Data Collaboratives

Trusted Data Intermediary Business Models

Delft University of Technology

TU Delft

Maartje Flipsen

Data Collaboratives

Trusted Data Intermediary Business Models

by

Maartje Flipsen

to obtain the degree of Master of Science

at the Delft University of Technology,

to be defended publicly on Friday August 9th, 2019 at 11:00 AM.

Student number:	4206444	
Project duration:	February 1, 2019 – August 9th, 2019	
Thesis committee:	Dr.ir. G.A. de Reuver,	TU Delft, first supervisor
	Dr. S. Cunningham,	TU Delft, second supervisor
	Dr. I. Susha,	Örebro University, additional supervisor

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft

Preface

You're braver than you believe, stronger than you seem, and smarter than you think

Winnie the Pooh

With this thesis I conclude my education in Engineering & Policy Analysis at Delft University of Technology. What started out as just a Master Program has finished to be two life-changing years. I am grateful to my friends, colleagues, teachers and family who helped me deal with every curve ball in life and who supported me in achieving my goal of graduating in nominal time. I did not expect to meet the amazing people that I've met. I did not expect to learn as much as I've learnt. And, I did not expect to achieve the things that I achieved. Most of all, I did not expect to believe what a silly old bear told me a very long time ago. I am celebrating this moment without my two biggest supporters. As I could not have reached this point without their faith in me, I dedicate this thesis to *Papa* and *Buuf*.

Maartje Flipsen
Delft, July 2019

Executive Summary

To address complex societal issues, cross-sector partnerships are needed that specifically aim to create value to address such challenges. Data collaboratives are initiatives that encourage and promote such partnerships, by the collection, sharing, or processing of data. In a data collaborative, data transactions occur between data contributors and data users. This transaction creates value for respectively the data contributor and data user, as well as value that can be leveraged in addressing societal issues. However, data collaboratives are faced with barriers that complicate collaboration between data contributors and data users. As a result, the potential of public value creation is not being reached. To overcome these barriers, decision-makers in data collaboratives need a better understanding of the Trusted Data Intermediary as a coordination mechanism.

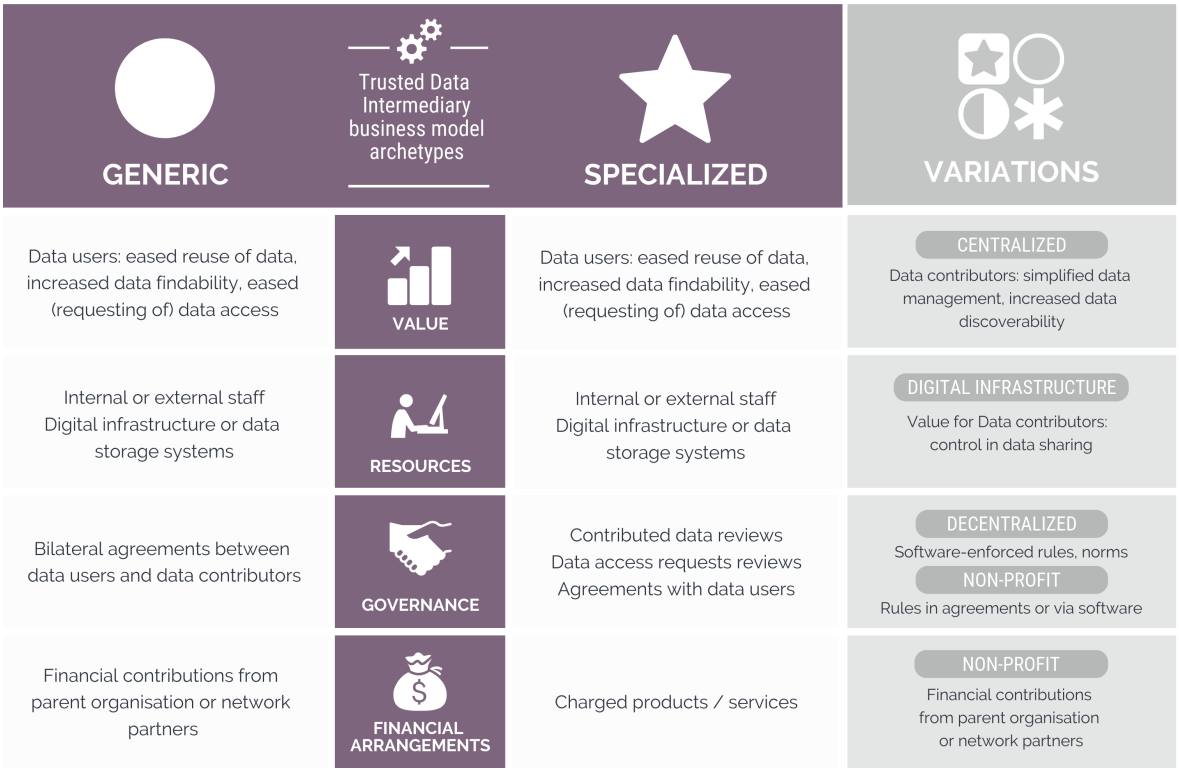
Trusted Data Intermediaries are entities entrusted with enabling data transactions between data contributors and data users across sectors, thereby generating value for collaborators, and enabling the creation of public value. Currently, academic knowledge on how these intermediary entities is absent. This thesis addresses this knowledge gap by researching Trusted Data Intermediaries' business models. Business models are abstract representations how Trusted Data Intermediaries operate. The models describe what kinds of value are created for resp. data contributors and data users, what resources are utilized, which governance mechanisms are leveraged in coordination of the network and what financial arrangements are in place. By identifying Trusted Data Intermediaries' business model archetypes, this thesis contributes to a better understanding of Trusted Data Intermediaries.

Following a systematic literature review, a theoretical framework is developed. This framework serves as the basis for describing the business models of Trusted Data Intermediaries. These business models concern the value proposition, architecture, network and finance dimensions. By a qualitative, exploratory multiple-case study, six descriptive business models of Trusted Data Intermediaries are developed. These business models are contrasted, and the findings are discussed in relation to Trusted Data Intermediary characteristics in specialization, profit motive and centrality of data storage.

From the discussion of findings in relation to characteristics, relations are proposed that form the basis of the proposed business model archetypes. Two archetypes are offered: the **Generic Trusted Data Intermediary business model archetype** and the **Specialized Trusted Data Intermediary business model archetype**. Both archetypes create value for data users in the form of eased reuse of data, and increased data findability. Furthermore, both archetypes utilize technical and (substitute) human resources. Agreements are used to coordinate the network of data contributors, data users and network partners. Depend-

ing on the specialization of the Trusted Data Intermediary to a specific customer segment, additional kinds of value may be created for data users in the form of eased (requesting of) data access. In addition, Specialized Trusted Data Intermediaries leverage review processes for data access requests and contributed data. While, Generic Trusted Data Intermediaries specifically leverage bilateral agreements between data users and data contributors. Furthermore, the choice for a centralized data storage system can create value for data contributors in the form of simplified data management and increased data discoverability. In contrast, a decentralized system with software-enforced rules may create trust, communication and information sharing. Last, a non-profit based Trusted Data Intermediary will leverage rules to coordinate the network and receive funding from its network.

By describing two archetypes, this thesis has made a contribution to both the academic field and practice. For the academic field, an initial academic theory is developed on a new type of intermediary. For decision-makers in data collaboratives, this thesis demonstrates the role that Trusted Data Intermediaries can play in abating barriers for collaboration between data contributors and data users. As next steps, future academic endeavours may test the developed theory on Trusted Data Intermediary business models, adapt the research design to accommodate quantitative analysis, and research other characteristics. In practice, decision-makers in Trusted Data Intermediaries can make use of the theoretical framework and archetypes to reflect on their own business model. A thorough understanding of the business model will help in communicating how data contributors and data users can derive value from making data transaction. As a result, new cross-sector partnerships may form that address grand societal challenges.



Trusted Data Intermediary business model archetypes.

Contents

List of Figures

List of Tables

1

Introduction

Society is challenged by grand, utter complex issues of 'wicked' nature which are almost impossible to resolve (Rittel and Webber, 1973). In response of these issues, the United Nations Member states adopted the 2030 Agenda for Sustainable Development. The seventeenth Sustainable Development Goal (SDG) focuses on strengthening "the means of implementation" and revitalizing "the global partnership for sustainable development" (General Assembly Resolution 70/1, 2015, p. 14). A part of this resolution is the encouragement and promotion of effective public-private and civil society partnerships that mobilize and disseminate knowledge, expertise and technology to advance the achievement of the SDGs (General Assembly Resolution 70/1, 2015). This thesis discusses initiatives that encourage and promote cross-sector partnerships that specifically aim to create value that addresses a complex societal issue, by the collection, sharing, or processing of data (Susha, Janssen, and Verhulst, 2017a). To describe these initiatives, the concept *data collaboratives* is most commonly applied (Susha, Grönlund, and Van Tulder, 2019). As such, this term is adopted in this thesis.

1.1. Data collaboratives

Data collaboratives are "cross-sector (and public-private) collaboration initiatives aimed at data collection, sharing, or processing for the purpose of addressing a societal challenge" (Susha, Janssen, and Verhulst, 2017b, p. 2691). For instance, using private sector search engine data to complement official public sector statistics on suicide rates in India (Adler, Cattuto, Kalimeri, Paolotti, Tizzoni, Verhulst, Yom-Tov, and Young, 2019). Following Susha et al. (2017a), data collaboratives can be perceived as "bazaars", a market form where entities exchange commodities on the basis of supply and demand. Henceforth, the entity that supplies data will be referred to as *data contributor*, and the entity that demands data will be referred to as *data user*.

Data collaboratives are challenged by barriers that impede on collaboration by data contributors and data users (Susha et al., 2019). Data collaboratives and similar concepts are mostly challenged by a lack of consistent and comprehensive legal provisions, lacking or misaligned incentives, low uptake of data providers, lack of financing, lacking discoverability and (low-cost) accessibility of data, and privacy issues (Susha et al., 2019). Current academic literature on data collaboratives identifies intermediaries as a coordination mechanism to overcome barriers in matching data contribution and data use (Susha et al., 2017a). This thesis will focus on such entities that can function as coordination mechanisms to overcome barriers in data collaboratives. This thesis stipulates them as *Trusted Data Intermediaries*.

1.2. Trusted Data Intermediaries

The working definition for Trusted Data Intermediaries in this thesis is as follows: "Trusted Data Intermediaries are entities entrusted with enabling data transactions between data contributors and data users across sectors, thereby generating value for collaborators, and enabling the creation of public value". A tentative example of value for a data contributor can be increased discoverability of their available data. For data users, value may come from increased accessibility of data owned by others. Similar, the Trusted Data Intermediary can create value by eliminating or lowering other barriers faced by data collaboratives. Further, the Trusted Data Intermediary can enable public value creation by reducing barriers for data contributors and data users, thereby enabling them to participate in or form a data collaborative. Tentative examples of public value may be knowledge creation and transfer, design and delivery of public services, situational awareness and response, impact assessment and evaluation, or prediction and forecasting (Gasco-Hernandez, Feng, and Gil-Garcia, 2018).

In enabling participation of data contributors and data users in data collaboratives, it is relevant that Trusted Data Intermediaries understand and communicate what value they offer. Research shows that private organisations are more inclined to contribute data once they see how collaboration creates value for them (Robin, Klein, and Jutting, 2016; Sayogo, Zhang, Pardo, Tayi, Hrdinova, Andersen, and Luna-Reyes, 2014). While research has indicated that Trusted Data Intermediaries can function as a coordination mechanism for data collaboratives, knowledge on how they operate is lacking. Among others, it is unknown what resources are utilized, what mechanisms are leveraged to coordinate network actors, and what financial arrangements are for Trusted Data Intermediaries. In addition, it is unclear whether all Trusted Data Intermediaries operate the same, or whether this can differ. These gaps can be addressed by business models.

1.2.1. Knowledge gap: Trusted Data Intermediaries' business models

A business model depicts how the content, governance and structure of transactions create value (Amit and Zott, 2001). The content of the transaction relates to what is being exchanged, and what resources are required. The structure refers to the network actors

that participate in the exchange. Governance refers to how parties control information flows, resources and goods (Amit and Zott, 2001). As such, business models provide ways to understand, analyze, communicate and manage strategic choices among stakeholders (Al-Debei and Avison, 2010). Although many definitions for business models exist (Al-Debei and Avison, 2010; El Sawy and Pereira, 2013a; Zott, Amit, and Massa, 2011), this thesis opts for the definition of the business model as "an abstract representation of an organisation, be it conceptual, textual, and/or graphical, of all core interrelated architectural, co-operational, and financial arrangements designed and developed by an organisation presently and in the future, as well all core products and/or services the organisation offers, or will offer, based on these arrangements that are needed to achieve its strategic goals and objectives." (Al-Debei and Avison, 2010, p. 372). This definition was formed from 22 business model conceptualizations found by content analysis, followed by a comprehensive literature review. As such, it integrates multiple facets of business models' conceptualizations. Although this definition emphasizes the value logic of for-profit organisations, business models can pertain to companies, organizations or persons (Nielsen and Lund, 2015), as well as non-profit entities (Janssen and Zuiderwijk, 2014). The business model perspective has been long adopted by the domains of eBusiness, eCommerce, eGovernment, as well as business management and strategy, the telecom sector and the software-industry (Al-Debei and Avison, 2010). Yet, business models have been lacking from data collaboratives literature.

Business models are relevant for gaining a better understanding of how Trusted Data Intermediaries operate and create value for collaborators. As such, knowledge on Trusted Data Intermediary business models can illuminate their role as a coordination mechanism for data collaboratives. Currently, there is no academic research available that investigates Trusted Data Intermediaries or similar concepts. As such, knowledge on Trusted Data Intermediary business models is absent.

1.2.2. Research objective: Trusted Data Intermediary business model archetypes

The objective of this thesis is to create a better understanding of the Trusted Data Intermediary as a coordination mechanism in data collaboratives, by describing Trusted Data Intermediary business model archetypes. Hence, this thesis will address the following research question: *What are business models for value creation by Trusted Data Intermediaries for resp. data contributors and data users.*

In the context of the working definition of the Trusted Data Intermediary, the business models for data transactions need to be described. The setting in which these intermediaries operate can be described as virtual or electronic markets (Amit and Zott, 2001), because transactions can be conducted over open networks based on an infrastructure of fixed and wireless Internet. These markets are characterized by high levels of reach and connectivity, as well as low-cost information processing power. In such a setting, business model components comprise of what is exchanged, what resources are required to do so, which network partners are involved and how are information flows, resources and goods controlled (Amit

and Zott, 2001). In relation to the adopted definition of business models by Al-Debei and Avison (2010), the business model should describe what kinds of value is offered and the resource, network and financial arrangements that allow for the provision of value. While Al-Debei and Avison (2010) offer an extensive conceptualization for these dimensions, for the purposes of this thesis it was chosen to focus on each dimension's core component for two reasons. Foremost, the thesis is limited in time and resources. Furthermore, academic literature on Trusted Data Intermediaries or similar concepts is absent. These constraints in mind, it is chosen to focus the research on specific parts of the dimensions. To begin, the business model should describe what kinds of value is offered. Creating knowledge on this matter can help position Trusted Data Intermediaries in literature as a separate type of intermediary entity. Furthermore, identifying and communicating the created value for data contributors and data users, which could stimulate uptake. Other questions relate to the arrangements that allow for the provision of value. According to Amit and Zott (2001), a business model depicts the design of transaction content, structure, and governance so as to create value through the exploitation of business opportunities. As such, the second questions focuses on the key resources utilized by Trusted Data Intermediaries. For instance, are these mostly technical resources, given the commodity being data? Or, does this differ per Trusted Data Intermediary? Third, given the positioning of Trusted Data Intermediaries as a coordination mechanisms for data collaboratives, what are the governance mechanisms leveraged to coordinate the collaboration? Tentative examples of governance mechanisms are rules, audits, agreements and many more. Last, what are the financial arrangements? For instance, are Trusted Data Intermediaries non-profit entities? Do they charge fees for the provision of services? Table 1.1 displays the chosen research questions of this thesis.

Table 1.1: Research questions.

Research question	
Main research question	What are business models for value creation by Trusted Data Intermediaries for resp. data contributors and data users?
Sub question 1	What kinds of value do Trusted Data Intermediaries offer data contributors and data users?
Sub question 2	What are key resources for value creation by Trusted Data Intermediaries?
Sub question 3	What are key governance mechanisms for value creation by Trusted Data Intermediaries?
Sub question 4	What are financial arrangements of Trusted Data Intermediaries?

The next chapter discusses the state of the art on Trusted Data Intermediaries and their business models. The theoretical framework in chapter 3 is applied in an exploratory case study of six cases that represent Trusted Data Intermediaries. For each case, information is collected and qualitatively analyzed. Chapter 5 describes the business model of each case. Subsequently, the business models are compared and discussed in light of literature (chapter 6). The differences and similarities are explained by the means of case characteristics. Based on this discussion, two archetypes for Trusted Data Intermediary business models are proposed. This thesis concludes with recommendations to the academic field and practice.

2

State of the Art

In this chapter, the state of the art of literature on Trusted Data Intermediaries is laid out. Section 2.1 explores concepts for intermediary organisations for the domain of data collaboratives. From this exploration, the working definition for such intermediaries is argued for. Next, in section 1.2.1 the state of the art on the intermediary business models, and the various kinds of value, resources, governance mechanisms and financial arrangements of intermediaries is discussed.

To ensure the accumulation of the most relevant literature, a systematic approach as described by Webster and Watson (2002) was applied. First, the most relevant sources were consulted for keywords related to the topic. Due to the multidisciplinary nature of data collaboratives, the literature search focused on journals in the disciplines *business research*, *social and decision sciences*, and *ICT*. The literature was also limited to peer-reviewed and fully published articles. The databases linked to the TU Delft Library were used to search for keywords. Among others, the library is connected to databases like Web of Science, JSTOR, ABI/INFORM, IEEE, ScienceDirect, SpringerLink, WorldCat and Wiley Online Library. In addition, Scopus was searched for keywords. Unfortunately, not all results shown in the databases were accessible, which limited the amount of results yielded from the searches. To collect information on the context of this thesis' topic, the keywords "data collaboratives" and "trusted data intermediaries" were entered first. The search for "data collaboratives" yielded five results, which were mostly from the same author(s) and institution. For the keyword "trusted data intermediaries" no academic literature exists that uses this term. The term "data intermediaries" does yield results. From this list of 50 results, 5 are deemed relevant for the literature review based on their application of intermediaries. However, this literature all refer to the open (government) data domain. Next, literature was found by consulting the reference lists of the yielded articles. From the articles that were found, the concepts "trusted third party", "electronic intermediaries", "partnership brokers" and "information intermedi-

aries" came forth. The related references were consulted and subsequently, the databases were searched for these concepts as well. Of these concepts, "electronic intermediaries" and "information intermediaries" yielded the most results applicable to the type of intermediaries discussed in this thesis. Another point of interest to accumulate literature for is the concept of "business models". The amount of resultant literature for this term is overwhelming, so to limit these results I looked at the most cited literature and only adopt literature regarding the business model concept. When limiting the search concept to "business models and intermediaries", the search yields around 1000 results. Of these results, much literature is sector-specific (e.g. tourism, insurance, banking, country-specific). As a result, four articles were selected. Lastly, iterations were made in this approach, where synonyms were searched, and Google Scholar was accessed as well. Furthermore, the list of literature has been complemented with suggested literature by thesis supervisors. Table 2.1 provides an overview of the primary searched keywords, number of results in each database and the findings from going through these results.

