, three manual steps were followed. First, the predefined codes for the business characteristics were analyzed through deductive coding. These predefined codes were deducted from the three scientific articles that studied similar taxonomies (Bergman, 2020, Fruhwirth et al., 2020, M. van de Ven et al., 2021). Second, the documents were analyzed, and relevant quotes were subtracted. These quotes assigned the first set of codes. Third, the codes were categorized according to the frame of deductive coding. The categorization process exposed which themes come up the most and defined the ending coding for the business characteristics.

Table 11: Content analysis Ocean Protocol

Dimension	Characteristics	Quote
Industry domain	Any data	*"Ocean Protocol is a business, technical, and governance framework that is brought together to serve the needs of **all stakeholders in the data ecosystem**."* (OP-2b)
Participants	Data providers, data consumers, community, marketplaces, internal and external developers	*"Each component of the Ocean Protocol, **network**, and Ocean Token is designed to give **data providers** full control over how they publish and share their data. **Marketplaces** and **intermediaries** can provide tools to offer discovery and value-added services to data **consumers**."* (OP-2b)
Data service	Data brokering service	*"Ocean **aims to unlock data**, for more equitable outcomes, using a thoughtful application of both technology and governance."* (OP-4) *"Ocean smart contracts and libraries **make it easy to publish data services** (deploy and mint datatokens) **and consume data services** (spend datatokens)."* (OP-2a)
Data quality	User reviews	*"**Anyone** can add liquidity, aka stake (equivalent in AMMs). This is a curation, as stake is a proxy to **dataset quality**."* (OP-2a)
Privacy	Encrypted	*"…it stores the URL on-chain. So that others don't see that URL, **it encrypts it**."* (OP-2a)
Time frame	Multiple	*"Data providers – **any** enterprise, government, group or data custodian that possesses valuable but under-utilized data"* (OP-2b)
Smart contract	Configurable	*"Keeper-Contracts: **Support the requirements coming from Keeper-Contracts** that hold the Ocean Protocol Business Logic."* (OP-4) *"Contracts: **Support smart contract upgrades** and configuration with data persistence and immutability."* (OP-4)
Data mediating expertise	Provided by intermediaries	*"…**intermediaries** can provide tools to offer discovery and value-added services to data consumers"* (OP-2b)
Data output	(Both) standardized and aggregated	*"In the publish step, the publisher **invokes Ocean Datatoken Factory** to deploy a new datatoken to the chain."* (OP-2a) *"Data Union framework **provides a way to bundle a user's real-time data together** with others' and distribute a share of the revenue."* (OP-4)
Data access	Multiple options	*"The publisher can optionally **use IPFS for a content-addressable URL**."* (OP-2a)

		*"Compute-to-Data: let the data stay on-premise, yet allow 3rd parties to **run specific compute jobs** on it to get useful analytics results like averaging or building an AI model."* (OP-2a)
Data sources	Customer-provided data	*"New actors. Ocean Protocol has these actors: **Data Providers, who want to sell their data**; Data Consumers, who want to buy data; and **Marketplaces, to facilitate data exchange**."* (OP-2a)
Blockchain infrastructure	Public permissionless	*"Starting with Ocean V3.0 release, Ocean is **decentralized and permissionless**."* (OP-2a)
Cryptocurrency	(Both) payment method and utility tokens	*"We can formalize this right: the datatoken would typically automatically have a license to use that data."* (OP-2a) *"OCEAN is the default unit-of-exchange in Ocean Market and in Ocean libraries"* (OP-2a)
Data processing activities	Limited	*"Complementary to Ocean Market, Ocean **has reference code to ease building third-party data marketplaces**, such as for logistics or mobility."* (OP-2a)
Open-source governance	Meritocracy	*"The **community-driven OceanDAO** funds software development, outreach, and more."* (OP-2a)
Revenue model	Usage fees	*"…there will be **small transaction fees** that go to the Ocean community via a bridge."* (OP-2a)
Cryptocurrency governance	Cryptocurrency investment, token burning, maintain cryptocurrency stake, participants incentivization	*"Here's the loop: more usage of Ocean tools in the Data Ecosystem leads to **more OCEAN being staked**, leading to more OCEAN demand, **growing $OCEAN**.* ***More usage** also leads to **more Network Revenue**, which goes to (i) burning and (ii) OceanDAO. Burning OCEAN reduces supply to grow $OCEAN. **Funds go through OceanDAO to workers** who have the mandate to grow usage of Ocean tools. And the loop repeats."* (OP-2a)
Pricing model	Fixed pricing	*"Ocean Market supports **fixed pricing** and automatic price discovery."* (OP-2a)
Price discovery	Set by the marketplace provider or automatic price discovery	*"AMMs provide **automated price discovery** without the disadvantages of order books or auctions. AMMs work for an initial asset offering and throughout the asset's lifetime."* (OP-2a)

A.2. dHealth Foundation (dHF) Codes

The following table presents the codes that are subtracted from content analysis on the case documents of dHealth Foundation. The codes are aligned according to the business dimensions for the taxonomy of business models researched in chapter 4. In the coding process, three manual steps were followed. First, the predefined codes for the business characteristics were analyzed through deductive coding. These predefined codes were deducted from the three scientific articles that studied similar taxonomies (Bergman, 2020, Fruhwirth et al., 2020, M. van de Ven et al., 2021). Second, the documents were analyzed, and relevant quotes were subtracted. These quotes assigned the first set of codes. Third, the codes were categorized according to the frame of deductive coding. The categorization process exposed which themes come up the most and defined the ending coding for the business characteristics.