Table 2.1: Literature search results from databases.

Keywords	**TULib**	**Scopus**	**Findings**
"Data collaboratives"	5 results	96 results	Literature from same authors and institution (e.g., Adler et al., 2019; Klievink et al., 2018; Susha et al., 2017a,1,1)
"Trusted data intermediaries"	0 results	0 results	No academic literature found
"Data intermediaries"	28 results	22 results	Much literature focuses on open data domain: (e.g., da Silva Craveiro and Albano, 2017; Magalhaes et al., 2013; Mercado-lara and Gil-garcia, 2014; van Schalkwyk et al., 2016; Yoon et al., 2019)
"Trusted third party"	846 results	300 results	Academic literature focuses on cryptography, which is out of scope
"Electronic intermediaries"	96 results	64 results	Some articles were inaccessible, other results were out of scope, except for (Bailey and Bakos, 1997; Janssen and Verbraeck, 2005; Muylle and Basu, 2008)
"Partnership brokers"	3 results	1 result	Resultant literature out of scope: impact of partnerships or marketing, education
"Information intermediary"	187 results	203 results	Much literature is on domains out of scope, except for (Barghava and Choudhary, 2004; Womack, 2002)
"Business models"	14.000 results	12,541 results	Overwhelming amount of literature, so most cited literature relevant to research scope was assessed (e.g., Al-Debei and Avison, 2010; Amit and Zott, 2001; Teece, 2010; Zott et al., 2011)
"Business models" + "Intermediaries"	461 results	495 results	Limited literature on intermediaries that is not sector-specific (i.e., Bouwman et al., 2008; Janssen and Zuiderwijk, 2014; Ranerup et al., 2016; Täuscher and Laudien, 2018).

2.1. Intermediaries for data collaboratives

A clear definition for intermediaries in data collaboratives is lacking. Nevertheless, various concepts that describe (facets of) intermediaries in a similar context exist. This section identifies intermediary concepts in virtual or electronic markets and discusses the definitions offered by literature. The section concludes with a justification of the chosen working definition for intermediaries in data collaboratives, stipulated as Trusted Data Intermediaries. An overview of the concepts is provided in Table 2.2 at the end of this section.

2.1.1. Concepts and definitions for intermediaries

Recently, a non-academic study from the Digital Civil Society Lab at the Stanford Center on Philanthropy and Civil Society proposed the term "Trusted Data Intermediaries". The term was created to describe non-profit organisations that collect, aggregate, open up large private data sets for public purposes by negotiating the relationships between data users and data contributors (Digital Civil Society, 2017). The term has not yet been applied in the academic field, as a systematic literature search for "Trusted Data Intermediaries" yielded no results. As such, the state of the art on Trusted Data Intermediaries is dictated by a research paper that provides a summary of a workshop held in 2016. The contents of the research paper were published to the Digital Civil Society's blog on the Medium platform (Digital Civil Society, 2017). In the workshop, scholars, practitioners and leaders of Trusted Data Intermediary-like enterprises convened to identify commonalities and distinctions for Trusted Data Intermediaries. A small group of organisations was studied, including Artstor, LearnSphere, NPC Data Labs, and the Mastercard Center for Inclusive Growth. From this study, it was derived that Trusted Data Intermediaries can add value by making timely data available, or by storing data for the long term. In addition, Trusted Data Intermediaries can offer security and performance. Furthermore, data contributors may provide data aggregations, which results in different layers of aggregated data offered by the Trusted Data Intermediary. It is the different levels of aggregated data and inter-mediation between parties that offers a new dimension of value. On the contrary, the workshop also revealed distinctions between Trusted Data Intermediaries. For example, Trusted Data Intermediaries vary in openness, since some Trusted Data Intermediaries emphasize control and limited access rather than an open platform. Further, data ownership may vary between Trusted Data Intermediaries, i.e. contributors can be individuals or organisations. Similarly, the data subjects may range from organisational data to individual (personal) data (Digital Civil Society, 2017). As well, some Trusted Data Intermediaries may generate revenues, although this is not the case for all.

Although an academic definition for intermediaries in data collaboratives is absent, various other concepts have been researched. To begin, the concept of the *trusted third party* (abbr. TTP) in electronic commerce was studied by Froomkin in 1996. As he argues, much electronic commerce require a trusted third party in the bilateral relationship between buyer and seller, who identifies, authenticates and vouches for the parties in the transaction. A year after, Skevington and Hart (1997) build upon this concept. The authors argue that trust is an essential part in any transaction. TTPs help in the facilitation of trust between parties that wish to conduct the transaction, by acting as an intermediary in the transaction or by providing some service. Well known examples of TTPs are banks, credit card companies and financial advisors. Electronic commerce requires the development of new TTP mechanisms (e.g. cryptography and digital signatures) and TTP services (e.g. authentication services and verifying the existence of digital documents). Thus, in electronic commerce, TTPs assume new roles and responsibilities: authentication of parties' identities (1), checking of parties' credentials (2), guarantee

integrity and confidentiality of messaging between parties (3) and settlement of disputes (4) (Skevington and Hart, 1997).

The concept *electronic intermediaries* was introduced by Bailey and Bakos in 1997. This concept explicates intermediaries in electronic markets: markets where suppliers provide goods and services to customers in transactions that are partially or fully automated by information technology. Furthermore, the concept goes beyond the role of establishing trust in transactions. The authors found that electronic intermediaries have roles as: aggregator of information (1), facilitator of the market (2), matchmaker between buyers and sellers (3), and being a trusted agent (4). Barghava and Choudhary (2004) build upon the roles identified by Bailey and Bakos (1997). They found that electronic intermediaries, in particular those in the role of aggregator of information, provide *matching services* and *value-added services*. Matching services facilitate the establishment of a buyer-seller agreement, for instance through listing, price discovery and matching. Value-added services provide additional value, for example through workflow coordination, transaction or account management, industry reporting and consultation (Barghava and Choudhary, 2004). Weill and Vitale (2001) also discuss the primary function of electronic intermediaries to be bringing together buyers and sellers and concentrating information. Further, Weill and Vitale (2001) describe electronic intermediaries as six types: electronic markets and auctions, specialty auctions, portals, shopping agents and electronic mall. Muylle and Basu (2008) developed a typology for electronic intermediaries (EIMs). They argue that, depending on their role, such intermediaries support trade processes (e.g. search, authentication, valuation, payment and logistics), decision support processes (e.g. configuration, collaboration and business intelligence), and integration processes (e.g. data integration and application integration). However, this research excluded government-driven EIMs or EIMs that operate by a non-profit motive.

While Barghava and Choudhary (2004) focus on the intermediary role of aggregator, the concept of *partnership brokers* emphasizes the role of intermediaries as matchmaker between organisations. According to Gould and Fernandez (1989), partnership brokers can assume five social roles in systems of exchange or networks of resource flows: coordinator (1), itinerant broker (2), gatekeeper (3), representative (4) or liaison (5). In the first, third and fourth type of brokerage, the broker organisation takes part in the same group of at least one partner organisation. While, in the second and fifth type, the broker organisation belongs to a different group than the partner organisations (itinerant broker), or all organisations are from distinct groups (i.e. the broker organisation is a liaison) (Gould and Fernandez, 1989).

Analogous to Gould and Fernandez (1989), Stadtler and Probst (2012) study brokers of the liaison type in public-private partnerships. The authors define brokers of public-private partnerships (PPPs) as “organizations that have specific experience and capacity to build and/or facilitate PPPs” (Stadtler and Probst, 2012, p. 32). Three roles of PPP brokers are identified: convener (1), mediator (2) and learning catalyst (3). In the role of convener, brokers must invest in public-private partnerships continuously. Limited resources and capacities can impede on this continuous investment, so the broker must take a focused approach to bridging relationships between partners. In the role of mediator, a broker can motivate partners to develop “bonding social capital”. Social capital is described as a set of resources that encompass trust, norms and values. Social capital is accessed through a network of social relations. Strong social ties from strong relationships and norms that create trust, communication, and information sharing, is described as bonding social capital. The broker’s mediator role can be challenged by a lack of willingness and internal readiness of partner organisations. This requires the broker to balance a pro-active attitude with leaving the partners with room for autonomous decision-making. In any case, continuous investment in the development of bonding social capital and communication with partners

is a pre-requisite for broker organisations. Finally, in the role of learning catalyst, the broker offers three modes of learning. First, by sharing its knowledge pool, the broker can offer passive learning possibilities to partner organisations. Second, a broker can offer active learning sessions in the forms of workshops or training. Last, interactive learning can be offered through fostering discussion and continuous communication between partners. In this sense, the broker acts more as a moderator that challenges the partners to seek opportunities to benefit from the partnerships (Stadtler and Probst, 2012).

Contrary to the concept of partnership brokers, the concept of *information intermediaries* emphasizes the role of intermediaries as aggregators of information. Womack defines an information intermediary (abbr. infomediaries) as "any system that mediates between the producers and consumers of information" (Womack, 2002, p. 133) by collecting information and disseminating it to a defined client group. Which information is selected, and how this is organized and distributed depends on the needs of the clients. The services offered to the client need to be of added value for the client to choose the intermediary's service over the option of collecting information directly from the source. This added value may be that information is of higher quality, more complete, more easily accessible, better organized, cheaper or other factors (Womack, 2002). Contrary to Womack (2002), Durrance, Walker, Souden, and Fisher (2007) describe information intermediaries as organized local groups in the context of communities. They specify the concept to *problem-centered information intermediaries*, defining them as "organized groups and non-profits whose primary function and mission emanated from a particular problem and whose secondary function became to provide relevant, targeted information" (Durrance et al., 2007, p. 8).

Alternatively, Hagel and Rayport (as cited in Janssen and Zuiderwijk (2014), p. 695) define *infomediaries* as "a broker or agent who helps users to manage the vast amount of information and safeguard their privacy". Alternatively, *infomediaries* are seen as providers of products and services based on public sector data (Magalhaes et al., 2013). As such, Magalhaes et al. (2013) categorize infomediaries as open government data intermediaries. Janssen and Zuiderwijk (2014) researched this category of infomediaries and positions them between data providers and data users. According to Janssen and Zuiderwijk (2014), infomediaries are characterized by specialization, economies of scale and economies of scope. Each of these characteristics can make other organisations dependent to the infomediary (Janssen and Zuiderwijk, 2014). For instance, an infomediary may only handle data for a very specific topic (i.e. highly specialized). Or, an infomediary may be a data repository linked to a variety of data sources, making it useful to large customer segments (i.e. economy of scope). Similar, an infomediary may be an application that enables the offering of multiple products and services with the same resources (i.e. economies of scale).

Similar to Janssen and Zuiderwijk (2014) and Magalhaes et al. (2013), *Multi-Sided Platform (MSP)* are public-private platforms positioned between two or more groups of agents (Klievink, Bharosa, and Tan, 2016). Such a platform creates value by enabling direct interactions between different types of customers. As such, the platform enables collaborative value creation (Klievink et al., 2016).

Other researches on intermediaries in the open data domain by van Schalkwyk et al. (2016) and Mercado-lara and Gil-garcia (2014) refer to intermediaries as *data intermediaries*. According to Mercado-lara and Gil-garcia (2014), data intermediaries abate barriers for data users by managing big and complex data sets and presenting them in a user-friendly way that is indifferent to the user's skill. van Schalkwyk et al. (2016, p. 77) define data intermediaries as keystone species that create "value in ecosystems by creating platforms, services, tools or technologies that offer solutions to other actors in

the ecosystem". This conceptualization of an entity that offers services through platforms is similar to that used by Janssen and Zuiderwijk (2014) and Klievink et al. (2016). In other research on data intermediaries for communities, data intermediaries are broadly defined as organizations that facilitate (re)use of data to impact communities (Yoon et al., 2019).

Table 2.2: Concepts and definitions for intermediaries.

Concept	Definition
Trusted Data Intermediary	Nonprofit organisation that collects, aggregates, opens up large private data sets for public purposes by negotiating the relationships between data users and data contributors (Digital Civil Society, 2017)
Trusted Third Party	A third party to a bilateral relationship (Froomkin, 1996) that help in the facilitation of trust between the other two parties that wish to conduct a transaction, by acting as intermediary in the transaction or by providing some service (Skevington and Hart, 1997)
Electronic Intermediary	Intermediaries in markets where suppliers provide goods and services to customers through transactions that are partially or fully automated by information technology (Bailey and Bakos, 1997)
Public-private partnership broker	Organizations that have specific experience and capacity to build and/or facilitate public-private partnerships (Stadtler and Probst, 2012)
Information Intermediary	Any system that mediates between the producers and consumers of information by collecting information and disseminating it to a defined client group (Womack, 2002)
Problem-centered information intermediaries	"Organized groups and non-profits whose primary function and mission emanated from a particular problem and whose secondary function became to provide relevant, targeted information" (Durrance et al., 2007, p. 8)
Infomediary	A broker or an agent that helps users manage vast amounts of information while safeguarding their privacy (Hagel and Rayport as cited in Janssen and Zuiderwijk, 2014). Or, providers of products and services based on public sector data (Magalhaes et al., 2013)
Multi-Sided Platform (MSP)	A public-private platform that enables direct interactions between different types of customers (Klievink et al., 2016)
Data Intermediary	Keystone species in the ecosystem that create value by creating platforms, services, tools or technologies that offer solutions to other actors in the ecosystem (van Schalkwyk et al., 2016). Or, an organization that facilitates (re)use of data to impact communities (Yoon et al., 2019).

2.1.2. Working definition for Trusted Data Intermediaries

From Table 2.2 it can be concluded that there are many conceptualizations and definitions for intermediaries. Although academic literature provides at least eight concepts and definitions for intermediaries in contexts similar to that of data collaboratives, none truly captures the core elements of data collaboratives. Firstly, data collaboratives are characterized by cross-sector collaboration (Susha et al., 2017b). This characterization lacks in the concepts *Trusted Third Party*, *electronic intermediary*, *information intermediary*, *infomediary*, and *data intermediary*. Second, data collaboratives are aimed at collecting, sharing and processing data (Susha et al., 2017a). This dimension is not addressed by the concepts *Trusted Third Party*, *Multi-Sided Platforms* and *public-private partnership broker*, nor by the concepts *electronic intermediary*, or *data intermediary*. Consequently, the academic concepts do not sufficiently cover the dimensions that an intermediary in a data collaborative encounters. In contrast, the non-academic concept of *Trusted Data Intermediaries* is more suitable. This concept addresses both the dimension of public-private collaboration and the dimension of data collection, sharing and process-

ing. However, this definition is centered on non-profit organisations. This stipulation of intermediaries would potentially exclude cases from this study, where the intermediary is a platform, individual, consortium et cetera. Further, the definition restricts the activities of the intermediary to the negotiation of relationships. Thereby, other potential roles of intermediaries (e.g. facilitator, aggregator, learning catalyst) are excluded.

Rather, a definition for Trusted Data Intermediaries needs to account for all forms of intermediaries, e.g. an individual, organisation, consortium, platform et cetera. Therefore, Trusted Data Intermediaries are entities rather than organisations. Further, much literature discusses trust as a dimension of intermediaries (e.g., Bailey and Bakos, 1997; Digital Civil Society, 2017; Froomkin, 1996; Skevington and Hart, 1997; Stadtler and Probst, 2012; van Schalkwyk et al., 2016; Womack, 2002), therefore this should be included in the definition. Furthermore, the variety of roles that intermediaries can assume should be accommodated by the definition. In conclusion, the definition for Trusted Data Intermediaries needs adjusting to make it work for intermediaries in data collaboratives. The definition answers the questions what a Trusted Data Intermediary is, what they do, why they do it and how they do it. The following working definition will be used throughout this thesis to stipulate intermediaries in data collaboratives: *Trusted Data Intermediaries are entities entrusted with enabling data transactions between data contributors and data users across sectors, thereby generating value for collaborators, and enabling the creation of public value.*

2.2. Business models of intermediaries

The objective of this thesis is to describe business model archetypes of Trusted Data Intermediaries. To do so, the answers of Table 1.1 demand the explication of the state of the art on intermediaries' business models, value creation for data contributors and data users, resources, governance mechanisms and financial arrangements of Trusted Data Intermediaries. Business models can be used to explicate the content, structure, and governance of transactions for value creation (Amit and Zott, 2001) by an entity, either operating on a for-profit or non-profit basis (as shown by Janssen and Zuiderwijk (2014); Nielsen and Lund (2015)). Business models for electronic markets can be referred to as e-business models (Amit and Zott, 2001). Literature on these business models highlight the notions of value, architectural aspects and financial aspects Zott et al. (2011). The following sections will address the state of the art on intermediary business models, and the separate aspects that may constitute an intermediary's business model.

2.2.1. Intermediary business models

The literature found in the search for "business models and intermediaries" (i.e., Bouwman et al., 2008; Janssen and Zuiderwijk, 2014; Ranerup et al., 2016) all cite the work by Weill and Vitale (2001). Weill and Vitale (2001) discuss the "generic business model" of *electronic intermediaries* (see Figure 2.1). According to the authors, the most important services offered in this business model are knowledge management, policy enforcement, centralized management of applications, information systems, information search services, storage, security. These services offer buyers and sellers lowered search and transaction costs. Buyers and sellers can come from any customer segment, e.g. businesses and consumers or other entities.

In a similar approach, Janssen and Zuiderwijk (2014) explicate *infomediary business models*. Six

atomic business models for infomediaries are formulated. The first business model is based on single-source open data applications (single-purpose apps). In the second business model, data users can contribute content themselves (interactive apps). Third, published open data may be collected, combined and processed (information aggregators). If this is done for the purpose of comparison, the fourth business model is found (comparison models). Fifth, a government may publish their data online (open data repository). Last, published data may be combined and complemented with features for data searching, processing and analysis (service platforms). The six business models differentiate along the levels of data accessibility and dialogue between data users and data contributors (Janssen and Zuiderwijk, 2014).

While applying the same research design, Ranerup et al. (2016) did not find the classification of business models by Janssen and Zuiderwijk (2014) for Public Service Platforms in quasi-markets. Rather, business models are classified along a traditional versus emergent view. In the traditional view, public services may be perceived as common pool resources. Contrary in the emergent view, public services are positioned in a “free market”.

Most recently, intermediaries as marketplaces enabled by digital technologies were studied by Täuscher and Laudien (2018). The authors developed a taxonomy of six types of web-based platform business models (see Figure 2.2).

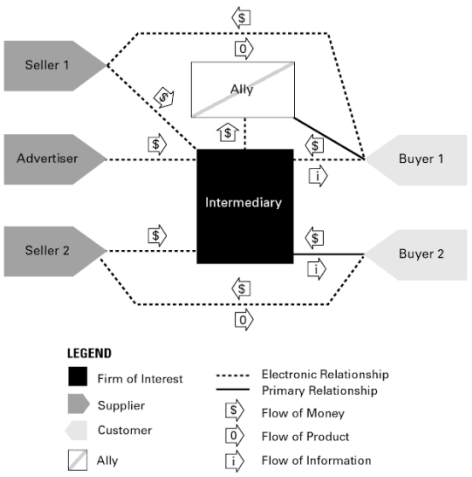


Figure 2.1: Generic business model of an electronic intermediary (Weill and Vitale, 2001).

Business model type	Platform type	Platform participants	Value proposition	Transaction type & good	Revenue model	Example
Efficient product transactions	Web-based platform	C2C, B2B	Large product variety	Physical products	Commission fee, subscription fee; mostly supply side	*Beepi* (eBay for used cars)
Digital product community	Web-based platform	C2C	Being part of a primary non-commercial community	Digital products; physical products	Commission fee; supply side, demand side	*Sellfy* (connecting ‘neighbors’ to share durable goods with each other)
Product aficionados	Web-based & mobile platform	B2C, C2C	Exchanging knowledge about niche products with community of like-minded people	Physical products	Commission fee; supply side	*HobbyDB* (Knowledge database and transaction platform for collectibles)
On-demand offline services	Web-based & mobile platform	B2C	Large service variety in a novel form	Offline services	Commission fee, subscription fee; mostly supply side	*StyleSeat* (connecting beauty salons & consumers)
Online services	Web-based platform	C2C, B2C	Novel online services with social networking character	Online services	Commission fee, subscription fee; mostly supply side	*iTalki* (connecting language learners with teachers for 1-on-1 online lessons)
Peer-to-peer offline services	Web-based & mobile platform	C2C	Novel services with community feeling within & outside digital platform	Offline services	Commission fee (demand & supply side), subscription fee (third parties)	*Airbnb* (connecting people to list, discover, and book private accommodations)

Figure 2.2: Business models of electronic marketplaces (Täuscher and Laudien, 2018).

Based on the information above, I argue that literature on intermediary business models is scarce. Moreover, knowledge on business models for Trusted Data Intermediaries is absent. While the business model proposed by Weill and Vitale (2001) is highly generic and accommodates different types of electronic intermediaries, the model presupposes that the intermediary owns the customer data and operates on a for-profit basis (Weill and Vitale, 2001). For Trusted Data Intermediaries, both assumptions are not always true (Digital Civil Society, 2017). Furthermore, the business models identified by Janssen and Zuiderwijk (2014) and Ranerup et al. (2016) are not directly applicable to Trusted Data Intermediaries, as they relate to the open government data and public service domain. Open government data is different from the data exchanged in domain of Trusted Data Intermediaries, because open data refers to public sector data made available to a large public (Janssen and Zuiderwijk, 2014). This is not necessarily the case for Trusted Data Intermediaries, as data contributors may also come from the private sector, and data is not required to be published publicly. Similar, public service platforms relate to E-government, while Trusted Data Intermediaries are not limited to the this domain alone. Last, Täuscher and Laudien (2018) offer insights in business models for intermediaries as electronic marketplaces. However, the Trusted Data Intermediary does not meet all the conditions set by the authors to be classified as an electronic marketplace. For instance, Trusted Data Intermediaries may produce or trade products and services itself. Also, data contributors and data users are not necessarily in direct contact for the initiation and realization of data transactions. As well, an institutional and regulatory frame for transactions may not necessarily be offered by Trusted Data Intermediaries. Thus, current literature on intermediaries lack to explain business models for entities like Trusted Data Intermediaries.

2.2.2. Kinds of value created and offered by intermediaries

Value is a much discussed, but semantically ambiguous concept (Austin and Seitanidi, 2012; Quelin, Kivleniece, and Lazzarini, 2017). Academic literature from various disciplines provide concepts to distinguish value creation by various entities. An overview of these concepts and their definitions is provided in Table 2.3. To begin, created value in a more general sense is referred to by Jetzek, Avital, and Bjørn-Andersen (2013) as **generated value**. Value is generated by activities that use resources and gain utility for a target group. The difference between the increased utility and resources used in the activity determines the generated value (Jetzek et al., 2013). Janssen and Zuiderwijk (2014) state that the role fulfilled by the intermediary generates value for the data contributor and data users.

Value created by public organisations can be stipulated as **public value**. Moore introduced the notion of public value creation in 1995 (Quelin et al., 2017). From a public management perspective, public organisations' managerial activities create value to the public (Quelin et al., 2017). Contrary, in public administration literature, public value is referred to as the benefits that accrue society minus the costs to achieve these benefits (Quelin et al., 2017). Different from public organisations, private organisations create **shared value**. Porter and Kramer discuss shared value creation from the discipline of strategic management (Quelin et al., 2017). Therefore, it looks at value from the perspective of the private sector rather than the public sector. Porter and Kramer (2011) introduce shared value to overcome the perception that public value creation is philanthropic by heart. They show that business activities that benefit society do not exclude business profits and call upon collective action by private companies to create both economic value as value for society (Porter and Kramer, 2011). However, this concept has been criticized for its uncertainty in accommodating various forms of public-private and cross-sector organisations (Quelin et al., 2017).

Value created by collaborations between organisations is stipulated by Austin and Seitanidi (2012) as **collaborative value**. Collaborative value is defined as the resulting benefits relative to costs that accrue society, organisations and individuals, as a result from interaction between collaborators (Austin and Seitanidi, 2012). This concept therefore is a hybrid between the concepts of public value and shared value, in the sense that it looks at the costs minus revenues as well as value for organisations. The authors view value as a spectrum from "sole creation" to "co-creation". Whereas the independent activity of an organisation is referred to as sole value creation, co-created value originates from joint activity by collaborators. To shed light on the dimensions of value created by collaborations, the authors suggest four types of value that organisations can derive from collaborations: **associational value**, **transferred resource value**, **interaction value** and **synergistic value** (Austin and Seitanidi, 2012). These types of value are best explained in the scenario where two organisations collaborate. In this scenario, associational value is the value derived by an organisation from being associated with the other organisation. Transferred resource value is the value that an organisation derives from receiving resources from the other organisation. The processes that occur during the collaboration between the organisations generate interaction value (e.g. trust, knowledge adoption, communication and conflict resolution). Last, synergistic value is the term used to describe the value derived by organisations from the achievement of outcomes that they would not achieve alone (Austin and Seitanidi, 2012). In the framework created by Austin and Seitanidi (2012), the four types of value are higher as more co-creation (i.e. collaboration between organisations) occurs. The four types of value described by Austin and Seitanidi (2012) are conceptualized by Jetzek et al. (2013) as **appropriated value**. When an organization is able to capture a portion of the value created by an activity, value is appropriated (Jetzek et al., 2013). As such, in the context of data collaboratives value appropriation is what can be understood as how an organisation benefits from participating.

Table 2.3: Concepts for value created by organisations.

Concept	Definition
Generated value	The difference between the increased utility for a target group and resources used in the activity (Jetzek et al., 2013)
Public value	Value created for the public from public organisations' managerial activities (Quelin et al., 2017) (Public management perspective)
Public value	The benefits that accrue society minus the costs to achieve these benefits (Quelin et al., 2017) (Public administration perspective)
Shared value	Policies and operating practices that enhance the competitiveness of a company while simultaneously advancing the economic and social conditions in the communities in which it operates ((Porter and Kramer, 2011), p. 6)
Collaborative value creation	The resulting benefits relative to costs that accrue society, organisations and individuals, as a result from interaction between collaborators (Austin and Seitanidi, 2012)
Associational value	The value derived by an organisation from being associated with the collaborating organisation (Austin and Seitanidi, 2012)
Transferred resource value	The value that an organisation derives from receiving resources from the collaborating organisation (Austin and Seitanidi, 2012)
Interaction value	Value generated in the processes that occur during the collaboration between the organisations (e.g. trust, knowledge adoption, communication and conflict resolution) (Austin and Seitanidi, 2012)
Synergistic value	The value derived by organisations from the achievement of outcomes that they would not achieve alone, as a result of collaboration (Austin and Seitanidi, 2012)
Appropriated value	A portion of the value that is created by an activity, which is captured by an actor (Jetzek et al., 2013)

Literature on value creation by intermediaries, specifically entities like Trusted Data Intermediaries, is scarce. Of the literature that accommodates the chosen definition of Trusted Data Intermediaries, some focus solely on public value creation (e.g., Janssen and Zuiderwijk, 2014). Or, authors lack to stipulate their findings as value. For instance, Yoon et al. (2019) found that data intermediaries assert roles for connecting communities with data through easy access, data aggregation and analysis and helping communities use data for improved decision-making. In addition, they offered different types of data services, serve as an information resource and bridge communities and different partners (Yoon et al., 2019). Similar, van Schalkwyk et al. (2016), Durrance et al. (2007) and Mercado-lara and Gil-garcia (2014) found that data intermediaries offer solutions to other actors. While these references provide an indication of what kinds of value Trusted Data Intermediaries may create, they refer to the domain of open data. Therefore as described in section 2.2.1, the results are not directly applicable to Trusted Data Intermediaries.

To summarize, the concepts in Table 2.3 address the types of value created on different levels. Collaborative value, in my view, corresponds to the value creation by a data collaborative. On a lower level, the value *derived by* a collaborating organisation can be classified as appropriated value, which can be either associational, transferred resource, interaction or synergistic value. On the other side, the value *created by* a single organisation may be stipulated as public value for public organisations, or shared value for private organisations. However, Trusted Data Intermediaries can be more than organisations, e.g. a consortium, individuals et cetera. Thus, a concept for the value created by intermediary *entities* is lacking. Because the four types of values suggested by Austin and Seitanidi (2012) are described as ways in which collaborators derive value from the collaboration, it remains unclear whether this value is *offered by* intermediary entities as well. Literature on value creation by data intermediaries does not link findings with these value concepts, nor can they be directly applied to Trusted Data Intermediaries due to their pertinence to open data. As such, knowledge lacks on what kinds of value is created by Trusted Data Intermediaries and offered to data contributors and data users.

2.2.3. Intermediaries' resources for value creation

Business models that follow a resource-based view of an organisation on value creation, presume that value creation comes from unique combinations and implementation of complementary and specialized resources and capabilities (Amit and Zott, 2001). In their application of this view on virtual markets, Amit and Zott (2001) explicate **information-based resources**. Opposed to other types of resources and capabilities, information-based resources have a higher degree of mobility. This characteristic makes that value migration can increase, thus making such resources more important in e-business firms. However, it can also make newly created value may become less sustainable, and this prospect may impede on the incentive to create value. To overcome this, the organisation may acquire **substitute resources** from network partners (Amit and Zott, 2001). This will be elaborated in the next sub section.

Specific to resources of intermediaries, the literature on value creation by organisations discusses costs or resources for value creation. However, academic literature on resources of intermediaries is ambiguous. For instance, in the article of van Schalkwyk et al. (2016) the term resources refers once to data sources, although the term is later used to refer to assets needed to create value. In addition, literature lacks to specify resources, as only a classification of their nature is offered (e.g., Digital Civil Society, 2017; van Schalkwyk et al., 2016; Yoon et al., 2019). Janssen and Verbraeck (2005) and Stadtler and Probst (2012) do specify technical and human resources, but do not refer to them as such.

Sometimes, the term **organizational resources** is used as a collective term to describe organizational assets deployed to create value (e.g., Stadtler and Probst, 2012; Yoon et al., 2019). These assets can include human resources, technical resources (Yoon et al., 2019), reputation and social networks (Stadtler and Probst, 2012). When described separately, literature on intermediaries mostly discuss **human resources** and **financial resources**, but fail to specify them. For instance, in their study of intermediaries in an open data ecosystem, van Schalkwyk et al. (2016) briefly mention financial and human resources for data supply, yet they do not specify of what these resources consist exactly. Similarly, Mercado-lara and Gil-garcia (2014) state that data intermediaries need skilled human resources and the use of new technologies. In addition, in the non-academic study of Digital Civil Society (2017), Trusted Data Intermediaries are described as dedicating financial and human resources as well.

Although not described as resources of a specific nature, Janssen and Verbraeck (2005) discuss technical and human resources. The authors argue that intermediaries in electronic markets need to use information and communication technology for efficiently coordinating the exchange of information between entities. In order to develop such technology, considerable technical and domain-specific expertise is necessary (Janssen and Verbraeck, 2005). Similarly, Stadtler and Probst (2012) touch upon specialized staff and processes, which can be classified as human resources. Further, the authors discuss knowledge management systems, which can be classified as both **technical resources** and information-based resources. Moreover, the authors discuss **bridging social capital** as a resource of intermediaries as brokers of public-private partnerships. Social capital encompasses trust, norms and values accessed through a network of social relations. The authors argue that intermediaries have bridging social capital, which is embedded in large networks and organizational legitimacy (Stadtler and Probst, 2012). However, due to this embedment, bridging social capital can only exist if the broker is institutionalized in a network. Therefore, bridging social capital can not be a resource for an intermediary that has not yet established a network.

In summation, business model literature offers a classification of resources as information-based resources and substitute resources. Further, intermediaries literature identify resources of financial, human, technical and social nature. However, a specification of these resources lacks. As such, the knowledge gap on key resources utilized by Trusted Data Intermediaries for the creation of value for collaborators remains.

2.2.4. Intermediaries in networks of social relations

Networks are an important part of business models, because networks can offer access to new sources of information, capabilities, markets and technologies. It also enables an organisation to share risks, knowledge, facilitate learning, generate economies of scope and scale and to reap benefits that accrue from interdependent activities. Furthermore, networks enhance the range of possible organizational arrangements for value creation (Amit and Zott, 2001). As described above, the network creates the possibility for an organisation to choose substitute resources of network partners. Weill and Vitale (2001) refers to such partner actors as “allies”. More specific, Bouwman et al. (2008) argue that partners can be distinguished in **critical partners** and **supporting partners**. Critical partners provide **critical resources** (i.e. resources that are indispensable and irreplaceable), while supporting partners provide **supporting resources** and capabilities that can be substituted by others.

For intermediaries’ business models, literature on networks is limited to the notion that intermediary networks consist of data contributors, data users and other entities. For instance, the research of

Janssen and Zuiderwijk (2014), shows that besides the data contributors and data users, the network of infomediaries can exist of governments, businesses, social institutions, experts, as well as citizens. These actors can assert roles such as providing funds and developing or operating the business model. Sometimes an actor fulfills a more crucial role, other times a supporting role (Janssen and Zuiderwijk, 2014). While these results were obtained in the domain of open government data, it identifies the possibility of additional actors in the network of intermediaries, as well as the notion that network partners may be critical or supporting. However, for entities like Trusted Data Intermediaries, an identification of network partners lacks.

In business model literature, networks are discussed in combination with the concept of governance. However, different implementations of governance are used. One focuses on the governance of transactions, while the other focuses on the governance of the network. To start, Amit and Zott (2001) discuss governance mechanisms as influences for the efficiency of transactions, hereby following Williamson's transaction cost theory (Williamson, 1975, 1979, 1983 as cited in Amit and Zott, 2001). According to transaction cost economics, governance refers to the institutional framework that determines the integrity of a transaction (Williamson, 1979). In later work, Williamson defines governance as "the means by which order is accomplished in a relation in which potential conflict threatens to undo or upset opportunities to realize mutual gains" (Williamson, 1998, p. 37). Thus, governance describes the "play of the game" (Williamson, 1998): contractual institutions between actors that govern the transaction (Williamson, 1998). According to Amit and Zott (2001), transaction cost economics is critical for the understanding of electronic business, because a reduction of transaction costs can lead to higher value creation. Therefore, the institutional framework that governs the transaction costs are highly important. Contrary to the governance of transactions in the network, more recent business model literature looks at network governance. This focuses primarily on which network actor takes a dominant role and how they exercise this role (e.g., Al-Debei and Avison, 2010; Bouwman et al., 2008).

Literature on intermediary business models only discusses network actors' dominance (e.g., Janssen and Zuiderwijk, 2014). Although the literature on intermediary business models does not explicate institutions that govern transactions, the literature used to define Trusted Data Intermediaries does. To begin, social capital explicates the idea that an intermediary is positioned in a network of social relations that is tied by trust, norms and values (Stadtler and Probst, 2012). As described by Stadtler and Probst (2012), intermediaries can develop "bonding social capital" in the form of **norms that create trust, communication and information sharing**. In addition, bonding social capital are strong social ties that originate from strong relationships with others in the network (Stadtler and Probst, 2012). Further, Klievink et al. (2016) describes governance as the agreements on standards and procedures that guide the activities of network partners. They find governance instruments for public-private platforms to consist of **standards** and **agreements**. In addition, Digital Civil Society (2017) describes negotiation mechanisms identified in the workshop on Trusted Data Intermediaries. The mechanisms include permission policies (1), rights management software (2), licenses (3), contract law (4), technological strategies for security, privacy and self-reporting (5), self-regulation and review processes (6), regulatory statutes (7), law (8), self-certification (9) and codes of ethics (10) (Digital Civil Society, 2017). These mechanisms correspond to the findings of Klievink et al. (2016) and the notion of governance as contractual institutions that govern transactions as described by Williamson (1998).

To summarize, intermediaries operate in a network of partners. Partners can be critical or supportive to value creation. Transactions in the intermediary's network are governed by institutions which thereby affect the value creation. Literature identifies such institutions to be standards, agreements or norms that create trust, communication and information sharing. However, the literature on intermediary

business models so far has not accommodated these institutions. Rather, current literature focuses on the dominance of actors in a network. Therefore, the state of the art regarding governance mechanisms of Trusted Data Intermediaries lacks.

2.2.5. Intermediaries' financial arrangements

The last aspect to business models concerns the financial arrangements. These can relate to revenue streams and cost structures (Zott et al., 2011), pricing methods (Al-Debei and Avison, 2010), as well as the division of costs, revenues and investments over the network (Bouwman et al., 2008). As indicated before, value creation incurs costs (Austin and Seitanidi, 2012; Jetzek et al., 2013; Quelin et al., 2017). To be financially relevant (Al-Debei and Avison, 2010), costs must be covered by revenues. When the revenues exceed the costs, a profit can be made.

There is very little literature on financial arrangements of intermediaries, let alone entities like Trusted Data Intermediaries. Weill and Vitale (2001)'s business model supposes that **revenues** come from buyers, sellers or both. However, Digital Civil Society (2017) identified that not all Trusted Data Intermediaries create revenues and defines Trusted Data Intermediaries as **non-profit** organisations. Janssen and Zuiderwijk (2014) solely studied non-profit infomediaries' business models and focused on public value creation rather than financial arrangements. However, Ranerup et al. (2016) studied public service platforms. In their study, they found that most public service platforms are financed by **grants** or project **funds**. Also, they show that private organisations can generate revenue streams as a result of a **pricing model**.

While there is really not much to go on, literature indicates that entities like Trusted Data Intermediaries can create revenues, or operate on a non-profit basis. In the case of a for-profit entity, a pricing model can lead to revenue streams. This pricing model may be imposed on data contributors, data users or both. Overall, knowledge on financial arrangements is severely lacking, resulting in a knowledge gap.