Table 12: Content analysis dHealth Foundation

Dimension	Characteristics	Quote
Industry domain	Heath & Personal data	*"The dHealth Network can align the incentives of* ***all stakeholders in the healthcare system*** *by making its processes more efficient, transparent, and valuable."* (dHF-1)
Participants	Patients, industry, researchers, healthcare providers, supernode operators, internal developers	*"The dHealth Network launched 29 March 2021 with a limited number of* ***Supernodes operated by partners*** *that share the dHealth Foundation's vision of an equitable healthcare system."* (dHF-2) *"Health Network is open to* ***third-party developers*** *to create their applications."* (dHF-2)
Data service	Healthcare data brokering service	*"These technological advances will allow the network to provide the health* ***information infrastructure of the future that is interoperable with other blockchains****."* (dHF-2)
Data quality	Marketplace providers review	*"Supernode Operators enjoy the privilege of* ***defining their own Namespace*** *and* ***creating their own network tokens*** *at no additional costs. They have* ***direct access to the blockchain*** *for their applications and* ***possess special voting rights*** *regarding the technical development of the network."* (dHF-4)
Privacy	Encrypted	*"All information you provide to us is* ***stored on our secure servers****. Any web-based payment transactions* ***will be encrypted****."* (dHF-3)
Time frame	Multiple	*"It provides the pillars of* ***a real-time and efficient*** *data-transaction healthcare ecosystem."* (dHF-1)
Smart contract	Standardized	*"dHealth Network continues to use the Symbol protocol for its L1 infrastructure. Symbol is well suited for* ***transactional use cases, high throughput, and plug & play capabilities*** *for asset management."* (dHF-4)
Data mediating expertise	Provided by intermediaries	*"Health Network is open to* ***third-party developers to create their applications****."* (dHF-2) *"Supernode operators enjoy the privilege of* ***defining their own Namespace*** *and* ***creating their own network*** *tokens"* (dHF-4) *"Validator Nodes operate as access nodes (API gateway). The primary responsibility of an API gateway is* ***to make the data accessible in a readable format****."* (dHF-4)
Data output	Both standardized and aggregated	*"dHealth Network comes with functionality such as* ***aggregated transactions and cross-chain swaps****."* (dHF-2)
Data access	API	*"Every Supernode in the* ***network provides API*** *endpoints that can be called by applications to perform functions on the blockchain."* (dHF-2)

Data sources	Customer-provided data	*"…most individuals are not interested in their health-related data, unless they are sick. For them, tokens can be used as incentives to make their data shareable."* (dHF-2)
Blockchain infrastructure	Private permissioned	*"The healthcare-dedicated network of Supernodes serves as **an alternative solution for permissioned**…"* (dHF-1) *"The nodes of **the public dHealth Network are spread out geographically** to provide network stability and accessibility."* (dHF-1)
Cryptocurrency	Payment method and utility tokens	*"The native currency of the network is the Digital Heath Point (DHP). **DHP is a utility and payment token**."* (dHF-2)
Data processing activities	Limited	*"dHealth Network comes with functionality such as **aggregated transactions** and cross-chain swaps. These technological advances will allow the network to provide the health information infrastructure of the future that is interoperable with other blockchains. **dHealth Network is open to third-party developers to create their applications**."* (dHF-2)
Open-source governance	Benevolent Dictator	*"As a **community-owned and distributed platform** for health-related transactions and data-access, the dHealth Network puts the individual at its center."* (dHF-2) *"This organization has no public members. **You must be a membe**r to see who's a part of this organization."* (GitHub)
Revenue model	Usage fees	*"**Transaction fees** generated on the dHealth Network must be paid in DHP. It **will be divided according to their importance among the accounts participating** in the Delegated Harvesting."* (dHF-2)
Cryptocurrency governance	Cryptocurrency sale, incentivization, staking reward	*"A total of 2 billion (2'000'000'000) DHP will be the circulating supply on the dHealth Network. **Half of it will be created as Staking Rewards during 10 epochs** or 12.5 million blocks and distributed to eligible harvesting accounts (see Figure 1). Besides, **20% is reserved for the Innovation Fund** to initiate and facilitate the implementation of dHealth Network use cases."* (dHF-2)
Pricing model	Fixed pricing	*"Operators of Supernodes **can issue their own tokens on the network including stable coins**."* (dHF-1)
Price discovery	Set by marketplace healthcare providers	*"Operators of Supernodes can issue their **own tokens on the network including stable coins**."* (dHF-1)

A.3. Streamr (STR) Codes

The following table presents the codes that are subtracted from content analysis on the case documents of Streamr. The codes are aligned according to the business dimensions for the taxonomy of business models researched in chapter 4. In the coding process, three manual steps were followed. First, the predefined codes for the business characteristics were analyzed through deductive coding. These predefined codes were deducted from the three scientific articles that studied similar taxonomies (Bergman, 2020, Fruhwirth et al., 2020, M. van de Ven et al., 2021). Second, the documents were analyzed, and relevant quotes were subtracted. These quotes assigned the first set of codes. Third, the codes were categorized according to the frame of deductive coding. The categorization process exposed which themes come up the most and defined the ending coding for the business characteristics.