3

Theoretical framework

This chapter lays out the theoretical framework which will be applied to answer the research questions. As delineated by the research questions, the focus lies solely on the value created for collaborating organisations, so value created for society will not be addressed by the theoretical framework. In this chapter, section 3.1 explains why the Unified Business Model framework of Al-Debei and Avison (2010) is chosen as the basis for the theoretical framework developed in section 3.2. Finally, Table 3.2 provides the chosen theoretical framework.

3.1. Business Model ontologies

To describe a business model, many ontologies exist (e.g., Ballon, 2009; Bouwman et al., 2008; El Sawy and Pereira, 2013b; Gordijn, 2004; Heikkila, Heikkila, and Tinnila, 2005; Joyce and Paquin, 2016; Osterwalder and Pigneur, 2010). A business model ontology is an explicit, simple specification of a conceptualization of components of a business model and the relationships between them (Heikkila et al., 2005; Kort and Gordijn, 2008).

An ontology that will be applied to answer the research questions (see Table 1.1) needs to meet certain criteria in order to be used to answer the research questions. Firstly, the ontology needs to accommodate the kinds of value proposed by the Trusted Data Intermediary for different customer segments (i.e. data contributors and data users). Second, the ontology needs to address the (different types of) resources utilized for the value creation. Third, it should explicitly address the network of social relations and accommodate the identification of governance mechanisms for different network partners. Fourth, the ontology needs to allow for the identification of a finance structure. Furthermore, the context of Trusted Data Intermediaries in data collaboratives should be taken into account. Data collaboratives do not necessarily have to be long-term relationships (Susha et al., 2017a), but it has been found to be a success factor (Susha et al., 2019). Therefore, an additional criterion is formulated that the ontology should accommodate possible long-term relationships. Based on these criteria, the ontology needs a level of abstraction that is sufficiently broad to capture a variety of Trusted Data Intermediary

business models, while sufficiently detailed to answer the research questions and enable new findings. Table 3.1 provides an overview of potential ontologies found by the author and a discussion offered by literature, as well as an assessment of why an ontology is more or less applicable as a theoretical framework in this thesis.

From the discussions in Table 3.1 it becomes clear that ontologies are out of scope (e.g., Joyce and Paquin, 2016; Osterwalder and Pigneur, 2010), or that a trade-off is made between a low level of abstraction and generalization. Ontologies are either lacking elements that are demanded by the research questions (e.g., El Sawy and Pereira, 2013b; Gordijn, 2004; Heikkila et al., 2005). Or, ontologies contain more elements than needed (e.g., Ballon, 2009; Bouwman et al., 2008). Contrary, the Unified Business Model framework as proposed by Al-Debei and Avison (2010) falls within scope and finds balance in this trade-off. Therefore, this ontology is selected as the foundation of the theoretical framework.

3.1.1. The Unified Business Model framework

The Unified Business Model Framework (UBM Framework) by Al-Debei and Avison was published in 2010. The framework is deduced from 22 business model conceptualizations found by content analysis following a comprehensive literature review. The framework's ontological structure comprises of four dimensions, each explained by specific constituent elements (see Figure 3.1). Studies by Janssen and Zuiderwijk in 2014 and by Ranerup et al. in 2016 show that the Unified Business Model Framework can be applied to entities that act as intermediaries. More particular, Ranerup et al. (2016) show that the framework can be applied to service platforms. We should recall from section 1.2.1 that service platforms is a data intermediary business model found in the research by Janssen and Zuiderwijk (2014). As such, the framework has been applied in research very similar to this thesis' objectives. Therefore, to analyze Trusted Data Intermediaries' value creation for collaborators, the Unified Business Model Framework (Al-Debei and Avison, 2010) is chosen to base the theoretical framework on for answering the research questions.

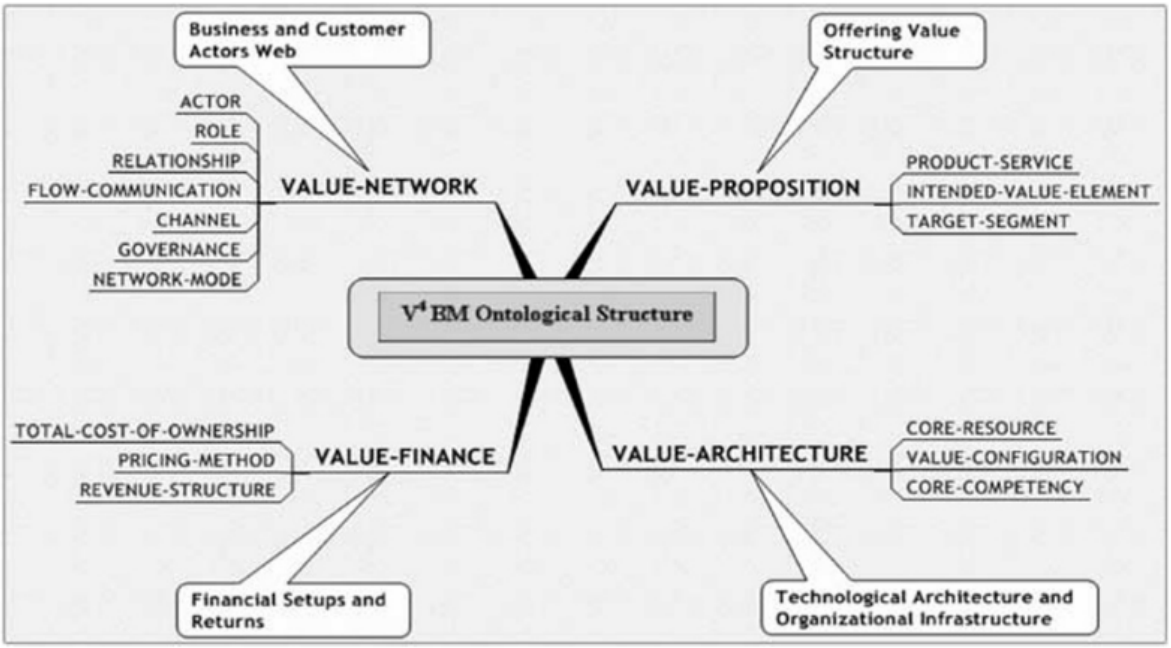


Figure 3.1: The Unified Business Model Framework as proposed by Al-Debei and Avison (2010).

Table 3.1: Discussion of Business Model ontologies.

Ontology	Discussion
Business Model Canvas (Osterwalder and Pigneur, 2010)	Very accessible and easy to learn because of its wide adoption and tools created for it (e.g. https://www.strategyzer.com/). Easy to apply, but little detail to design variables and leaves room for interpretation (Bouwman et al., 2012). Primary focus on individual companies (Bouwman et al., 2012) and short term transactions (Heikkila et al., 2005). Because of the latter, the ontology is deemed less appropriate for this thesis.
Triple Layered Canvas (Joyce and Paquin, 2016)	Integrates economic, environmental and social domains by adding two layers to the Business Model Canvas. Allows for easy exploration, development, visualization and communication of business models. But like the Business Model Canvas, it lacks detail to analyze on lower levels (Joyce and Paquin, 2016). Further, the environmental and social layers are out of scope for Trusted Data Intermediaries. Therefore, this ontology will not be used.
E3-value model (Gordijn, 2004)	Particularly looks at e-commerce business models (Gordijn, 2004), but has been applied in business-to-business scenarios (Kort and Gordijn, 2008). Can be visually represented and further analyzed (Gordijn, 2004; Kort and Gordijn, 2008). Focus lies on the exchange of economic value throughout a network. The ontology does not accommodate the governance mechanisms sought for by the research questions, nor the resources or financial structures of one entity. As such, the ontology will not be chosen.
C-soft (Heikkila et al., 2005)	Emphasis on customer relationship and adjustments necessary to reach a joint business model. Primary focus on long-term relationships within a network of companies (Heikkila et al., 2005) who have come to a joint business logic. The constructs of the business model (service, technology, organization of network, and finance) accommodate the research questions. However, the ontology focuses on technical resources. Further, because of its emphasis on joint business logic, an understanding of business models of data contributors and data users is demanded. This is more complex than necessary to answer the research questions. Thus, the ontology will not be adopted.

Ontology	Discussion
VISOR (El Sawy and Pereira, 2013b)	A business model that focuses on digital platforms and ecosystems. It comprises of five components (value proposition, interface, service platform, organizing model, and revenue model) (El Sawy and Pereira, 2013b). Each component has specific elements (descriptors), but governance is lacking from these descriptors. Further, the model juxtaposes maximization of the customers' willingness to pay with the minimization of business costs. This juxtaposition assumes a profit-driven business model, while this is not necessarily the case for a Trusted Data Intermediary. Thus, the ontology is not selected.
Ballon's BM (Ballon, 2009)	Looks at value configurations rather than a single value proposition, in particular for an ICT-enabled service (Ballon, 2009). Lacks detail to analyze the business model on lower levels of abstraction (Bouwman et al., 2012). Although the dimensions accommodate the topics of the research questions, each dimension contains multiple elements not necessary to answer the questions. This makes the ontology overly complex as a theoretical framework, as such the ontology is dismissed as a candidate.
STOF-ontology (Bouwman et al., 2008)	Considers the ICT-enabled services in a network of companies required to deliver a service (Bouwman et al., 2008) and operates on a detailed level of abstraction. Mainly focuses on business-to-consumer electronic services. Returns a descriptive model containing design variables. Model can be applied in a within-case analysis, as well as cross-case analysis and quantitative analysis (Bouwman et al., 2008). Emphasizes customer value and exposes interdependencies between components over four domains (service, technology, organization, finance) that can be analyzed along the critical design issues established for each domain (Bouwman et al., 2008). The STOF ontology as a dynamic conceptual framework over time can be succeeded by the STOF method for the design of business models (Bouwman et al., 2008). While the ontology contains all aspects needed to answer the research questions, the ontology operates on a lower level of abstraction than demanded. Thus, the application of the ontology would be an effort larger than made possible by the time constraints for this thesis. Therefore, the ontology is not chosen.
Unified BM Framework (Al-Debei and Avison, 2010)	Framework specific to information systems. The ontology synthesizes key aspects of 22 business model descriptions from the period 1998-2008. The definitions originate from the domains of information systems, eCommerce, eBusiness, technology and telecom industry, business management and eGovernment. It can be used for both the design as the assessment of business models. The ontology focuses on four primary constructs of business models: value proposition, architecture, finance and network (Al-Debei and Avison, 2010). The elements in these dimensions are concise and can be used to answer the research questions. Based on the foundation of the ontology in different domains, and the fit of the ontology's dimensions with the research questions, this ontology is selected.

3.2. Theoretical framework for Trusted Data Intermediary business models

The research questions (see Table 1.1) demand the description of kinds of value, resources, governance mechanisms and financial arrangements of Trusted Data Intermediaries. As can be seen in Figure 3.1, the Unified Business Model framework as proposed by Al-Debei and Avison (2010) contains more dimensions than required to answer these research questions. As such, elements from the framework will be adopted and adapted in the framework for the evaluation of Trusted Data Intermediary business models. The following sub sections describe which elements are chosen, and how they are operationalized. Table 3.2 shows the Unified Business Model Framework adapted to Trusted Data Intermediaries.

Although portrayed as separate dimensions, the dimensions are interrelated and interdependent (Al-Debei and Avison, 2010). As such, they can only be applied correctly when these interaction effects are considered and discussed in the results.

3.2.1. Value proposition: kinds of value

Firstly, the question on kinds of value offered to data contributors and data users can be answered in the value proposition dimension. This dimension describes the *target segments*, *intended-value elements* and *products-services*.

Target segments refer to the entity to whom value is being offered. In order to answer the first questions, the target segments of Trusted Data Intermediaries will consist of data contributors and data users. I broadly define a data contributor as an entity that shares, opens or provides access to data. Opposing, a data user is defined as an entity that uses, downloads, processes or gains access to data. *Products or services* are offered by the Trusted Data Intermediary to target segments. Products can be both tangible and intangible, for instance physical objects (e.g. hardware), or virtual or electronic products (e.g. tools or technologies (van Schalkwyk et al., 2016)). Services are intangible, in the sense that the Trusted Data Intermediary executes a task or function in favor of the target segment. Overall, this sub dimension can be described as the offering for, or receiving, obtaining, getting of a product, service by a data contributor resp. data user. The *intended value element* refers to what value the products and services are supposed to offer the target segments. From Table 2.3 we recall that many different concepts for value exist. In my interpretation, I see a common theme in these concepts of increased utility or benefits that accrue a target segment. Thus, the intended value element is described as the enabling of utility, gains, benefits, opportunities, possibilities for a data contributor resp. data user.

3.2.2. Value architecture: resources

Secondly, to answer the second research sub question, the value architecture dimension needs to address resources only. Section 2.2.3 indicates different terms to describe resources. Broadly, Trusted Data Intermediary resources for value creation can be defined as the organizational assets deployed to create value. While a specification lacks, literature (e.g., Janssen and Verbraeck, 2005; Stadtler and Probst, 2012) exemplifies specialized staff and processes, expertise, communication and information technology, knowledge management systems, and a network's trust, norms and values. I expand these examples with the financial capital of a Trusted Data Intermediary utilized to create value, as well as synonyms like software, digital infrastructure, data management systems.

3.2.3. Value network: governance mechanisms

Third, to answer the question regarding governance mechanisms, we look at the value network dimension. Specifically, I focus on the elements of *governance* and *network role*.

I define the *governance* element as the mechanisms leveraged by the Trusted Data Intermediary to coordinate data contributors, data users and network partners. Following the state of the art (sub section 2.2.4), these mechanisms can range from norms to contracts to processes. My definition adopts a conceptualization of governance that deviates from that of Al-Debei and Avison (2010). Al-Debei

and Avison (2010) refer to the governance element as which actor is governing or being dominant. This focus will not answer the research question. Thus, a different approach is adopted, namely the conceptualization of governance by Williamson (1998). In this conceptualization, governance refers to the "play of the game", primarily based on contractual institutions between actors that governs the transaction cost economics. According to Amit and Zott (2001), transaction cost economics is critical for the understanding of electronic business. I deem this application of governance appropriate for answering the research question, because it allows for the identification of mechanisms on the same level of abstraction on which Digital Civil Society (2017) and Klievink et al. (2016) identified negotiation mechanisms and governance instruments. The state of the art review indicates that the network of the Trusted Data Intermediary possible exists of more than customer actors. Furthermore, Al-Debei and Avison (2010) indicate that the value network dimension is interrelated and interdependent with other dimensions, such as the value architecture. For example, resources may be acquired from the network (Al-Debei and Avison, 2010). These substitute resources may range from resources that support the product-service offering, or resources that finance the business model, e.g. funds or grants. Therefore, another element will be the network *role* of actors, to be described as the provision of resources or execution of processes by an entity which are leveraged in the process of value creation for data users and data contributors.

3.2.4. Value finance: financial arrangements

In the value finance dimension, the financial arrangements of Trusted Data Intermediaries is addressed. Elements such as *cost- and revenue structures* and *pricing methods* will be addressed.

I define *cost- and revenue structures* as the arrangements that lead to revenues for the Trusted Data Intermediary from offering products and services or receiving of funds or grants, and costs incurred from the deployment of resources. Contrary to the original framework, I choose to combine the elements of cost- and revenue structures. Because, the state of the art review identified that some Trusted Data Intermediaries generate revenues, while others operate on a non-profit basis, or receive grants or funds. The combination of cost-revenues into one element accommodates the discussion on the level of for-profit versus non-profit organisation and accommodates the income from funds. An alternative application of the value finance dimension for intermediaries was exercised by Janssen and Zuiderwijk (2014). To accommodate the non-profit nature of the studied intermediaries, Janssen and Zuiderwijk (2014) extended the value finance dimension to discuss the public value rather than the pricing, and cost-revenue streams. However, an extension like that of Janssen and Zuiderwijk (2014) would not fall within the set scope of this research questions. Therefore, this application is not adopted. The second element discussed in this dimension is the *pricing method* element. A Trusted Data Intermediary may generate revenue streams from charging the products and services offered to data contributors and data users. This came forth in the research of public service platforms by Ranerup et al. (2016). A product or service can be completely free of charge, or has a cost associated to its use. In addition, well-known examples of intermediary platforms such as Flickr introduce the possibility of "freemium" (free and premium) business models (Teece, 2010). Because knowledge on financial arrangements of Trusted Data Intermediaries is lacking, the pricing method is broadly defined as the fees, charges, subscription, costs of use that are or aren't imposed on data contributors or data users.

Table 3.2: Theoretical framework for Trusted Data Intermediary business models.

Dimension	Elements	Description
Value Proposition	*Target-segment*	An entity that shares, opens or provides access to data (data contributor) or an entity that uses, downloads, processes or gains access to data (data user)
	Product-service	The offering for, or receiving, obtaining, getting of a product, service by a data contributor resp. data user
	Intended-value-element	The enabling of utility, gains, benefits, opportunities, possibilities for a data contributor resp. data user
Value Architecture	*Core-resource*	The organizational assets deployed to create value, e.g. financial capital, software, digital infrastructure, staff, processes, systems, trust, norms and values in the network
Value Network	*Governance*	The mechanisms or contractual institutions leveraged by the Trusted Data Intermediary to coordinate data contributors, data users and network partners, e.g. contracts, agreements, licenses
	Role	The provision of resources or execution of processes by an entity which are leveraged in the process of value creation for data users and data contributors
Value Finance	*Cost-revenue-structure*	The arrangements that lead to revenues for the Trusted Data Intermediary from offering products and services or receiving of funds or grants, and costs incurred from the deployment of resources
	Pricing-method	The fees, charges, subscription, costs of use that are or aren't imposed on data contributors or data users

4

Methodology

To answer the research questions, the case study method is applied. Section 4.1 explains how the case study method takes form. Section 4.2 explains how data is collected from three sources of evidence, while section 4.3 elaborates on the method for making inferences from the collected data. Section 4.4 explains how this data is analysed such that it answers the research questions. For each method, limitations are discussed and reflected upon in section 4.5. In Figure 4.1, the research design as applied in this thesis is provided.

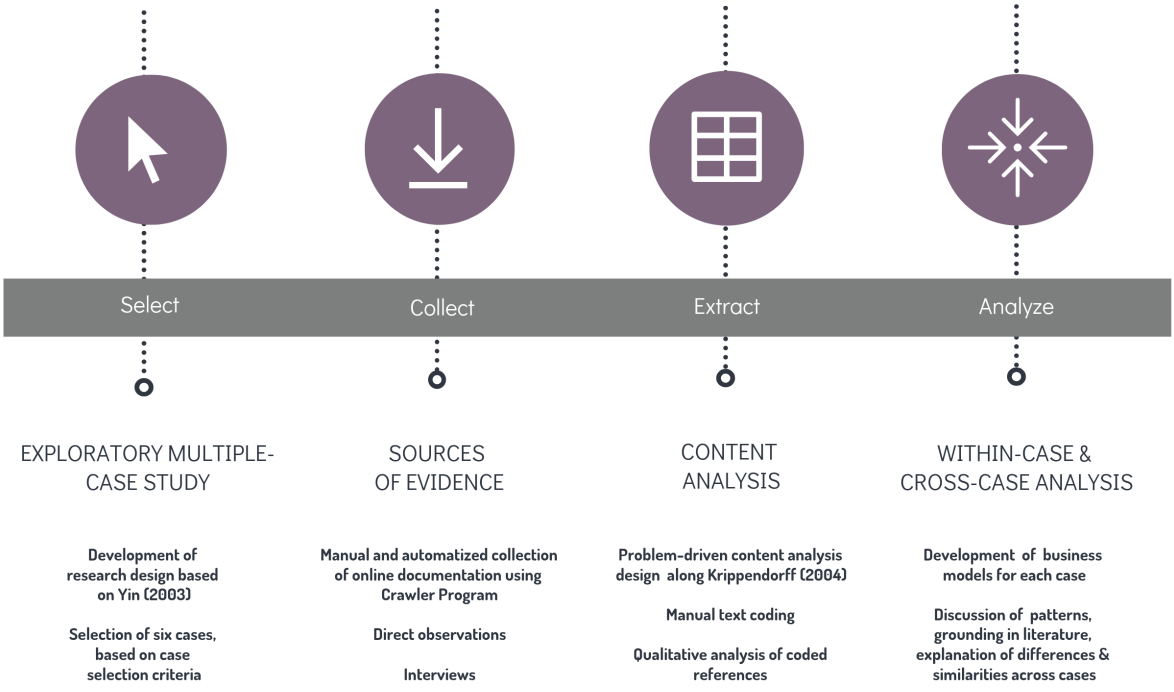


Figure 4.1: Research design.

4.1. Exploratory multiple-case study

The case study method is chosen based on the type of research questions and novelty of the research area. To begin, the research questions of Table 1.1 demand the description of Trusted Data Intermediaries' business models through "what" questions. According to Yin (2003), these types of questions favor methods like archival analysis and surveys. However, Trusted Data Intermediaries are a relatively new research area. As such, there is only limited data available to apply archival analysis. The novelty of the research area and the scope of the object of study also limits the amount of potential respondents of a survey. In contrast, the case study method does not rely on prior empirical quantitative or qualitative data, allows for the analysis of a small sample and is applicable to "what" questions as well (Eisenhardt, 1989; Yin, 2003). As such, qualitative research in the form of case studies is preferred over competing research methods. To conclude, the case study method is chosen to answer the research questions, as this is suitable for the accomplishment of demanded descriptions in novel research areas (Eisenhardt, 1989).

The case study method is limited in several ways which can impede on thesis results. First, an unsystematic approach can create a lack of rigor (Yin, 2003). To address this drawback, the next sections will address the research design, methods for data collection and methods for data analysis extensively. Second, generalization of results is seen as a challenge in case study research (Eisenhardt, 1989; Yin, 2013,0). Yin (2003) emphasizes that qualitative case study results cannot be generalized to populations (statistical generalization), but can be generalized to theoretical propositions (analytical generalization). The latter is especially the case in multiple-case study designs. Moreover, a multiple-case study design addresses the limitation of premature or false conclusions drawn from limited amounts of data analysed in unique contexts of a single case study (Eisenhardt, 1989). Although a multiple-case study requires more resources, time and requires a replication logic, it is favorable for the inquiry of typical cases and results in more compelling, robust findings (Yin, 2003). Thus, to overcome the second drawback, this thesis opts for a multiple-case study design. A third possible drawback of the case study method is that its results lack parsimony, logical coherence and testability (Eisenhardt, 1989; Yin, 2003). Thus, in the presentation of results, the parsimony of developed theory on Trusted Data Intermediary business models needs to be guarded. For logical coherence and testability, the research design will be tested for *construct validity*, *external validity*, as well as *reliability*. Section 4.5 will describe how the research design accommodates these tests. A reflection on the methodological limitations' implications for the research results is offered in section 7.3.

The research design explicates the logical sequence which links the collection and analysis of data and the conclusions drawn from the analysis to the research questions (Yin, 2003). In this thesis, the research design has been constructed along the proposed approach by Yin (2003). This approach is one of the leading methodologies in case study research (Yazan, 2015). Yin's positivist approach (Yazan, 2015) is chosen over competing approaches, because Yin's methodology has been recommended literature throughout research methodology courses at my university. Further, other methods focus solely on the use of qualitative methods (Yazan, 2015). Given the early state of academic knowledge on Trusted Data Intermediaries and data collaboratives, Yin's approach creates an advantage for future research as they are enabled to apply the research design to quantitative data.

Due to the lack of knowledge on Trusted Data Intermediary business models, the case study will assume the characteristic of an *exploratory case study* (Yin, 2003). This case study will explore business models of entities entrusted with enabling data transactions between data contributors and data users across sectors, thereby generating value for collaborators, as well as public value (i.e. Trusted Data

Intermediaries). The purpose of the exploration is to describe business models of such entities that create value for data contributors and data users. Last, the exploration will be deemed successful when the case study's questions (the sub research question in Table 1.1) have each been answered and one or more business model archetypes can be derived from these answers. Further, this thesis' case study follows an embedded multiple-case research design. The rationale for this design is provided by two considerations. First, as described before, a multiple-case study design is chosen to overcome limitations regarding analytical generalization and robustness of the findings. Second, the units of analysis searched for in the cases are the dimensions of the theoretical framework in Table 3.2. As such, there are multiple units of analysis to be studied within a case.

4.1.1. Case selection

In deviance of Yin research design approach, cases are selected on the basis of theoretical sampling as suggested by Eisenhardt (1989). Practical and methodological approaches approaches for case selection are possible. Regarding the first, case study research text books and encyclopedia focus on practical considerations (Bleijenbergh, 2010; Dul and Hak, 2008, e.g.). A drawback of a practical approach is that, similar to random sampling of cases, these considerations can lead to sample bias (Seawright and Gerring, 2008). Contrary, methodological approaches prevention sample bias by offering various case selection techniques. For instance, Yin (2003) focuses on replication logic and Eisenhardt (1989) emphasizes theoretical sampling. Yin's replication logic requires a theoretical foundation which explicates under which conditions a phenomenon is likely to be found or not. Because literature on Trusted Data Intermediaries lacks, a theoretical foundation is absent. Yin's approach does not explicate how the replication logic applies in such a situation. Contrary to Yin (2003), Eisenhardt's theoretical sampling approach does not require theoretical propositions in advance (Eisenhardt, 1989). As such, the case selection will occur on the basis of theoretical sampling. To allow for theoretical sampling, case selection criteria are defined (see Table 4.1). Firstly, cases need to meet the definition of a Trusted Data Intermediary, that is they are entities entrusted with enabling data transactions between data contributors and data users across sectors, thereby generating value for collaborators, and enabling the creation of value. Second, in adoption of the practical considerations opted by Dul and Hak (2008), information on the cases must be available and accessible. Furthermore, to allow for cross-case analysis and the development of initial theories, the cases should be diverse. Tentative examples of diversity relate to the non-profit or for-profit nature of the Trusted Data Intermediary, or where they are active (e.g. global or local). Most importantly, the cases should be selected that cover the breadth of the concepts in the theoretical framework sufficiently. Based on the case selection criteria, six cases are selected. These cases comprise of *Social Science One*, *Vivli*, *Humanitarian Data Exchange*, *Amsterdam Data Exchange*, *City Innovation Platform* and *Dataverse*. Table 4.2 provides a brief description of each case.

Table 4.1: Criteria for case selection.

Case selection criterion
1. Cases need to meet the definition of a Trusted Data Intermediary, that is they are entities entrusted with enabling data transactions between data contributors and data users across sectors, thereby generating value for collaborators, as well as public value.
2. Information on the cases (case data) must be available and accessible.
3. Cases should be diverse, for instance in their geographic activity (local, global) or economic objective (for-profit, non-profit).
4. Cases should be selected that cover the breadth of the concepts in the theoretical framework sufficiently.

Table 4.2: Selected cases and their descriptions.

Case	Case description
Social Science One (SSO)	Social Science One (SSO) is an independent organization of academics whose goal is to connect independent academic experts from around the world in social science to corporate organizations while meeting both interests. SSO is incubated at the Institute for Quantitative Social Science at Harvard University.
Vivli	Vivli is an independent non-profit organization that created a data sharing and analytics platform for clinical research data. The platform originated from research by the Multi-Regional Clinical Trials center. Vivli aims to contribute to improved human health and advance science by enabling global data sharing.
Humanitarian Data Exchange (HDX)	The Humanitarian Data Exchange (HDX) is an online platform for humanitarians to share and use data on a global scale. The HDX is a service managed by the Center for Humanitarian Data. The Center is located in The Hague, and is a part of the Office for the Coordination of Humanitarian Affairs (OCHA), which in turn is a part of the United Nations Secretariat. Due to the UN-OCHA mandate, HDX is exclusive for data that holds relevance to the humanitarian sector.
Amsterdam Data Exchange (AMdEX)	The Amsterdam Data Exchange (AMdEX) is a project in development, that aims to function as a secure and trusted open marketplace for data. The project was initiated by the Amsterdam Economic Board, to be an original contribution to the European Open Science Cloud. Although currently focused on actors in the Netherlands, the case aims to contribute to the European network
City Innovation Platform (CIP)	Civity's City Innovation Platform (CIP) is a Dutch product focused on enabling smart cities. CIP core elements are a data market (1), security and privacy (2) and a Data Management Framework (3) which comprises of five modules: "Dataplatform", "Datasafe", "Dataplatform IoT", "Slim melden" and "Smart City Inzicht".
Dataverse (DV)	Dataverse is open source research data repository software, developed at Harvard. It is globally active and open for all research subjects.

4.2. Sources of evidence

To answer the research questions, information on the objects of analysis represented in Table 3.2 is collected. As such, information on target segments, products and service offering, intended-value, governance mechanisms, network actor roles, core resources, cost-revenue structure and pricing methods is searched. In the selection of cases, I found that almost each case has a web site and is active on social media, using the internet to explain their objectives and activities. Thus, information is collected from online documentation. In this section, I will explain why manual and automatized collection of online documentation is selected as a primary method for the retrieval of online documentation and how they are applied. Further, I will address a secondary source of data, namely observations from attending meetings and presentations. As well, interviews with case representatives is selected as a third source. In order to conduct data analyses, data on the business models' components of Trusted Data Intermediaries need to be extracted from the collected information. To do so, *content analysis* is applied (section 4.3). The limitations to these methods are addressed later on, in section 4.5.

4.2.1. Manual and automatized collection of online documentation

The primary source of evidence for this research concerns online documentation. Specifically, this documentation comprises the web contents from case websites, online articles mentioning the cases

or reports or other files published online. The method for collecting online documentation can be referred to as online desk research. Online desk research is suitable, because it allows to save time and resources for gaining access to the needed information. For instance, this method bypasses the potential problem that organisations are unwilling or incapable of granting information within the short time frame. This problem was encountered in the exploration of other research questions relating to data collaboratives. Further, the collected information is not tainted by the act of measurement (Krippendorff, 2004; Yin, 2003). On the other hand, document contents may be written for other purposes and audiences than those of this case study (Yin, 2003), which may hamper how representative the information is to the case study. Nevertheless, given the time frame and resources for conducting this research, online desk research is chosen as the primary data collection method.

In the application of online desk research in a multiple-case study, a systematic implementation is ensured. Therefore, three steps are sequentially applied for each case. First, I assess the case organisation's web site. This can be either the homepage, or a child page that discusses the objectives in more detail, e.g. an "about"-page. Second, the case website is searched for any recent documentation. If found, this document is downloaded. Third, a self-developed computer program will be executed that automatically collects and searches websites linked to the case organisation for textual data. The program is written in Python and is available through *https://github.com/mflipsen/Crawler*. The processes executed by the program are known as web crawling. In the remainder of this thesis I will refer to this program as the Crawler program. Should after these three steps information still be missing, manual online documentation is collected again, mainly focusing on web pages or documents in formats which are inaccessible by the Crawler program.

Web crawling is categorized as automated information retrieval (Saini and Arora, 2016). A web crawler is a program or script that automatically traverses web pages, copies the contents and downloads this to a specified location (Saini and Arora, 2016). In the research by Kim and Ha (2016), web crawling is applied to gather online data for the purpose of building a knowledge base on small businesses. In their study, online data entails content from social media and business web pages. The former is considered to change over time and is thus referred to as dynamic data. The latter is considered to be constant over time, e.g. name of the business, and is therefore referred to as static data (Kim and Ha, 2016). As for all methods, the implementation of web crawling requires a systematic approach. This can be referred to as a web crawling strategy.

In this thesis, a simple web crawling strategy is implemented to collect dynamic and static data. Here, the strategy entails that the program only selects web pages that are relevant to the case. To create an initial list of web pages that are relevant, I use the Google Search Engine API to return a maximum number of results specified by the user for a query of the case name. Next, these web pages are each traversed (i.e. crawled) by so-called spiders that execute a Breadth First Search algorithm. This algorithm collects all hyperlinks mentioned on a web page and queues this to be searched for more hyperlinks. Normally, this loop continues until all hyperlinks have been found (Nigam, 2014). However, in the Crawler program I have chosen to break the loop after one execution. This was done to reduce the run time and computational costs for each case. Subsequently, all the web pages found by the spiders are accessed and all data that belongs to user-specified HTML tags is extracted ("scraped"). The collected data is saved in a docx-format, which makes it readable through text-processor programs like Microsoft Word. The data is categorized by the source url. Figure 4.2 conceptualizes the Crawler program's main steps and where the loop breaks. Figure 4.3 provides an example of how the Crawler program output is represented per url.

While this strategy also incurs limitations (see section 4.5.2), the added value of the Crawler program is that it adds to the transparency and reliability of the data collection, as well as the efficiency of the research. Transparency and reliability are enhanced, because the Crawler creates files on the user's computer that reads which settings were used, and orders the output per web page. In addition, the efficiency is increased due to the fact that web pages are selected based on case relevance.

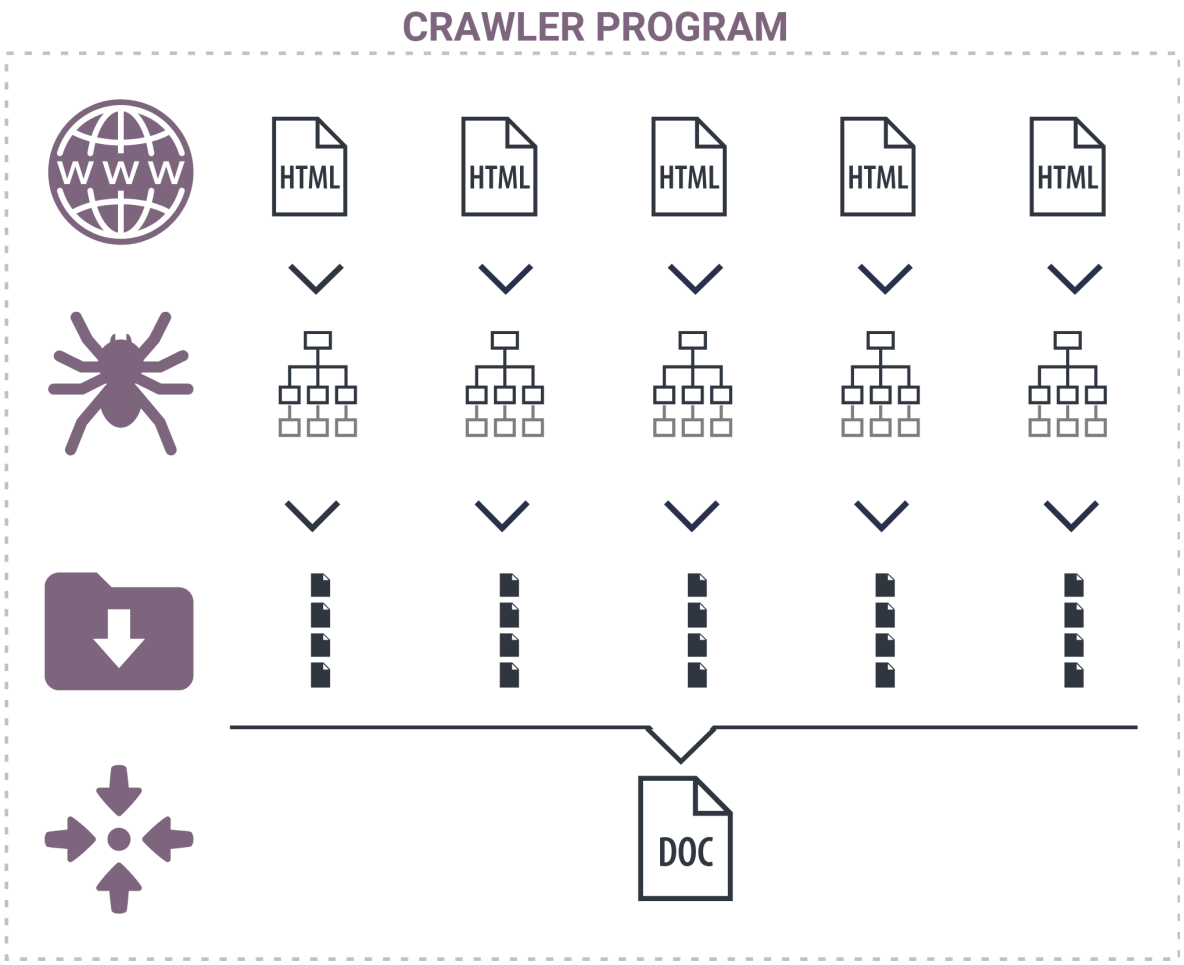


Figure 4.2: Conceptualization of automatized online documents collection (Crawler program).

4.2.2. Direct observations

A second source of evidence are (direct) observations. For some cases, online desk research does not yield sufficient information. Therefore, by attending events offered by the cases, direct observations are used as a supplement. In these situations, direct observations concern observations of meetings. Strengths of this evidence source are that information is offered in reality and in context (Yin, 2003). However, attending meetings is time-consuming (Yin, 2003), and not all cases offered events within the time frame of this thesis. Also, contrary to documentation, observational methods aren't unobtrusive. This entails that the information yielded may be influenced by the fact that the information sources know they are being observed. Yin (2003) refers to this as the weakness of reflexivity.

Notes taken in attendance of the meetings and any supplementary documents obtained at the meeting are summarized, stored and formatted as documents similar to the documents yielded by online

desk research. Written notes are summarized in paragraphs that cover a specific topic. For example, in the case of *Amsterdam Data Exchange*, a discussion was held among attendees following a topic list. The notes on these matters are summarized and rewritten to make a coherent paragraph regarding the specific topic. In Nvivo software, I created a document file that contains the summary of the meeting notes. As such, the notes are stored in the case study database and are available in a format that can be coded directly using the same software.

4.2.3. Interviews

A third source of evidence are interviews. Similar to direct observations, in the situation where online desk research does not yield sufficient information to create the case business model, structured interviews are conducted with case representatives. The strengths of interviews are that they are both targeted and insightful (Yin, 2003). On the other hand, interviews are vulnerable to bias in questions and response. Similar to direct observations, interviews aren't unobtrusive and reflexivity may occur (Yin, 2003).