Table 13: Content analysis Streamr

Dimension	Characteristics	Quote
Industry domain	Any data	*"**Anyone – or anything – can publish new data** to data streams, and others can subscribe to these streams."* (STR-2a)
Participants	Data publisher, data subscriber, streamr broker nodes, marketplaces, internal developers, external developers	*"Streamr Network is the data transport layer in the technology stack. It consists of **Streamr Broker nodes** which establish a P2P network. The **network hosts a publish/subscribe mechanism** and supports decentralized storage of events."* (STR-2a)
Data service	Customized map service with publish-subscribe system	*"The Streamr **Network is a decentralized, topic-based publish-subscribe system**. Each publish-subscribe topic, referred to as a stream in the Streamr Network, has its own peer-to-peer overlay network that is built and maintained by a set of BitTorrent-like trackers."* (STR-2a)
Data quality	Marketplace review	*"There may also be a **mechanism for subscribers to flag bad data**, negatively affecting the publisher's reputation. These safety features ensure that the **publisher cannot get paid without delivering quality data** as promised."* (STR-2a) *"A data producer **gains karma when events she published are delivered** to subscribers. Subscribers earn karma by receiving events. Broker nodes earn karma for helping with data delivery and persistence. Bookkeeping is easy: The amount of new karma equals the amount of DATAcoin exchanged. The difference is that karma decays and eventually expires, while the token balance does not."* (STR-2a)
Privacy	Encrypted	*"Given that anyone can participate in the Streamr network by running a broker node, event payloads of non-public streams in the Streamr **network are strongly encrypted using asymmetric key cryptography**."* (STR-2a)
Time frame	(Near) real-time	*"Streamr. The decentralized platform for **real-time data**"* (STR-1)
Smart contract	Configurable	*"Streamr Smart Contracts enable nodes in the Streamr network to **reach consensus, hold stream metadata, handle permissioning and integrity checking, and facilitate secure token transfers**."* (STR-2a)
Data mediating expertise	Provided by marketplace	*"We foresee an ecosystem where there are several usability platforms and tools available. The existing **Streamr platform already implements some elements** of the usability layer, with **more functionality being added** in the coming months and years. **The aim is to reach a stage where you can build**

		and deploy a useful and functioning data-driven *smart contract in minutes."* (STR-2a)
Data output	Standardized	*"Each publish-subscribe topic, referred to as a stream in the Streamr Network,* ***has its own peer-to-peer overlay network that is built and maintained by a set of BitTorrent-like trackers****."* (STR-2b)
Data access	Specialized software	*"Alongside the decentralized data network and marketplace, the full Streamr stack includes a* ***powerful analytics engine and a UI for rapid development of real-time Ðapps****"* (STR-2a and STR-1)
Data sources	Customer-provided data	*"****Anyone – or anything – can publish new data*** *to data streams, and others can subscribe to these streams to power Ðapps.* *To* ***incentivize user participation*** *in the network, there's a built-in mechanism for data monetization"* (STR-2a)
Blockchain infrastructure	Public permissionless	*"Instead of a centralized party, the Network is run by its community of users, and it heavily relies on cryptography to remove the need for trust. This creates a* ***permissionless and neutral network for real-time data****."* (STR-4)
Cryptocurrency	Payment method	*"DATAcoin is implemented as an ERC20 token on Ethereum. The token smart contract maintains DATAcoin balances, and* ***ensures that payments are handled in a trustless and secure way****. Following the ERC20 standard ensures interoperability with wallets and other tokens."* (STR-2a)
Data processing activities	(All) provided by marketplace	*"Streamr Editor enables rapid development of data-driven smart contracts, lowers the threshold for Ðapp creation, and comes with* ***ready-made templates for common use cases built in****."* (STR-2a and STR-1) *"Streamr Engine is the* ***high-performance analytics engine*** *that executes off-chain within a decentralized computing provider."* (STR-2a)
Open-source governance	Benevolent Dictator	*"The open-source contribution model does* ***not necessarily have any connection to open-source governance****. In practice, governance style ranges from centralized models to decentralized meritocracies."* (STR-2a)
Revenue model	Commissions	*"Admins have the power to add and remove members. They are responsible for maintaining their Data Unions, including ensuring good data quality and removing members that are not contributing data as they're expected to.* ***Admins are incentivized to perform this work by the Admin fee parameter, a fraction of the incoming Data Union revenue****."* (STR-1)
Cryptocurrency governance	Sustained network, monetary policy	*"One objective of governance can be a creation of a model where participants contribute to common goals even when they act in their own interests. The incentive structure is related to the* ***monetary policy*** *and* ***token economics in the network, and its design is a non-trivial problem*** *in its own right."* (STR-2b)
Pricing model	Flat fee tariff	*"A license gives the right to access the data for a specific* ***period of time****, on certain conditions, and* ***for a fee****."* (STR-2a)
Price discovery	Set by the marketplace provider	*"****You can make your product free or paid****. For paid products,* ***set price by minute, hour, day or week intervals****. You can switch between viewing price in Streamr's DATA token or fiat currency for your convenience. You can also choose to fix price in fiat to protect against sudden shifts in the DATA price"* (STR-4)

A.4. Vetri (VET) Codes

The following table presents the codes that are subtracted from content analysis on the case documents of Vetri. The codes are aligned according to the business dimensions for the taxonomy of business models researched in chapter 4. In the coding process, three manual steps were followed. First, the predefined codes for the business characteristics were analyzed through deductive coding. These predefined codes were deducted from the three scientific articles that studied similar taxonomies (Bergman, 2020, Fruhwirth et al., 2020, M. van de Ven et al., 2021). Second, the documents were analyzed, and relevant quotes were subtracted. These quotes assigned the first set of codes. Third, the codes were categorized according to the frame of deductive coding. The categorization process exposed which themes come up the most and defined the ending coding for the business characteristics.