In the application of interviews in this thesis, structured interviews are chosen. For this thesis, structured interviews are applicable due to their fixed, highly standardized order of questions (McLaughlin, 2006), which is complementary to the business model ontology that underlies the theoretical framework. To refrain from imposing bias in the questions, the interview is formatted with both open-ended and closed-ended questions. The set of questions represent the elements of each dimension in the theoretical framework. By doing so, the interview serves a dual purpose. On the one hand it fills information gaps on the case business model. On the other hand it creates secondary data points for information which was retrieved through online desk research. Table 4.3 provides the interview structure and the type of questions per business model element. These questions represent the core of each question which corresponds to each busines model element.

I extended requests for interviews with *Vivli*, *Amsterdam Data Exchange*, *City Innovation Platform*, *Dataverse* and *Social Science One* through the contact information published on the case websites. For the case *Humanitarian Data Exchange*, all information needed for creating the business model was retrieved already from online documentation and attendance of a meeting in February 2019. After this meeting, I spoke with case representatives who confirmed that the information on the case website (https://data.humdata.org/) is a correct representation of their organisation's information and business model. Due to time constraints for this thesis, I chose not to request an interview as well. For the other cases, I received responses from *Amsterdam Data Exchange*, *City Innovation Platform* and *Vivli*. In lack of a response, I sent reminders to *Dataverse* and *Social Science One*. *Social Science One* responded after the reminder, requesting to receive the questions to internally distribute them. After sending these questions, I did not receive a response too late to hold an interview within the remaining time for this thesis. Similar, from *Dataverse* I received a response two days before the thesis deadline, thus it was too late to conduct an interview and process the data prior to the deadline. *Vivli* indicated that all publicly available information is available on their website and in their report. *Amsterdam Data Exchange* replied with an invitation for me to attend a workshop on July 4th 2019 in Amsterdam (see sub section 4.2.2). *City Innovation Platform* did accept my request for an interview. For this case, the interview questions of Table 4.3 were specified for the case. For example, on their website the content refers to data users as "vragers naar data" (eng: requester of data). Furthermore, the questions were translated to Dutch, since the case representative is Dutch. The interview was conducted via phone, due to time and travel constraints. The interview transcript is stored and formatted in a similar fashion

to the online documentation, and notes and documents obtained from direct observations. As such, the (anonymized) interview transcript is part of the case study database.

Table 4.3: Interview questions.

Dimension	**Element**	**Interview question**
Value proposition	*Target-segment*	1. Do you focus on specific target segments for data users and data contributors?
		2. Who are contributors of data?
		3. Who are the users of data?
	Product-service	4. What products or services are offered to data users?
		5. What products or services are offered to data contributors?
	Intended-value-element	6. What is the added value of these products and/or services to users of data?
		7. What is the added value of these products and/or services to data contributors?
Value architecture	*Core-resource*	8. What resources are used to offer products and/or services?
Value network	*Role*	9. Are all resources self-developed?
		10. Are all products/services self-developed?
		10a. If not, what partners are involved?
	Governance	11. Is there interaction between data contributors and users of data?
		11a. If so, in what way is this interaction coordinated?
		12. How are possible conflicts in data transactions resolved?
Value finance	*Cost-revenue-structure*	13. Does your business have a profit motive?
	Pricing-method	14. Do customers pay for products and/or services?
		14a. If so, how are prices determined?

4.3. Content analysis

To enable data analysis, data needs to be inferred from the collected information generated by the three aforementioned sources. Content analysis is a method that uses a set of procedures to make replicable and valid inferences from textual data to the contexts of their use (Krippendorff, 2004; Weber, 1990). Hence, content analysis creates data from chosen procedures for answering specific research questions (Krippendorff, 2004).

With the objective of describing Trusted Data Intermediary business models in mind, the content analysis method is appropriate, because it allows for the observance of phenomena that cannot be observed directly (Krippendorff, 2004). In addition, it allows for classifying textual data in predefined categories that reflect the business model components. Other studies have already applied content analysis to describe business models (e.g., Al-Debei and Avison, 2010; Täuscher and Laudien, 2018). Furthermore, advantages of content analysis compared to competing choices are its applicability to a broad range of organizational phenomena, and that it allows analytical flexibility for studying both manifest content and latent content (Duriau, Reger, and Pfarrer, 2007), as well as unstructured or large volumes of data (Krippendorff, 2004). As well, it can be used to conduct inductive and deductive research (Duriau et al., 2007). Also, content analysis is unobtrusive (Krippendorff, 2004). Lastly, the method is cost-efficient and applicable to small-scale studies with minimal requirements, such as this thesis (Duriau et al., 2007).

In the design of the content analysis, I take a qualitative approach. This entails that the results will be interpreted on a qualitative rather than quantitative basis. The qualitative approach follows from the choice for a multiple-case study oriented for analytical generalization rather than statistical generalization. In other words, a quantitative technique for content analysis would require a larger, independent or random sample. Given the selected cases, such a sample is not feasible. Further, in this thesis, content analysis is driven by the research questions that follow from the research gaps concerning Trusted Data Intermediary business models. As such, the approach for conducting problem-driven content analysis as described by Krippendorff (2004) will be applied. While other approaches exist (e.g. Weber (1990) and McMillan, 2000 as cited Herring (2009)), these approaches are less detailed than the approach by Krippendorff (2004). The application of Krippendorff's approach is described in the following paragraph.

4.3.1. Questions and assumptions

The questions to be answered by the problem-driven content analysis concern the context of the content analysis. The purpose of the data analysis will be to construct and compare business models for each case. As such, the inferences made by the content analysis need to create data on these business models. Therefore, these questions differ from the thesis research questions as established in the Introduction (section 1.2.2). Based on these considerations, the research questions for the content analysis will concern the theoretical framework described in section 3.2 to explicate the context. For each case, the following content analysis questions need to be answered:

1. What constitutes the value proposition of this case?
    (a) What are the target segments? Who are data contributors? Who are data users?
    (b) What are the products-services offered to each target segment?
    (c) What is the intended value from offering these products and services?
2. What constitutes the value architecture of this case?
    (a) What are the core resources utilized to offer products and services to target segments?
3. What constitutes the value network of this case?
    (a) What are roles of network partners?
    (b) What mechanisms or contractual institutions are leveraged to govern the network?
4. What constitutes the value finance of this case?
    (a) What is the cost-revenue structure related to the provision of products and services to target segments?
    (b) What pricing methods are imposed on target segments?

I assume that what is stated or expressed in a (selection of) text will be directly correlated with an answer to one of the questions. For instance, a text that refers to or contains "data use agreements", is assumed to directly correlate with an answer to the question "What mechanisms or contractual institutions are leveraged to govern the network?" and thus takes part of answering the question "What constitutes the value network of this case?". Especially for the case *Amsterdam Data Exchange*, a second assumption is made. The case is still in development, which means that much of the case information is described on a more hypothetical basis rather than a definitive basis. As such, I assume

that what is provided in the documentation can be interpreted as the final plan. For instance, the text "Members can create agreements with algorithm developers to share data" is interpreted as "Members create agreements with algorithm developers to share data". By doing so, the same coding rules can be applied to this case as for other cases.

4.3.2. Collected data documents

The three information sources (online desk research, direct observations and interviews) generated relevant texts. Sections 4.2.1, 4.2.2 and 4.2.3 elaborate on the particular methods. In this paragraph, I elaborate on the specific data sources. The primary data collection method, *manual and automatized online desk research*, is applicable to all cases. Unfortunately only a selection of cases could accommodate a secondary method of information collection. To illustrate, *direct observations*, proved only possible for the cases *Humanitarian Data Exchange* and *Amsterdam Data Exchange*. Similarly, only the case *City Innovation Platform* was available for an interview. Table 4.4 lists all the sources of data documents collected in the multiple-case study, categorized by source type and in sequential order. Each data document is assigned a reference code, so that in the explanation of results examples of contents can be referenced easily.

Regarding online desk research, a tri-step, iterative approach was chosen. Firstly, contents from the case's web site is retrieved by downloading the page to a pdf-file. Each case has its own web site design, therefore the homepage differs per case. Table 4.4 explicates for each case which web site was accessed when. Second, when a document is published on the case's web site, this document is downloaded and stored in the case study database. Third, the Crawler program (see section 4.2.1) is executed for the queries specified in Table 4.4. Should after these three steps information still be missing, manual online desk research is performed again, mainly focusing on web pages or documents in formats which are inaccessible by the Crawler program.

A secondary data source was available for the cases *Humanitarian Data Exchange*, *Amsterdam Data Exchange* and *City Innovation Platform*. Regarding the first two cases, notes were taken from attending meetings. A meeting between TU Delft's Humanitarian Technology Lab and members of the HDX team was held on February 21st 2019 in The Hague, Netherlands. During this meeting, I took notes of what was being discussed, thus creating direct observations notes. Similarly, on July 4th I attended a Workshop held by *Amsterdam Data Exchange* in Amsterdam, Netherlands. The purpose of this meeting / workshop aimed at closing the exploration phase and kick starting the construction phase of the Amsterdam Data Exchange. As part of this meeting, working groups presented their findings. The corresponding presentation slides and supplementary documents were shared with me after the meeting. These documents thus serve as additional documentation on the case *Amsterdam Data Exchange*. For the case of *City Innovation Platform* I interviewed a case representative over the phone on on June 24th 2019. The transcript of the interview was anonymized, summarized and added to the case study database.

Table 4.4: Data sources with reference codes.

Case	Source type	Sources of data documents	Reference code
Social Science One (SSO)	*Online documentation*	1. *https://socialscience.one/*, accessed on May 31st 2019	SSO-1
		2. Paper by King and Persily (2019)	SSO-2
		3. Crawler program output for query "Social Science One"	SSO-3

Case	Source type	Sources of data documents	Reference code
Vivli	*Online documentation*	1. *https://vivli.org/about/overview/*, accessed May 31st 2019	Vivli-1
		2. Vivli 2018 Progress Report (Vivli, 2018)	Vivli-2
		3. Crawler program output for query "Vivli"	Vivli-3
Humanitarian Data Exchange (HDX)	*Online documentation*	1. *https://data.humdata.org/faq*, accessed on June 7th 2019	HDX-1
		2. Crawler program output for query "Humanitarian Data Exchange"	HDX-2
		3. *https://www.unocha.org/story/centre-humanitarian-data-four-takeaways-first-year*, accessed June 13th 2019	HDX-3
	Observations	4. Notes taken by the author at a meeting between TU Delft's Humanitarian Technology Lab and members of the HDX team on February 21st 2019 in The Hague	HDX-4
Amsterdam Data Exchange (AMdEX)	*Online documentation*	1. *https://towardsamdex.org/about/*, accessed June 12th 2019	AMdEX-1
		2. Report 'Toward an internationally trusted exchange of data' by Los (2018)	AMdEX-2
		3. Crawler program output for query "Amsterdam Data Exchange"	AMdEX-3
	Observations	4. Presentation slides of AMdEX work group Infrastructure, presented at AMdEX meeting July 4th 2019.	AMdEX-4
		5. Presentation slides of AMdEX work group "Ontwikkeling Datamarkten" (eng: data markets development), presented at AMdEX meeting July 4th 2019.	AMdEX-5
		6. Summary report of work group "Regelgeving" (eng: regulation), findings presented at AMdEX meeting July 4th 2019	AMdEX-6
		7. Notes taken by the author at "Workshop AMdEX", a meeting between the work groups and the coordination team of the Amsterdam Data Exchange, on July 4th 2019 in Amsterdam	AMdEX-7
City Innovation Platform (CIP)	*Online documentation*	1. *https://www.civity.nl/city-innovation-platform*, accessed June 12th 2019	CIP-1
		2. Crawler program output for query "Civity City Innovation Platform"	CIP-2
		3. *https://www.civity.nl/even-voorstellen-onze-partners*, accessed June 14th 2019	CIP-3
		4. *https://www.civity.nl/dienst-dataplatform*, accessed June 18th 2019	CIP-4
		5. *https://www.civity.nl/dienst-dataplatform-iot*, accessed June 18th 2019	CIP-5
		6. *https://www.civity.nl/dienst-datasafe*, accessed June 18th 2019	CIP-6
		7. *https://www.civity.nl/dienst-smart-city-inzicht*, accessed June 18th 2019	CIP-7
		8. *https://www.civity.nl/dienst-slim-melden*, accessed June 18th 2019	CIP-8
	Interview	9. Phone interview with case representative on June 24th 2019	CIP-9

Case	Source type	Sources of data documents	Reference code
Dataverse (DV)	*Online documentation*	1. *https://dataverse.org/about*, which was accessed June 12th 2019	DV-1
		2. Journal article by King (2007)	DV-2
		3. Crawler program output for query "Dataverse"	DV-3

4.3.3. Text samples and units of (content) analysis

In this thesis, I have chosen to sample texts generated by the Crawler program, to reduce the volume of text analyzed in the following steps. In the visual inspection of the located relevant texts generated by the Crawler program, I found that several web pages contain the same information, that is they hold identical paragraphs. For instance, the Crawler program yielded the same text data for the urls *socialscience.one/about-us#main-content* and *socialscience.one/#main-content*. As such, the latter is not used in the recording of units. Further, many of the texts yielded by the Crawler program for the case *Social Science One* contained the same sentence "Social Science One is an LLC operating on a not-for-profit basis". As such, recording all of these units of analysis would create duplicates that are not needed in the qualitative content analysis design. Rather, recording each of these units would be time-consuming. Therefore, in the sampling of the texts I have chosen to only consider "unique" units of analysis. In addition, the Crawler program yielded text from websites unrelated to the case. Mostly, these texts pertain to urls from Twitter or Facebook, and state general information such as "Embedded Tweets. bring your pick of content from Twitter into your website articles"[1]. These texts are therefore also not sampled.

To break the volumes of sampled text into more manageable bits of text (Krippendorff, 2004), units of analysis are chosen. For this thesis, sampling units are the web contents retrieved by the Crawler program based on their text-related tag. Other html-tags could have also retrieved information, but this information but these were not selected as sample units. For example, the web content related to a <*figure*>-tag, which specifies self-contained content[2], is not a sample unit. While, the web content that belongs to a <*caption*>-tag is, because this represents textual data that may be relevant. Other illustrations of sampling units for this content analysis are the interview transcripts, reports or papers, and meeting notes. I choose sentences as the recording unit of analysis. The syntactical meaning of the sentence are retained by selecting the sample unit as context unit: the paragraph in which the recording unit (i.e. the sentence) is located, as well as any other textual data in the sample. For instance, a (caption of) a table or figure may also serve as context unit.

4.3.4. Categories and coding rules

Categories and recording instructions are developed to ensure that the units of analysis are described in terms such that an answer to the content analysis questions can be created, and that the research is replicable. In this thesis, the definition of coding categories is done along the protocol for creating and testing a coding scheme for content analysis as suggested by Weber (1990). Although Weber (1990) argues for computer-assisted coding, I choose to code manually. The rationale for manual coding comes from the fact that Trusted Data Intermediaries is a novel research theme and this research takes an

[1] dev.twitter.com/web/embedded-tweets
[2] https://html.com/tags/figure/

exploratory objective. Computer-assisted coding would possibly miss any emergent themes from the textual data.

The categories are defined along the elements in the theoretical framework (see Table 3.2). As such, the elements are defined as nominal variables. The operationalization of these variables has been offered in Table 3.2. This operationalization rests on verbal designation of concepts. This is illustrated by the category for the *Governance* element, which offers examples such as *agreements, contracts, licenses* et cetera. By choosing a longer definition of a category, making distinctions in coding the units of analysis becomes theoretically motivated (Krippendorff, 2004). Coding for the categories was conducted using NVivo software. In this software, each category (i.e. element) is represented by a "node". Furthermore, the dimensions are represented as parental nodes to the corresponding elements. For instance, when a unit of text is "coded" for *target-segment*, it is assigned the node *target-segment*. Due to the child-parent noding, the unit of text is automatically also assigned the node *value proposition*.

Under the assumption that there is a one-to-one correlation between the text and the answers to content analysis questions, the following coding rule is applied: *if a recording unit (i.e. a sentence) corresponds to the description of an element, then the unit should be assigned to the element's corresponding node*. To illustrate, consider the following example of a unit of analysis: "The commission and partnering companies come to an agreement on the scope of a research project". The subject in this unit of analysis is *the commission and partnering companies*, while the object is *an agreement on the scope of the research project*. The object is a contractual institution that coordinates the subjects. As such, this unit is assigned the node that corresponds to this element, i.e. *Governance*. This element is part of the *value network* dimension. As such, this node should be assigned too.

The categorization was tested by coding and assessing a sample of text. The case *Social Science One* is used as the test case. In this thesis, coding is executed by one coder. This creates low accuracy and thus weak reliability (Weber, 1990). To address this matter, the coded texts were evaluated after the first assignment. The NVivo software allows for the inspection of coded text per node. From this inspection I found that in the value proposition, differences between data user and data contributor could be made more clear by distinguishing the *product-service* element and the *intended-value* per target segment. As such, the categorization was revised. Another insight from the test application is that some categories proved not to be mutually exclusive. Krippendorff (2004) emphasizes that categories need to be mutually exclusive for the purpose of semantic validity. However, it is inherent to the theoretical framework that dimensions are substantially interrelated and interdependent (Al-Debei and Avison, 2010). Considering the example above, the unit of analysis describes both a *core-resource* and *governance* element. This is because the governance (i.e. the agreements) is created using a resource (i.e. the commission) of Social Science One. Ways for dealing with this could be to assign the node that corresponds to the subject of the sentence. However, in the test application it became clear that for instance the word "agreements", which corresponds to the node *Governance* are often the object of a sentence, rather than the subject. As such, many of the units of analysis that inform on the *Governance* element, would be excluded from this category. The same applies when coding rules would be focused on the object of a sentence, or another syntactical structure. Taking these considerations into account, the coding rule is revised to: *if a recording unit (i.e. a sentence) corresponds to multiple descriptions of elements, then the unit should be assigned to the elements' corresponding nodes*. After the revision of the categories, another sample of text was test coded. The case *Vivli* was chosen as the second test case. In this test application, it was found that the categories were not fully exhaustive to variations to the terms in the dictionary. Therefore, the phrase "or variations to these terms" was adopted in the coding logic.

All textual data was coded per case along the final coding logic represented in Table 4.5. Similar to the test cases, the coding was done using NVivo software, where the dimensions and elements of the theoretical framework are represented by nodes. Figure 4.3 provides an example of how sample text was coded. Subsequently, the consistency in application of coding rules was assessed across cases. For each category, the coded references were inspected for consistency. For example, whether sentences describing the profit motive of the case is coded under "cost-revenue-structure" instead of "pricing-method". The result of these inspections is that the coding rules were consistently applied.

Table 4.5: Coding logic.