Table 14: Content analysis Vetri

Dimension	Characteristics	Quote
Industry domain	Personal data	*"VETRI consists of a data wallet and a marketplace. The **data wallet allows you to store and manage all your personal data**. The marketplace enables you to sell parts of your data to researchers, businesses, or marketers in an anonymized way."* (VET-1)
Participants	Data users, data consumers, third parties validators, internal developers	*"The VETRI platform will consist of a mobile wallet (VETRI wallet) for **users to manage their personal** data as well as a web application (VETRI marketplace) for **data consumers** to buy and access that data. Users and data consumers, who together form the "stakeholders", will be able to add data, **request data form third parties, get their data verified**, manage their privacy settings, buy services and finally share and monetize their data in a fully user-controlled fashion."* (VET-2)
Data service	Customized map services with specialized an application that groups data for different use cases	*"As a decentralized application interacting with users' personal data, privacy and security will be VETRI's number one priority. VETRI users will store their most sensitive data locally on their device by **using state-of-the-art encryption techniques and the application itself will be locked by multiple factors of user authentication**."* (VET-1)
Data quality	Intermediaries reviews	*"VETRI's **identity platform relies on third-party attribute verifications**, it can claim a high degree of (demographic) **data reliability**. And because VETRI is built around data privacy and security by design, users don't mind inputting their real personal data."* (VET-2)
Privacy	Anonymized	*"With the VETRI wallet, users can directly and easily start monetizing their personal data in a fully secure and controlled fashion on the VETRI marketplace. In return, data consumers **gain direct access to reliable, anonymized data** in a compliant and cost-efficient way."* (VET-2)
Time frame	Up-to-date	*"With VETRI, users will be able to **pro-actively share anonymized demographic and psychographic data**, thus enabling the creation of target audience filters on the VETRI platform for digital advertisers to use."* (VET-2)
Smart contract	Standardized	*"**The VETRI marketplace connects** identity holders and data consumers through blockchain smart contract technology and matches supply and demand commission-free."* (VET-2)
Data mediating expertise	Provided by intermediaries	*"Furthermore, depending on the success of the VETRI marketplace, Procivis AG, **along with third-party service providers, intends to build value on top** of the VETRI marketplace, for instance by offering VETRI wallet owners secure*

		storage and other services that can be paid for with VLD tokens" (VET-2)
Data output	Aggregated	*"Entities receiving and using data sets. In the context of VETRI, this would typically be* ***companies conducting research, surveys or digital advertising campaigns****."* (VET-2)
Data access	Specialized software	*"The VETRI platform will consist of a mobile wallet (VETRI wallet) for users to manage their personal data as well as a* ***web application (VETRI marketplace)****"* (VET-2)
Data source	Customer-provided data	*"With VETRI, users will be able to* ***pro-actively share anonymized demographic and psychographic data****."* (VET-2)
Blockchain infrastructure	Hybrid	*"As VETRI will be handling a very significant amount of data,* ***using a hybrid approach*** *is to our current knowledge the best option to overcome the processing limitations of public blockchains."* (VET-2)
Cryptocurrency	Payment method	*"The VETRI token enables* ***buy and sale transactions*** *between users and data consumers on the VETRI marketplace and provides a unit of value."* (VET-2)
Data processing activities	(All) provided by marketplace	*"****Services for data consumers could also include the development of sophisticated analytics tools*** *and smart contracts to optimize results from data gathering efforts."* (VET-2)
Open-source governance	Benevolent Dictator	*"All software and applications developed for the VETRI* ***platform will be rendered open source and transferred to the Foundation upon testing*** *and completion of the platform in 2021."* (VET-2)
Revenue model	Subscription	*"The Foundation's funding will be provided by a significant token allocation and be sustained by standard* ***open-source licensing revenue models****."* (VET-2)
Cryptocurrency governance	Cryptocurrency selling, users incentivization, burning	*"The total number of VLD tokens sold in the ITO will always make up for 50% of all minted tokens. All unsold tokens in the ITO will be burnt and proceeds from the token sale will be allocated proportionally as per the below cost category allocation."* (VET-2)
Pricing model	Pay-per-use	*"...users will be remunerated* ***in VLD tokens based on the desirability of the shared data****"* (VET-2) *"We create a variety of data that is useful for different types of research, but maybe you do not want to share all of it.* ***With VETRI you control what is shared****."* (VET-1)
Price discovery	Set by the marketplace	*"As a reward, users will be remunerated in VLD tokens based on the desirability of the shared data as perceived by data consumers and using straight-forward* ***market-based pricing mechanisms****."* (VET-2)

A.5. Enigma (EN) Codes

The following table presents the codes that are subtracted from content analysis on the case documents of Enigma. The codes are aligned according to the business dimensions for the taxonomy of business models researched in chapter 4. In the coding process, three manual steps were followed. First, the predefined codes for the business characteristics were analyzed through deductive coding. These predefined codes were deducted from the three scientific articles that studied similar taxonomies (Bergman, 2020, Fruhwirth et al., 2020, M. van de Ven et al., 2021). Second, the documents were analyzed, and relevant quotes were subtracted. These quotes assigned the first set of codes. Third, the codes were categorized according to the frame of deductive coding. The categorization process exposed which themes come up the most and defined the ending coding for the business characteristics.

Table 15: Content analysis Enigma

Dimension	Characteristics	Quote
Industry domain	Any data	*"A peer-to-peer network, enabling **different parties to jointly store and run computations on data** while keeping the data completely private."* (EN-2b)
Participants	Data providers, data consumers, known nodes, internal and external developers	*"…we launch our open off-chain network where anyone can become a node and provide storage and computational resources in return for ENG tokens."* (EN-2a) *"…we're introducing Secret Fellows, a program for developers interested in building dApps and tooling for Secret Network in return for competitive compensation and mentorship"* (EN-4)
Data service	Data brokering service	*"**Large organizations can use Enigma to protect their data and trade secrets** from corporate espionage and rogue employees. Employees can still use and analyze data for the benefit of the organization, but won't be able to steal any data."* (EN-2b)
Data quality	Market providers reviews	*"With the assumption that **providers are rational agents** who wish to maximize their utility, this ensures their **best strategy is to provide <u>quality data</u>** and guarantee its availability."* (EN-2a)
Privacy	Encrypted	*"In particular, **the off-chain component handles everything** from routing requests, verifying permissions, parsing and forwarding queries to the source and finally —routing the result back in the peer-to-peer network to the requesting client (**encrypted** with the client's private key)."* (EN-2a) *"…users are able to share their data with **cryptographic guarantees** regarding their privacy."* (EN-2b)
Time frame	Up-to-date	*"The data provider, through their off-chain client, submits a cron-like job with the data curation script and details on how frequently to execute the job (e.g., minutely, hourly, daily). **This ensures the data always remains up-to-date**."* (EN-2a)
Smart contract	Standardized	*"The smart contract allows data curators to **register new data** sets, and allows users to **subscribe to existing data** sets. The subscription **period defaults to one month, and prices for all data sets are designated accordingly as monthly costs**."* (EN-2a)
Data mediating expertise	Provided by marketplace	*"Enigma is private. Using secure multi-party computation (sMPC or MPC), data queries are computed in a distributed way, **without a trusted third party**."* (EN-2b) *"Our goal is to enable developers to build 'privacy by design', end-to-end decentralized applications, **without a trusted third party**."* (EN-2b)

Data output	Standardized	*"To register a new data set, download and install the **Catalyst client**. Catalyst will process any file with a **CSV extension**."* (EN-2a)
Data access	Specialized software	*"**Catalyst users** may subscribe to data sets from the Enigma Data Marketplace and use them in their trading algorithms. Depending on their properties, data sets may be usable for backtesting as well as live trading."* (EN-2a)
Data sources	Customer-provided data	*"**The data provider, through their off-chain client,** submits a cron-like job with the data curation script and details on how frequently to execute the job."* (EN-2a)
Blockchain infrastructure	Hybrid	*"Code is executed both on the blockchain (**public parts**) and on Enigma (**private or computationally intensive parts**)."* (EN-2b)
Cryptocurrency	Payment method	*"As a data provider, you can publish your data sets to the Enigma Data Marketplace and get **rewarded in ENG tokens**."* (EN-2a)
Data processing activities	Limited	*"The key new utility Enigma brings to the table is the **ability to run computations on data**, without having access to the raw data itself."* (EN-2b)
Open-source governance	Meritocracy	*"...the Secret Fellows developer program, which is designed to **involve more developers full-time in the Secret Network ecosystem**"* (EN-4)
Revenue model	Asset sales	*"Secret Network has a twenty-one day **unbonding period, with a circulating supply of approximately 70 million SCRT** and a total supply of approximately 190 million SCRT."* (EN-4)
Cryptocurrency governance	Cryptocurrency investment, block rewards, maintain cryptocurrency stake	*"Tokenomics. Secret Network leverages inflation, block **rewards, and staking to incentivize SCRT holders and validators to bond their tokens to the network**."* (EN-4)
Pricing model	Flat fee tariff	*"Once a user finds a data-set that they like and wish to subscribe to, they send a subscription transaction specifying the unique path of the data-set, as well as the required payment for it. **This transaction can be automated in the client so that it is repeated <u>every month</u>**."* (EN-2a)
Price discovery	Set by the marketplace provider	*"As a provider, **you set the price of your data sets** for monthly subscriptions."* (EN-2a)

B. Expert Interviews

In this appendix, two examples from the six conducted interviews with researchers and industry experts are presented. If the reader requires the other expert interviews, they could be made available on request. These expert interviews contributed to the assessment of the artifacts in chapters of formative and summative evaluation. To perceive the privacy of the interviewed researchers and experts, their names, contact details, or any other personal information are anonymized. Although, the only relevant information for the research is their recommendations according to the evaluation properties. In the following, the general information about the interviewed experts is presented. Thereafter, the interview transcripts are attached.

Table 16: Interviewed researchers and industry experts

Expert (E)	Role	Relevance	Subject Area
E1	Research partner working as part of the EU Horizon 2020 project TRUSTS, with the primary focus on deriving the business ecosystem modeling, commercialization, and innovation impact assurance for secure, trustworthy, GDPR-compliant data platforms.	Provided a critical point of view on the concept of decentralized data marketplaces and their business modeling characteristics in the data industry.	Decentralization, data marketplaces, business modeling, and platform ecosystems.
E2	Top manager at Ocean Protocol focusing on establishing strategic partnerships and ensuring growth opportunities for the Ocean Protocol ecosystem.	Provided insights on the concept of data marketplaces with blockchain architecture and their characteristics in building functional ecosystems for the data market.	Decentralized technologies, data marketplaces, and platform ecosystems.
E3	Senior researcher at the Research Institution of Sweden (RISE).	Provided interesting viewpoints in data sharing with practical reference to the transportation industry.	Data marketplaces and decentralized architecture.
E4	Senior researcher and advisor at the Research Institution of Sweden (RISE).	Provided insights on data sharing and management in compliance with mobility services.	Data marketplaces, business modeling, and mobility services.
E5	Research partner at the Institute for Software Technology IST within the Faculty for Computer Science of the University Koblenz-Landau.	Provided expertise on the concept of data marketplaces and the mechanisms of trading data assets.	Data marketplaces and decentralized architecture.
E6	Senior researcher at the Institute for Information Systems and New Media of the University of Vienna.	Provided insights on the data markets, human-centric information systems, and algorithmic accountability.	Data marketplaces decentralization and digital privacy.

B.1. Interview Transcript 1

General Questions

- From the research conducted in the last couple of months, I observed that data marketplaces are still emerging as digital platforms. Most marketplaces are trading domain-specific data or deliver technical solutions for data providers to exchange data with a closed group of parties where agreements are negotiated in bilateral contracts. Is this reluctance to trade data with more participants in many-to-many marketplaces related to the low trust in a single party, privacy concerns, transparency in exchange, possible security breaches? Or the data culture is not yet present?

It depends on the use case you are discussing. I don't think we can generalize. In my perception, this is not a question where we talk about centralized or decentralized marketplaces. There are, of course, aspects where trust plays a role. For example, think of sensitive data in healthcare or financial services; you expose yourself to high risk if you cannot control data access and data usage. The challenge is not that a data marketplace itself is not trusted but rather that such organizations want to keep relatively tight control. This requires a tight contract with trusted parties. For instance, if I am Deutsche Bank, then I want to know what is being done with data and by whom. It has nothing to do with the data marketplace but more with the data asset buyer. If there were a data leak, then I would face reputational and potential legal repercussions. As an organization using a marketplace, this would be something that I don't want to take on myself. Thus either I would reframe from providing access for this type of data, or obviously, I need to select and pre-process these data assets to a degree where it doesn't really provide the full scale of meaningful use cases.