If the unit of analysis describes, refers to, or contains...	**Then assign the unit to node...**
An entity that shares, opens, or provides access to data or variations to these terms	*Target-segment: data contributor*
The offering for or receiving, obtaining, getting or variations to these terms of a product, service, tool or variations to these terms by a data contributor	*Product-service: data contributor*
Utility, gains, benefits, opportunities, possibilities or variations to these terms for, or the enabling or connecting or variations to these terms of a data contributor	*Intended-value-element: data contributor*
An entity that uses, downloads, processes or gains access to data or variations to these terms	*Target-segment: data user*
The offering for or receiving, obtaining, getting or offering or variations to these terms of a product, service, tool or variations to these terms by a data user	*Product-service: data user*
Utility, gains, benefits, opportunities, possibilities or variations to these terms for, or the enabling or connecting or variations to these terms of a data users	*Intended-value-element: data user*
The organizational assets deployed to create value, e.g. financial capital, software, digital infrastructure, staff, processes, systems, trust, norms and values in the network, or variations to these terms	*Core-resource*
The mechanisms or contractual institutions leveraged by the Trusted Data Intermediary to coordinate data contributors, data users and network partners, e.g. contracts, agreements, licenses or variations to these terms	*Governance*
The provision of resources, execution of processes or variations to these terms by an entity which are leveraged (or variations to this term) in the process of value creation for data users and data contributors	*Role*
The arrangements that lead to revenues for the Trusted Data Intermediary from offering products and services or receiving of funds or grants or variations to these terms, and costs incurred from the deployment of resources	*Cost-revenue-structure*
The fees, charges, subscription, costs of use or variations to these terms that are or aren't imposed on data contributors or data users	*Pricing-method*

4.3.5. Analytical constructs and procedure for inference

Analytical constructs operationalize what I know, suspect or assume about the context of the text. They represent a model of the relationship between the text and what it is that we need to know about that context (Krippendorff, 2004). In a qualitative analysis, the description of analytical constructs is more

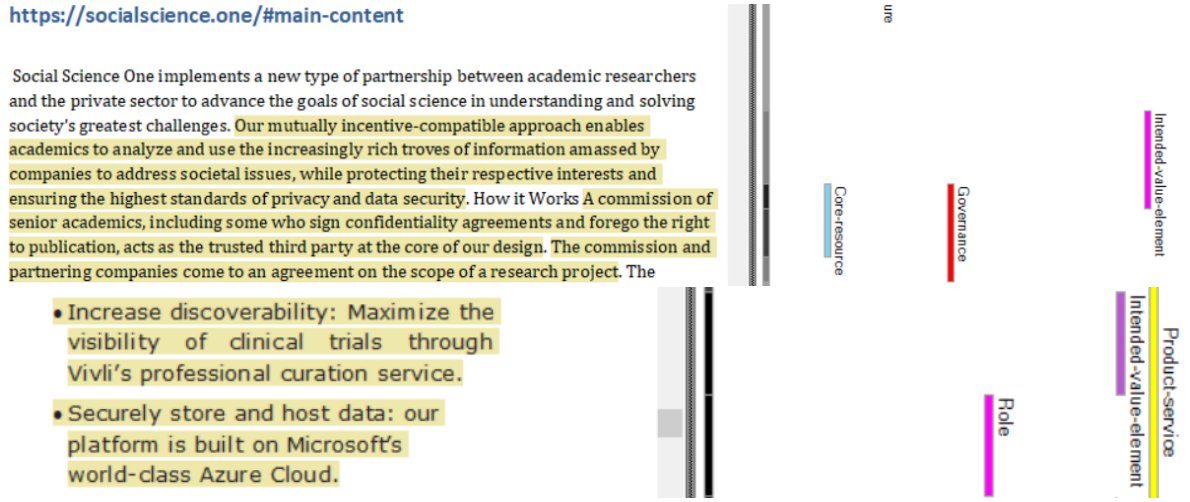


Figure 4.3: Examples of text coding applications for *Social Science One* and *Vivli* cases.

ambiguous than in a quantitative analysis, for in the latter equations and such are easily applied to each data point. Nonetheless, I attempt to describe my analytical constructs for qualitative content analysis as clear as possible by the means of the following illustration. Consider the context of the value architecture of a case. Here, the question "What are the core resources utilized to offer products and services to target segments?" is posed. To answer this question, an analytical construct is needed that relates the text "A commission of senior academics, including some who sign confidentiality agreements and forego the right to publication, acts as the trusted third party at the core of our design" to an answer to the question. In this example, the analytical construct comprises that *the commission of senior academics* is the same as *staff*. This I learnt from the inspection of the other units of analysis to the same node. Thus, the corresponding analytical procedure is that units of analysis that contain or refer to a *commission* should be interpreted as that they refer to *staff*. Due to the fact that analytical constructs can differ for the specific context (Krippendorff, 2004), I have created a list of analytical constructs as applied in this content analysis in Table 4.6. Especially for the *intended-value-element*, I found that many of the units of analysis described value in such a way that it did not need further generalization. For instance, consider the following unit of analysis found in the Vivli 2018 report: "Increase efficiency through harmonization: Take advantage of Vivli's harmonized policies and streamlined processes" (Vivli, 2018, p. 5). This unit of analysis thus directly describes value in the form of increased efficiency. The same applies to multiple other units of analysis (see Figure 4.3). Therefore, less analytical constructs were created here.

Table 4.6: Analytical constructs per business model element.

Content analysis question	**Element**	**Analytical constructs**
What are the target segments?	*Target-segment*	Private company / corporation / commercial parties = private-sector data contributor or user Government / municipality / = public-sector data contributor or user Individual / citizen / inhabitant (NL: inwoner) / developer / journalist = civilian data contributor or user Researcher / journal / research center / scholarly organisation = academic data contributor or user
What are the products-services offered to each target segment	*Product-service*	Customer training / support = customer service Research environment and (analytical / aggregation) tools = data analysis product Search engine / search filter / query-building tool / data grid = data search product Data repository / catalog / hosting = data storage product/service Infrastructure / APIs connection = data logistics service Data organisation / management / anonymization / evaluation tools = data management service Request form / following data sets = data access service Metadata / data set statistics / persistent identifier / dashboards / data reports / data visualizations = data informative product Grants / funds = financial product Peer pre-review feedback service = data publishing service Publishing free from pre-approval = data publishing service Data citation / accreditation = data publishing product Messaging tool / option = platform communication product
What is the intended value from offering these products and services?	*Intended-value-element*	Finding new value / creating possibilities / opening doors / room for innovation (NL: ruimte voor innovatieve oplossingen) / development new solutions / shape market = achievement of new possibilities Reduced workload / data aggregation = increased efficiency Enlist support / work together / cross-sectoral = creation of collaboration possibilities Improve business / new insights / better decision-making possibilities (NL: betere sturingsmogelijkheden) / data-driven decision-making (NL: datagedreven sturing) = improved decision-making Produce social good / advance societal good / increase data impact / create solutions for social issue = enabling public value creation Show data availability / data existence / make data discoverable = increased data discoverability Decide what to share / remain ownership / exercise control / have grip on data sharing (NL: grip op data delen) / ensure security / protected interests = increased security in data sharing Gain academic credit = increased visibility Easy to use / simplicity / familiarity / harmonization / reduce time = eased processes Access to data / research independence = increased data accessibility

Content analysis question	Element	Analytical constructs
		Search features / capabilities / query = increased data findability
What are the core resources utilized to offer products and services to target segments?	*Core-resource*	Commission / team = staff Infrastructure / platform / API / = technical system for the data logistics Data management system = technical system for data contributors to manage their data Data storage facilities / repository = technical system where data is stored or hosted Tools / search engine = technology for creating or accessing specific information in the technical systems Software / code = technology that powers products and services on a technical system and delivers this to customers
What are roles of network partners?	*Role*	Partner provides user support / administers / manages / executes processes = partner provides replaceable resources Entity provides funds = partner provides replaceable, possibly indispensable financial resources Partner contribute to software development = partner provides replaceable resources Partner participates in design of products/services = partner provides replaceable resources Partner manages the case organisation = partner is a parent organisation Partner provides hosting / data storage systems = partner provides indispensable, irreplaceable (due to customization) resources Partner powers technical systems = partner provides indispensable resources
What mechanisms or contractual institutions are leveraged to govern the network?	*Governance*	Market rules = rules imposed on data transactions and market participants Data (contributor/ use/ exchange) agreements = governance of data transactions through agreements Waiver / licenses (NL: licenties) = access / data use permissions for data users Norms = regulation of how products/services may be used Terms (of use) / conditions = rules that customers accept or agree to as conditional to the use of products/services Verification / review = series of steps to ensure compliance with rules and/or agreements Target segments can/cannot/must/should/are (not) allowed… = regulation of how products/services may be used Procedures / request forms = implementation of rules and/or agreements
What is the cost-revenue structure related to the provision of products and services to target segments?	*Cost-revenue-structure*	Funds / grants = financial resources from network partners Not-for-profit basis = non-profit motive Commercially oriented = for-profit motive Expenses = costs incurred from offering products/services
What pricing methods are imposed on target segments?	*Pricing-method*	Free = price of products/services = customers do not pay for products and services Commission = price of products/services = customers pay for products and services Cost to share data = price of products/services = customers pay for products and services Contributions = price of products/services = customers pay for products and services

4.4. Within-case & cross-case analysis

The multiple-case study research design allows for both within-case analysis and cross-case analysis (Eisenhardt, 1989). As such, each case can be analysed in separately (within-case analysis). As well, the case results can be compared in a cross-case analysis where differences and similarities in pairs of cases are discussed.

4.4.1. Within-case analysis: case business model development

Within-case analysis is the analysis of a case as a stand-alone entity (Eisenhardt, 1989). It is applied to cope with the volume of data, to identify unique emergent patterns in a case, and to gain familiarity with the case which will aid in the cross-case analysis later on (Eisenhardt, 1989).

The familiarity with the case creates tacit knowledge on how business model dimensions are interrelated and dependent. This tacit knowledge is used to infer relative importance of business model components. However useful, tacit knowledge is difficult to explicate, thus clouding the procedure of analyzing the data. For instance, in a quantitative approach, one may choose to use frequency as an indicator for the importance of a resource. However, in a qualitative approach, the sample is not representative enough to assess importance of a concept in this manner. Rather, the contextual units will play a role in assessing the importance of a concept. To illustrate for the case of *Vivli*, the phrase "The Vivli platform includes an independent data repository, in-depth search engine and a secure research environment", and the analytical construct "Infrastructure / platform / API / = technical system for the data logistics" indicate that the platform is a resource. However, the criticality of this resource to Vivli's business must be derived from the knowledge that "independent data repository, in-depth search engine and a secure research environment" are products and services offered to the customers, which were found by other units of analysis and analytical constructs. From the example above, it becomes clear that the analytical constructs from Table 4.6 are instrumental for making inferences on business model components, as well as making inferences on components' relative importance. For each case, a business model is constructed based on these inferences. The business models are created along the content analysis questions established in section 4.3.

4.4.2. Cross-case analysis: comparison along case characteristics

After each case has been analysed separately, cross-case analysis can be conducted. The rationale for cross-case analysis lies with the idea that humans are poor processors of information and therefore may be impeded in analysing proper emergent patterns (Eisenhardt, 1989). Cross-case analysis addresses this concern by juxtaposing cases. This allows for looking at the data from different perspectives, thus challenging potential bias, improving the likelihood of accurate and reliable theory and the probability that novel findings will be captured by the analysis (Eisenhardt, 1989). Eisenhardt (1989) describes three tactics for cross-case analysis. The first tactic relates to within-group analysis: cases in the same category or dimension are compared and similarities and differences are listed. The second tactic relates to pair-wise comparison and listing of similarities and differences. The third tactic divides data by data source, rather than a case categorization.

The first tactic for cross-case analysis is selected. Because not all cases have multiple, distinct data sources, the third tactic described by Eisenhardt (1989) is not applicable. While the second tactic is

applicable, the pair-wise comparison for all six cases would result in 15 comparisons. Given the time constraints and diversity of selected cases, this tactic is less preferred than the first tactic. In this thesis, cases were initially not selected to represent categories or dimensions. However, as demonstrated by Table 4.2, cases can be non-profit based (e.g. *Vivli*). Further, some cases focus on specific data domains (e.g. *Social Science One*, *Vivli*, *Humanitarian Data Exchange* and *Dataverse*). In adoption of the first tactic for cross-case analysis described by Eisenhardt (1989), the different findings per sub research questions for cases will be compared. I will attempt to explain the similarities and differences by the means of case characteristics in section 6.5.

4.5. Methodological limitations

Multiple research methods are leveraged in the attempt to reach the research objective. These methods all incur inherent limitations, which can impede on the quality of research results. As such, this section is devoted to describing the limitations of the methods regarding the research design, data collection methods, content analysis and data analysis methods.

4.5.1. Limitations of the exploratory multiple-case study method

Following the method of Yin (2003), the quality of an exploratory, embedded multiple-case study research design can be assessed on three criteria. For the research design to be deemed of good quality, it needs to guarantee *construct validity*, *external validity* and *reliability*. Firstly, construct validity pertains to the establishment of correct operationalization of concepts (Yin, 2003). In this research design, the studied concepts are the business model dimensions in the theoretical framework (see Table 3.2). The operational measures for these concepts are captured by the web content analysis approach described under section 4.4. In particular, the coding rules in Table 4.5 and the analytical constructs in Table 4.6 provide the operationalization of the concepts. To maximize construct validity, a chain of evidence is ensured through a systematic data collection procedure applied to each case. Table 4.4 discusses which documentation was accessed in what order. A drawback of this research design is the use of a single source of evidence for half of the cases (the exceptions are *Humanitarian Data Exchange*, *Amsterdam Data Exchange* and *City Innovation Platform*). Given the time limit for conducting this thesis and the considerations of availability of data, online documentation was chosen as the primary source of evidence. Because other potential sources were not used for all cases, data is not always triangulated and construct validity is therefore not maximized for all cases.

Second, external validity relates to establishing the domain to which the analytical generalization can be applied (Yin, 2003). Yin (2003) argues that a replication logic will maximize external validity in multiple-case study designs. However, as described in subsection 4.1.1, this thesis has opted for the theoretical sampling approach rather than replication logic. Thus, according to Yin (2003), external validity is not maximized. On the other hand, the cross-case analysis based on emergent case characteristics (see section 4.4.2) establishes to which domains the findings are applied. Nonetheless, external validity should be a point of attention in future research.

Last, reliability refers to demonstrating that errors and bias are minimized so that the procedures applied in data collection and analysis can be repeated and will produce the same results (Yin, 2003). Following Yin's suggestions, a case study protocol is constructed (see Appendix A). Also, by coding the text using NVivo software, a case study database is constructed that contains the data and the coding

of the data as applied in the data analysis.

4.5.2. Limitations of data collection methods

Regarding data collection methods, the research design is limited in several ways. Firstly, the single investigator research design has implications for the reliability of the information obtained by direct observations and interviews. Especially in the situation where no hand-outs or audio recordings are available, the notes are subject to the context interpretation of a single investigator. Further, the single investigator design limits the volume of online documentation accessed. For instance, the Crawler program was limited to retrieve 10 initial urls, gather the hyperlinks on these pages and stop. In a multiple investigator design, the Crawler program could run for more web pages, or with more consecutive loops as the volume of resulting data could be assessed by more investigators.

In addition, the online documentation collected through manual online desk research and automatized desk research may overlap. This entails that information is redundant and more work is to be done for one coder. This became apparent in the content analysis where texts were sampled for relevance and uniqueness.

Furthermore, web content is dynamic and change over time (Kallinikos, Aaltonen, and Marton, 2013). A way to cope with this limitation is to take "snapshots" that create new objects with limited interactivity. This ensures provenance of the web content (Kallinikos et al., 2013). To illustrate, the contents of a web page may be printed to a physical object, or downloaded as a digital object in the form of a PDF-file. In manual data collection, I've chosen to download web pages to digital, ineditable objects in PDF-format, assigned with the web page url and a timestamp for the date of access. Subsequently, I stored the objects in the case study database. In addition, for implementation of this approach in automatized data collection, I've designed the Crawler program such that textual data is stored per origin url. Further, the Crawler program creates a txt-file with the setup of the program. This file contains the search query string, the timestamp of when the program was run and the amount of results requested from the Google API (see Figure 4.4).

Most importantly, the implementation of automatized online documentation collection in this thesis has several implications. To begin, the web crawling strategy in this thesis entails that the program only selects web pages that are relevant to the case. This is done through the use of the "googlesearch" module for Python. This module connects to the Google API and will search Google for a given query string (Vilas, 2018). This technique is limited in the sense that the Google search engine influences the results obtained by the Crawler program. First, the search results contain potential bias incurred by the applied search algorithms. For instance, the algorithm may amplify already popular pages, or promote undesirable material in a lesser fashion (Kallinikos et al., 2013). Secondly, the search ranking may be influenced by third parties that offer products to web page owners that increase the likelihood of a high ranking in search results (Kallinikos et al., 2013). In this sense, the Crawler program contains a "black box" component where the exact operational steps are unknown.

4.5.3. Limitations of content analysis

Content analysis has been applied to infer the information necessary to analyze from the various data sources. These inferences need to be replicable and valid to the context of Trusted Data Intermediaries' business models. While these concepts overlap largely with the criteria for the case study research

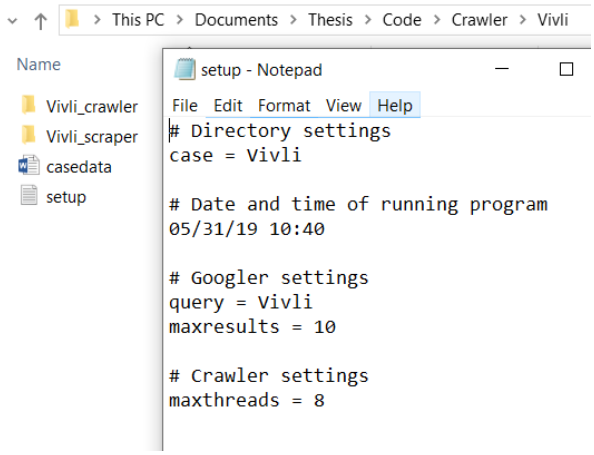


Figure 4.4: Storage of Crawler program settings for the Vivli case.

design, they have particular demands for content analysis (Krippendorff, 2004). Thus, I address them separately here.

Following Krippendorff (2004), in order for the analysis to be reliable a content analysis needs to be reproducable and validatable in principle. A content analysis is considered reproducable or replicable when the manner in which inferences are made from textual data, as well as the yielded results are replicable elsewhere. As such, reproducability or replicability is a form of reliability (Krippendorff, 2004). For this thesis' content analysis, reliability was assessed in the creation and testing of the coding categories along the protocol by Weber (1990). From this it followed that the single-coder design led to intracoder consistency in the application of coding rules for this specific thesis. Although the intracoder consistency caused that the stability and accuracy of coding was deemed sufficient, reliability was not maximized. Due to constraints in time and resources, intercoder reliability tests could not be conducted. In addition, due to the limited amount of data for each case, reliability data was not available. While reliability tests are absent, I argue that the research is reproducable in principle, based on the precautions taken for increasing the replicability of the content analysis research design. For example by explicating the data sources (Table 4.4), coding rules (Table 4.5), analytical constructs (Table 4.6) as well as the development of a case study database and case study protocol (Appendix A).