Another aspect is regarding cost and effort. More often than not, we see one-to-one use cases that emerge by using a search functionality on a platform. So, for one-to-one uses cases, the questions are: Do I have the internal reliance as far as data security or data governance is concerned? How long does the data preparation take? At what costs do I need to do this? And is the risk-return profile that I create adequate and beneficial?

Lastly, there is the organizational and human factor. This relates to some extent to the soft side of data governance. Ultimately, when we talk about data providers, we interact with complex social-dynamic systems. For example, the business owner of data assets who might be interested in exploiting them through data exchange or trading may be hesitant to take this ultimately personal risk and may be hesitant to concert the required multi-functional alignment, review, and approval processes required in large corporate set-ups.

- The focus of the research I am conducting is to visualize a many-to-many decentralized data marketplace. However, considering that most of the marketplaces trade domain-specific data could a many-to-many marketplace have a competitive chance of success? Making this question easier, Is there a chance of success for an eBay for data?

As for data trading: one of the central problems is our still incomplete understanding of data as an economic asset. I am very glad you mentioned eBay. Because often, people referring to data marketplaces don't think "this is an eBay for data" but rather imagine "this is an Amazon for data". Yet, most data markets are eBays for data. So, they are inadvertently in the business of supporting long-tail data sales. Obviously, to counter this, they may supplement with availing data for high-scale use cases, which exist in time-stable contexts of data. This means matching a marketplace with a multitude of actors with available data to increase the probability of usability and value creation with data. For instance, if you have an automotive-focused data market with car manufacturers and original equipment manufacturers (OEMs) on board, you could increase the probability of matching data supply and demand, e.g. to get relevant data that can be applied for the time-stable context of mobility and logistics solutions. In this regard, you automatically shift from data assets as a product to data assets as a problem solution. From what I observed, successful data markets are mostly domain-focused, or in the case of general marketplaces, they are trying to segment the data overall space through curation to create problem solutions as previously described. We should always look at ways that combine data for value creation within the ecosystem. Another attempt to achieve this in a cost-efficient way is automated harvesting and cleansing of open data. Personally, I would think twice before developing an infrastructure for a general-purpose data market as a standalone infrastructure, simply hoping an efficient scale of data exchange would occur. For such a platform to succeed takes an ecosystem, and they are very hard and take so long to build. Providing data market functionality is so much easier as an add-on to an existing ecosystem. Think hyper-scaled BigTech cloud solution ecosystems or emerging blockchain technology ecosystems. Going it alone is certainly more likely to succeed in a very focused domain-niche with common standards, and a sufficient number of players committed to leading the platform to success.

- From your expertise, do you think we could mitigate some concerns of the data providers with a decentralized architecture?

If we look at meaningfulness and adequacy, I struggle to see the superior benefits of decentralized architecture in many aspects, particularly if decentralization is single-mindedly equated with the use of blockchain technology.

Before I get to the core of your question, please consider two aspects at a systemic level of abstraction.

First, suppose you look at the European and country-based support for nurturing the data economy. Many of the initiatives we see in the last five years have been publicly funded, trying to utilize what was perceived as the upcoming opportunities, in a somewhat inefficient portfolio approach. So, if you as a researcher now observe existing data markets, is that a true reflection of what is the most meaningful? No, this reflects what was fueled by the inner mechanics of how public R&D programs and grants are managed. Massive amounts of fundings were funneled into this experimental space and of course, participating academia was also interested in exploring new fields.

Second, you should also consider the workings of venture funding. When the public blockchain hype took off, many initiatives arose from market players feeling "I don't fully understand what this is, but I can't miss this race", and thus a massive amount of funding went there. So when you do the research saying, "What is there? Is that a representative of what is meaningful?" no, it does not answer this question; it just really answers, "What is there?".

The problem of data markets is not one of lacking trust fueled by incidents along the lines of "…oh there was another data market that was hacked because it is centralized". Centralization has not led to a lack of trust. You still need central components like identity management, brokerage, execution, facilitation. It seems like we are trying to solve a problem that either doesn't need to have a solution, or we try to find a solution to applying decentralization – which indeed has benefits in many domains – and novel technology solutions (blockchain) whilst we still don't have an emerging understanding of what it could be.

Let us consider what decentralization means for trust. Just follow the news, and check "when did a major bank last time got hacked on a large scale?"; and then "when did the last crypto exchange got hacked?". What we believe to achieve with decentralization often is a projection of human nature because we all like to feel the illusion of being in control. Indiscriminately jumping onto decentralization without deeper reflection and a very clear weighing of all the pros and cons is akin to saying "Let’s not trust trains because these are the big things on a heavy, aged infrastructure"; but ultimately, car traffic is by far more unsafe, resulting just in more personal damage and, worse yet, externalities. Security in data exchange, as far as encryption is concerned, has long been solved. The real problem resolves in the systems, storage, and interfaces of data source and sink, and in the problem of data sovereignty. When we look at decentralized architecture and what is often being done, is that it just pushes the problem to the periphery. This means that marketplaces shift the risk from themselves to peer-to-peer exchange, to the providers and consumers managing data. From a risk perspective, and from a cost perspective – just think centralized storage and security – that is perfect. And from the user’s perspective, this removes another point of failure from the equation (let's hope that they use secure interfaces and connectors). But beyond that, I don't see how a decentralized architecture solves the existing fundamental problems of data trading in a meaningful way and what the new technology (blockchain) adds to it. Economics and sustainable business models of data markets at large are what require attention. How can we build ecosystems for value creation is a far more meaningful question than how to solve a postulated problem of trust in marketplace operators.

Validation of the High-Level Architecture

- We could move to the validation part of the interview, starting with the high-level architecture. Are the infrastructure components comprehensive in naming, definition, and functional requirements?

The slide shown provides a rather static representation. You would see if the architecture works by trying to put dependencies between the components. Another aspect I'd like to mention is that if you have a layered architecture between components, typically, you don't have broken (interacting) layers. The smart contract layer on the slide suggests that there are two different smart contracts, which I do not see.

Another thing is looking at the horizontal security layer; I would typically represent it as more of an infrastructure security aspect. Better add a horizontal layer of other components, such as authentication, assets discovery, and access negotiation within your architecture.

Validation of the Taxonomy of Business Models

- Are the characteristics described according to the business model taxonomy detailed and complete in the business model taxonomy? Or do we miss any essential components, dimensions, or not detailed characteristics of decentralized data marketplaces in the taxonomy?

I think it is always fit-for-purpose, so at an aggregate level, it is fine. I trust you already researched the previous taxonomies that we've created in TRUSTS papers.

- A change I made is related to contacts from the value creation that I labeled as smart contracts

If your research topic discusses decentralized data marketplaces and their business models, then I don't necessarily see the need to link it singularly to blockchain and smart contracts. This is because smart contracts are one of the most misunderstood things. Smart contracts are exactly not smart. Instead, they are programmed more often than not for small, non-complex contracts because you want to scale.

It gets interesting that even with legal contracts, they get complex and normal people do not understand them intuitively. So now you get this half-understandable agreement to a smart contract, where agreements are coded and hard to audit? Does this create an additional layer of trust between the business and users on both ends? I don't even know what is going on there; I have to blindly trust more parties involved. In the context of data markets, smart contracts are de facto just being used to auto-execute the most simple contracts. I understand that they are great conceptually, and possibly for other use-cases in the blockchain community, but it doesn't have to be the complexity of the context behind them.

That said: For the purpose of your research, leave it in, so that it does not side-track your research and others. But please be aware that this somewhat unduly narrows the outcome vis-à-vis the stated research intent.

- Another observation in the research process is that most decentralized data marketplaces are sustaining their business models through token economy principles. In my perception, these principles are based on (1) crowdfunding through initial coin offering, (2) strategies to raise the value of the digital assets, (3) increase the user base, and (4) continuously developing marketplace through applying some of the philosophies of the open-source projects till the marketplace could reach being considered a self-sustainable platform. Doesn't the hype and trend of cryptocurrency cover this process? Or, in

principle, could we deviate from cryptocurrency and evaluate the platform through the digital assets it offers? (Critical perception of the blockchain)

This question is more like what happens in real life vs. what happens in the research. It looks like a base rate fallacy; maybe the six decentralized data marketplaces you picked use tokenization, but they are not representative of the bulk of data market platforms. Of course, if your platforms inherently arise from blockchain ecosystems, rather than "just so happen to use blockchain technology", this could be true. However, if you want to draw a more general interpretation, this singular focus overly narrows the research.

There is another aspect related to when tokens could be meaningfully utilized. So either you do it because you are trying to operate in an established, wider blockchain ecosystem, interacting with other DAPs, etc. For instance, you're on Ethereum or better yet on the Internet Computer, and tokens could enable the interactions between services and assets within the ecosystem. In that case, then you utilize what is already there. However, if it is a standalone data market, I don't see the point of meaningfully utilizing tokens. This is because you mainly want to utilize tokens where you got repeat transactions or if it comes to voting rights or incentivization. And of course, a domain-focused data market tokenization might be a valid funding pathway for those heavy users. But for data markets that are more like eBay for data, e.g. you are "one-off" buying a big excel file from the University of Vienna: Wouldn't it be easier for everyone just for a simple money transfer of five euros? In this case, it is just another, in my view unnecessary layer of complexity to include tokens.

B.2. Interview Transcript 2

- From the research conducted in the last couple of months, I observed that data marketplaces are still emerging as digital platforms. Most marketplaces are trading domain-specific data or deliver technical solutions for data providers to exchange data with a closed group of parties where agreements are negotiated in bilateral contracts. Is this reluctance to trade data with more participants in many-to-many marketplaces related to the low trust in a single party, privacy concerns, transparency in exchange, possible security breaches, or the data culture is not yet present?

That is a good question; if I knew the answer, that would definitely help unlock a lot of value across Ocean Protocol and other marketplaces. But, regarding the second point – "if people are not comfortable trading data" – that plays a significant role. Having data is seen as an asset, although sharing data is seen as making yourself vulnerable or reducing your competitiveness. In general, that is not something that companies feel comfortable doing (trading data). They are reluctant to share data because they spent money, resources, and time accumulating data as their competitive advantage.

If companies have data, they are monetizing it themselves. For example, traditional Web2 companies are monetizing the data by providing other services and products to the market. In this regard, by making data available, companies are challenging their own business models.

Alternatively, business models are emerging through Web3, enabling individuals and organizations to monetize data. For example, Swash is a browser extension developed to collect browser search data. This extension can be connected to Streamr or Ocean Protocol, allowing individuals to offer their browser history data in return for monetization. So, giving more control to the individual in terms of sharing data. These are some services and plugins that Web3 enables.

Although Ocean Protocol has its own market as a reference marketplace, our core business is not a marketplace. Instead, we want to enable the ecosystem of marketplace**s** built on the Ocean Protocol technology by working with partners to build their own marketplaces and provide value add services. For example, storage capacity, computational power, or KYC (know your customer) could be plugged into the ecosystem. These services can be developed by third parties and not necessarily by the marketplace operator itself.

- Is the business model of Ocean Protocol associated with the marketplace itself, or does Ocean Protocol aims to develop an ecosystem of domain-specific data providers?

We are not trying to develop an ecosystem of domain-specific data providers. Instead, we are aiming to build an ecosystem of data service providers. So, either this is a marketplace, a service for data storage, a service of data computation, or anything connected to the data value chain. We are working to enable, integrate, and provide tools to deliver the data value chain using Ocean Protocol technology.

- What mechanism does ocean protocol use to match actors with available data to increase the usability of data? I mean, how does Ocean protocol shift from data as a product to data as a problem solution?

I would point to the staking mechanisms we have developed. These mechanisms place a stake on an asset (data or algorithms) used to signal the quality of this asset. So if you are a potential data consumer, you would have an indicator if others see the data you want to purchase as qualitative data or if others see it worth the price listed. For example, in Ocean Protocol, we have dynamic and static prices. The dynamic prices are determined by liquidity pools based on tokens and leveraging the staking aspect. This, in short, allows for real-time market pricing and enables the market to price an asset according to its demand. I would say that these are the tools based on staking mechanisms we have in place in Ocean Protocol. Currently, there are no active matchmakers or brokerage services that match data providers to data consumers in Ocean Protocol.

- In the context of peer-to-peer data exchange, do you think we shift the responsibility of problem solutions or the marketing aspects to the data providers?

In the sense that providers should market their data as being valuable?

- Yes

That is a relevant observation. As data is increasingly becoming an asset, there are also emerging data providers marketing their assets. Data providers aim to increase attention to their assets and increase the value of their assets. For example, in the Ocean ecosystem, data assets are attached to data tokens through liquidity pools. In this context, data providers are trying to gain attention for their data sets because this would also increase the number of individuals – that may not want to consume – but are interested in staking. So if they are able to attract attention for those individuals, then more people will stake, and the hope is that they will stake on qualitative data sets.

We had issues with bad actors on the open market where they are posting data assets looking to extract as much value from the assets as possible. These actors are not motivated to provide qualitative data but to increase the value of their data. So suddenly, they subtract the liquidity from the liquidity pool, leaving other liquidity providers out dry. We are working to prevent this, and we already addressed this issue in future releases to have more secure initial data offerings.

Back to the question, data providers are doing a lot to show the value of their data assets.

- If we try to touch a bit on the technical aspect, from your expertise, do you think we could mitigate some concerns of the data providers with a decentralized architecture?

I think that blockchain, at least from the ideological sense, places access control, rewards, and trust into individuals. There is a more general question, "is decentralization necessary for everything?". Sometimes there are different levels of decentralization that are useful. In the case of data sovereignty and data monetization, blockchain offers an interesting solution from my perspective. Particularly because it is transparent and traceable. Moreover, it is controllable in terms of "your keys, your bitcoin; your keys your data". Thus, an individual or organization has the necessary control and ability to determine how data is being used or been used. So, if you have offered data for usage, you can trace how it has been used, when, and by whom. In this context, I think that blockchain has a significant role in the data economy.

But, there are many data brokers and centralized data marketplaces, which are doing reasonably well. Although a lot of what centralized marketplaces have to deal with is overhead. So, you also mentioned all the detailed agreements between providers and consumers – this takes a lot of resources, time, and effort, sometimes slowing things considerably.

- It is clear that decentralization technically solves many problems in the data market. For example, data sovereignty or removal of the single point of failure in the infrastructure. But how does ocean protocol as a data marketplace builds the ecosystem for value creation in the data market? Furthermore, how does Ocen Protocol attracts more data providers in the ecosystem? Are there any clear steps?

Absolutely, we are continuously working on this. Unfortunately, I can reveal too much about it now. But we do have some programs that will be launched in the future to incentivize data providers and hopefully quality data providers.

- So the final purpose will be to build an ecosystem of data providers based on the infrastructure of Ocean Protocol?

I would not call an infrastructure of data providers but an ecosystem of data providers. We also want to incentivize service providers such as compute providers or storage providers to extend much more broadly from just data providers. In the same context, we are incentivizing the community to develop the Ocean Protocol through Ocean DAO. This curator fund grants individuals and teams according to the value they would like to bring to the Ocean Network.

Have you seen the Ocean DAO and looked at it?

- Yes, I did. From my perspective, Ocean DAO operated through the philosophy of open source projects. Isn't it?

That is what Ocean Protocol is at the end of the day – an open-source project. The funds initially raised to kick off the project were allocated to the core team. Now, the funds are allocated beyond the core team.

- Another observation in the research process is that most decentralized data marketplaces sustain their business models through token economy principles. These principles are based on (1) crowdfunding through initial coin offering, (2) strategies to raise the value of the digital assets, (3) increase the user base, and (4) continuously developing marketplace through applying some of the philosophies of the open-source projects. Doesn't the hype and trend of cryptocurrency cover this process? Or, in principle, could we deviate from cryptocurrency and evaluate the platform through the value of their digital assets? (Critical perception of the blockchain)

I would certainly opt for the latter, and I think your research also focuses on what value is created and delivered through decentralized marketplaces. Obviously, a part of the process has to be incentivized by those who travel down that path. They have to "keep the lights on" and survive so that their ecosystem can be self-sustainable.

To your first point, not all blockchain projects need to have a token. Obviously, it lends itself to tokenization. But you have to really assess either if you need a token in your environment.

Yes, you are right, there have been plenty of critical voices from the regulatory bodies concerned with the scams that came out. There are many scams promising some coin, and everyone buys it assuming that it is real or sounds great, then nothing happens in the end. But I would look at the serious projects and those who survive. These projects are really delivering something, and they are using tokens within the ecosystem to build value. Many of the projects are based on some utility. It is not just a storage of value or some currency that allows you to exchange value. These tokens are actually enabling you to perform certain actions. In the context of a marketplace setting, this is certainly relevant.

It is evident that decentralized marketplaces use tokens and are associated with cryptocurrency, which is seen as a dark market or darknet. That is a stigma that is difficult to break. But those who understood the value that blockchain could bring to digital ecosystems are certainly looking at this technology differently. Blockchain is being looked at and used in projects for private or public entities, and it is not necessary to have a token for all of these projects. So I think that this stigma of blockchain and cryptocurrency is starting to be reassessed.