Further, every content analysis ought to be validatable in principle, in order to contribute to literature on the matter (Krippendorff, 2004). While reliability focuses on results being trustworthy based on how they were derived, validity refers to results being deemed true and able to uphold in other situations or in light of new evidence (Krippendorff, 2004). As such, validation refers to the justification of research results (Krippendorff, 2004). The content analysis in this thesis is validatable in principle. For instance, in the application of Krippendorff (2004)'s problem-driven content analysis approach, standards for validity were created. These standards limit uncertainties regarding the nature of the context, the extent of knowledge on text-context relations and how the analysis is conducted (Krippendorff, 2004). Firstly, the nature of the context is structured by the business model ontology described in section 3.2. This framework informs on the text-context relations, as the categories in this content analysis were constructed along this framework. The framework is a derivative of a well-established theory on business models, which supports semantic validity of this content analysis. Furthermore, the theoretical framework as used in this thesis has been applied to other academic researches on intermediary entities. As such, the degree to which the categories accurately describe meanings and uses in the chosen context (i.e. semantic validity) is deemed sufficient. However, since those prior

researches (i.e., Janssen and Zuiderwijk, 2014; Ranerup et al., 2016) did not explicate their analytical constructs, the degree to which the chosen analytical constructs are appropriate (i.e. the functional validity) remains uncertain. The third uncertainty relates to internal validity that mirrors how the content analysis design leads to results. The coding phase is most vulnerable to this uncertainty (Krippendorff, 2004). Precautions were taken to enhance validity in this phase through increasing reliability, for instance the iterative creation and testing of the coding categories, the explication of the applied analytical constructs, as well as the development of a case study database in NVivo. In addition, because the multiple-case study has been designed such that quantitative analysis methods can be adopted in the research design, quantitative validation by future research is made possible.

4.5.4. Limitations of within-case and cross-case analysis

Regarding data analysis methods, a qualitative approach is chosen for the application of coding rules and interpretation of the results. This choice was based on the availability of data. For quantitative analysis, a larger sample of data documents would be required to infer results. Due to the novelty of the cases, the amount of available data is limited. Several implications of this choice have already been addressed, i.e. reliability, construct validity and external validity. In this sub section, I discuss three more limitations for data analysis in a multiple-case study research design.

To begin, as addressed in section 4.4.1, tacit knowledge of the case is needed for making inferences from textual data on how important a certain concept is relative to others. There is no straightforward equation that indicates whether a resource, product, service, partner or financial arrangement is more or less important than another. Although this is an inherent limitation of the qualitative data analysis approach, it is also a strength. Because, due to the familiarity with a case, new insights may emerge (Eisenhardt, 1989).

Secondly, Yin (2003) argues that a high-quality case study can only be conducted by a well-trained and experienced investigator. In this respect, the research design is limited by my lack of experience with qualitative research. Although I have conducted several case studies before, they mostly demanded quantitative data analysis. In response to the limitation that I impose on the research design, I've put an effort into applying a methodological approach to each part of the research design. In addition, by creating a case study protocol (see Appendix A), and carefully going through the steps of selecting cases, collecting data, and creating and testing the coding rules and categories, I hope to have minimized the errors in doing qualitative data analysis. Unfortunately, the coding and interpretation is not performed by multiple coders. This implicates that my work is not checked by a different party. As such, my skill-level may further impede on the accuracy and reliability of the coding (Weber, 1990).

Another limitation is that data documents comprise of two languages: English and Dutch. This is a result of the selected cases, i.e. *City Innovation Platform* is a Dutch business, and communicates mostly in this language. The differences in Dutch and English language complicate a consistent application of coding rules. For instance, *City Innovation Platform* discusses "betere sturingsmogelijkheden", which directly translates to "better control options". Rather, in my interpretation it is translated to "better decision-making", due to its context. As such, the level of subjectivity increases as a result of translating the data to English. The explication of the analytical constructs as applied in this thesis in Table 4.6 addresses this limitation.

5

Cases' business models

This chapter describes the business model inferred for each case. Besides an explanation in text of the constituent elements, examples of coded contents are provided that illustrate these elements. Each example is provided with the reference code of the corresponding data document, as provided in Table 4.4. At the end of the chapter, Table 5.1 tabulates the results for each case's business model dimension to ease cross-case comparison. Based on the cases' business models core elements, differences and similarities are discussed in the following chapter.

5.1. Social Science One

The first case analysed is *Social Science One* (SSO). For this case, online documentation content analysis (see Table 4.4) yielded the following business model input. To begin, SSO's target segments consists of independent academics (data users) that use data provided by private companies (data contributors) to answer research questions in the social science domain. Data users are offered eased processes for access to relevant, secure and privacy-preserving private sector data: "Even the formal "data use agreement" signed by the company and researcher's university, is pre-negotiated in a generic way by the commission so that researchers do not need to begin this often arduous process anew each time" (SSO-2). Also, eased publication of research results is offered due to the absence of pre-publication approval by the data contributor and the offering of a peer pre-review service. In addition, advice and support by respected senior academics in specific areas, as well as research funding is offered. Data contributors are enabled to improve their business decisions, while creating public value: "This novel approach enables companies to enlist the support of the scientific community to help them produce social good, while protecting their interests and improving their business" (SSO-1, SSO-3) and "This data enables companies to make better decisions by building on research and methods from the social sciences and other fields" (SSO-2). In support of this, both systems for secure data storage and a dedicated team of specialists is offered.

The primary resource of SSO is a group of people referred to as a commission: "A commission of se-

nior academics, including some who sign confidentiality agreements and forego the right to publication, acts as the trusted third party at the core of our design" (SSO-1). As such, the commission is primarily tasked with the protection of the interests of both parties: "The role of the commission at Social Science One is to simultaneously protect the company, the public, and academia" (SSO-2). Furthermore, the commission is both tasked with and subject to governance mechanisms, as some commission members may sign confidentiality agreements with the data contributor. At the same time, the commission oversees the compliance of the data contributor to the data use agreements between data contributors and data users. To illustrate, "If the company breaches this agreement in any way, such as withholding relevant information, the commission has the right and the obligation to publicly report the violation" (SSO-3). In addition, decision-making power on awarding grants lies with the commission due to their familiarity with the contributed data and agreed upon boundaries. On the other hand, this decision-making process is subject to review by independent ethicists. Other resources concern data storage systems, evidenced by "This has involved an enormous amount of effort: creating entirely new systems for securely storing the data, incorporating cutting-edge differential privacy and other techniques, and building a dedicated team of specialists" (SSO-3).

Agreements are the main governance mechanisms used to set boundaries to network actors' influence on the process of data exchange. Other mechanisms such as rules and procedures as part of the agreements. To begin, agreements are made between certain commission members and the data contributor in which the commission members, their responsibilities, the commission's access to company data and their confidentiality boundaries are established. Second, agreements are used where the data contributor agrees to the scope of the research project. Also, provision of access to specific data, as well as forfeiting the right to pre-publication approval as exemplified in a current project: "Facebook has also agreed to relinquish rights to pre-publication approval of research results within the scope of the project" (SSO-3). Similar, data users agree to follow specific rules of engagement: "Scholars legally agree to follow the rules and the rest of the community hopes they comply, to one of collective responsibility, where multiple people are always checking and the risk of improper actions by any one individual is greatly limited" (SSO-3). Further, data users agree to data access rules, replication standards and to go through review processes and audits, as demonstrated in "Social Science One said outside researchers will have to pass a "rigorous peer review process" and special "ethical review," conducted by the Social Science Research Council, an independent group led by Alondra Nelson, a sociology professor at Columbia University" (SSO-3). Agreements are also used for partner actors, specifically the non-profit foundations that provide funding for research projects: "The foundations have also agreed to the following additional processes to protect the academic integrity of our decision making" (SSO-2). Other partners in the research proposal review processes concern academic peers, who evaluate proposals based on scientific and social merit, as well as the researcher's merits and potential costs of the research. Last, SSO partners with ethics reviewers: "Finally, because of the ever-changing nature of ethical understandings, we are enlisting researchers in ethics to study the commission's decisions, how they are viewed by other academics and the general public, and how they might be improved" (SSO-2).

SSO does not have a profit motive, as is demonstrated by "Social Science One is an LLC operating on a not-for-profit basis" (SSO-1, SSO-3). The primary finance mechanisms include the pooling of funds, provided by partners. It is not excluded that data contributors may be charged or asked to provide funding in the future, as demonstrated by "The only real difference would be that the company may fund the commission, consultants, and outside experts directly" (SSO-2). An independent partner actor redistributes these funds as grants for data users and to cover the operational costs of SSO: "To be clear (you can never be too clear when funding is involved), the foundations put their money into SSRC's Social Data Initiative, from which shared fund it is then distributed both to cover Social

Science One's operations and the grants" (SSO-3). These operational costs include the payment of the commission members: "Commission members signing restrictive confidentiality agreements are compensated at fair market rates, and, in highly charged partisan or otherwise sensitive environments, are paid by nonprofit foundations independent of the firm" (SSO-2).

5.2. Vivli

The business model of *Vivli* was constructed based on the contents analysed from online documentation (see Table 4.4). In the value proposition dimension, Vivli's target segments consist of researchers in academic, industry, foundation and non-profit entities that have completed clinical trials (data contributors). At the same time, the data user segment consists of researchers in the international research community. Data contributors are offered services for secure data hosting, tools for anonymization and mapping of data and the service of reviewing data requests on their behalf. Data users are offered a data search engine, a form to request data with data contributors, as well as an environment and tools for data aggregation (from various sources) and analysis of this data: "Flexibility: Researchers may bring in their own data sets, statistical software, and scripts to use on an individual Secure Research Environment within the Vivli platform" (Vivli-2). Both data users and data contributors are offered customer support services. For data contributors, value in the form of increased efficiency and security of data hosting and sharing is offered: "The Vivli platform was designed and built to ensure the security of the anonymized clinical trial participant data that it hosts" (Vivli-1). As well, discoverability of data is increased, which makes increased impact of data possible, thereby creating more public value in the sense that researchers can advance clinical science. For data users, value lies in increased findability of data, harmonized processes for requesting access to data, security in sharing and analyzing data, and possibilities for aggregation of data from different contributors.

Vivli's core resource is the digital platform, including tools for the aggregation and analysis of data. The platform "includes an independent data repository, in-depth search engine and a secure research environment" (Vivli-3). Besides the platform, Vivli has staff for the customer support and data request review process. The platform, along with other technical resources like solutions for disclosure, security, data export and the search engine, are powered by partners. Furthermore, one of the partners optionally conducts reviews of data request forms: "Wellcome Trust collaborates with Vivli to act as the secretariat for the data sharing platform's Independent Review Panel (IRP)" (Vivli-3).

Vivli's uses agreements for data use and data contribution: "Total number of signed data use agreements (View details on Research Proposals) 5 Data use agreements in process 6 Data use agreements withdrawn, agreement not reached, no response 0" (Vivli-3). Before a data contributor can submit data, it is required to sign a Harmonized Data Contributor Agreement. Similarly, before a data user can request data, it is required create a profile and to sign a Harmonized Data User Agreement. Additionally, there are data request forms, which are reviewed by either Vivli' staff or those of a partner. As indicated before, partners power or provide Vivli's resources: "We have partnered with some of the biggest supporters of clinical trials to architect the Vivli platform to increase the impact of the trial participants" (Vivli-3). Partners also provide funding.

Vivli is a non-profit organization whose income originates from funding by partners. In addition, Vivli support a pricing model that charges data contributors and data users for using Vivli services (the search engine excluded): "The cost to share data on Vivli is similar to the cost for publication, according to Li. It is approximately $2,000 to share a single dataset if the contributor does not require a review

panel. User cost is "nominal" and does not kick in until after about 1 year, depending on configuration, Li said. It is always free to search and request data" (Vivli-3).

5.3. Humanitarian Data Exchange

Information for the business model of *Humanitarian Data Exchange* (HDX) was retrieved from online documentation and attendance of a meeting (see Table 4.4). The HDX target segments are focused on humanitarian aid workers (data users) and organisations that hold relevant data for humanitarian aid workers (data contributors). In principle, any entity may contribute or use data, but the quality of the data is reviewed to ensure relevance to data users: "Any organisation that has humanitarian data can request to join HDX" (HDX-4) and "We ask you to submit the following information: an organisation name, description and link to an organisation-related website (optional). We review this information and then either accept the request or ask for more information, such as a sample dataset" (HDX-1). Data contributors are offered services and tools that host data, impose access controls for published data sets, create metadata, and standardize, refine, statistically analyze and visualize data. For instance, "Data Check automatically detects and highlights common humanitarian data errors including validation against CODs and other vocabularies from your HXL-tagged spreadsheet" (HDX-1) and "If your data uses HXL hashtags, then the Quick Charts tool can automatically create customizable graphs and key figures to help you highlight the most important aspects of your dataset" (HDX-1). These offerings create value for data contributors by increasing discoverability of available data, even when data sets are not published to the platform yet: "Let others know your data is available by publishing your metadata without uploading any file(s) via HDX Connect. Once users request access, you decide what to share" (HDX-1). While simultaneously remaining control over who can access the data. For data users, services and tools for searching, following and requesting data are offered, as well as an API infrastructure to integrate the platform into user-own developments. The offerings increase the timeliness of identifying and requesting relevant data: "The Data Grid provides a quick way to find datasets that meet or partially meet the criteria for a set of core data categories, like internally displaced persons and refugee numbers, conflict events, transportation status, food prices, administrative divisions, health facilities, and baseline population" (HDX-1).

The primary resources of HDX consist of the platform and the HDX team. The platform operates on the CKAN open-source software and deploys several technological features: "HDX is based on CKAN, an open-source data management system for powering data hubs and data portals" (HDX-2). The features are developed by the HDX team, who also verifies data contributors, and evaluate contributed data sets for data quality and sensitivity levels: "Data quality is important to us, so we manually review every new dataset for relevance, timeliness, interpretability and comparability. We contact data contributors if we have any concerns or suggestions for improvement" (HDX-2).

The HDX value network consists of HDX's managing organisations, the UN OCHA's Center for Humanitarian Data: "HDX became part of the Centre's data services workstream" (HDX-3). The Center manages HDX and supplies staff. In addition, partners are enlisted for the design and testing of technical features: "Thank you to the partners who participated in the design research and provided feedback on initial versions of the Data Grid. And special thanks to our partners at Oblo for their work on the research" (HDX-2). Last, the network has partners that provide funding. The governance of the HDX platform comprise of terms of use. Further, software-embedded rules are used that design roles for who can share or edit data sets: "The requestor can not specify the role (i.e., admin, editor or member). Instead, the person receiving the request assigns the role" (HDX-1). Further, there are rules for which

data may be shared and how: “HDX does not allow personal data or personally identifiable information (PII) to be shared in public or private datasets” (HDX-1). As well as rules for sharing under which licensing (Creative Commons): “HDX promotes the use of licenses developed by the Creative Commons Foundation and the Open Data Foundation”. There are also reviews executed by the HDX team. These processes comprise of the evaluation of data contributors, and the evaluation of data sets’ sensitivity levels: “However within one day, an HDX team member will evaluate the dataset to determine if it sufficiently meets the criteria to remain in the Data Grid” (HDX-2).

The HDX platform’ services can be used free of charge. Through its managing organisation, The Centre for Humanitarian Data, HDX receives funding from Dutch governmental institutions: “The Centre receives funding from the Netherlands and the city of The Hague” (HDX-3).

5.4. Amsterdam Data Exchange

The information on the *Amsterdam Data Exchange* (AMdEX) business model originates from online documentation and documentation gathered from attending a workshop (see Table 4.4). The information from the workshop validated and elaborated on the online information. Most AMdEX document content are written in future tense as the project is still in development. For instance, “Such a trusted market will offer new opportunities for cooperation that each of the market participants are not likely to achieve separately or in the current data market practice” (AMdEX-1) and “Both, software code as by design compliant with generic data exchange agreements, and archived individual agreements will secure audit practices for performance and disputes” (AMdEX-3). Such sentences are interpreted as if the information constitutes present tense.

AMdEX’s does not aim at specific target segments (e.g. branche, sector, industry) for data contributors and data users. Data contributors can be individuals, organisations and communities. Data contributors are offered products and services to connect to a generic infrastructure, where they can shape the rules for data sharing themselves: “AMdEX supports the development of a Data Exchange Infrastructure that enables the deployment of secure, reliable and controlled digital data marketplaces for the exchange and linking of data that preserve the sovereignty of data asset owners” (AMdEX-4). During the workshop, it was added that this configuration of a decentralized infrastructure thus also offers independence for the data contributor, because it does not have to rely on a centralized cloud solution. Through these products and services, AMdEX aims to create value for data contributors in the form of security, control and trust in data transactions between data contributor and data user: “Similar to the Internet exchange mechanisms or inter-bank cash transfer services, the Data Exchange mechanisms provide neutral (infrastructure and rules) facilities, operating in the background, allowing for controlled, trusted and secure data transactions” (AMdEX-3). In the workshop it came forth that the AMdEX also creates value in the form of legal compliance and achievement of collaboration possibilities that one party would not be able to achieve alone. Lastly, efficiency from savings through shared costs and matching services have been described: “Daarnaast wordt toegevoegde waarde van AMDEX wordt ook bepaald door: – Matchmaking: reduceren van zoek en vindkosten. – Marketing: creëren van toegang tot marktactoren. – Besparing door gedeelde kosten” (AMdEX-5). Data users can be either providers of data products or consumers of data products. Data providers will offer data products based on the data made available in the Open Data Market, therefore they are classified as data user. Similar to data contributors, data users are offered neutral infrastructure and market rules that creates value in the form of new opportunities for cooperation: a data marketplace “...serves a common benefit no single organization can achieve on its own, offers data asset owners, data consumers new

opportunities for trade activities, service development and cooperation" (AMdEX-4).

AMdEX core resource relates to the digital infrastructure over which data transactions take place. The work group infrastructure proposed to create a generic infrastructure which data contributors can connect to: "Data Exchange Infrastructure will consist of generic facilities that bring data from sovereign data owners to consumers of data. Several different archetypes of Data Exchange Infrastructures are possible, as well as the exchange of different types of data (e.g. static versus dynamic data)" (AMdEX-7). The generic infrastructure would comprise (among others) identity management solutions and authorization solutions, as well it should support auditing, data analytics, data storage with persistent identifiers and metadata services, and support contracts and provision of cloud access. Thus, it should overcome interoperability issues and therefore would use standardized protocols: "Use of open standards for optimal interoperability, and for easy integration of new technologies and services (e.g. interoperable APIs to connect to services)" (AMdEX-4). The current idea is that data contributors can connect to the generic infrastructure by incorporating custom trust-by-design software: "Members can create agreements with algorithm developers to share data" (AMdEX-4) and "sovereign data contributors contact algorithm developers who based on a contract will create a path from the data owner to the generic AMdEX infrastructure" (AMdEX-7).

In the workshop, it became clear that the network of AMdEX is still being formed. Overall, it became clear that partners will be involved during the development process of the AMdEX and for the execution of processes. This is demonstrated by "Infrastructure development could be done with a consortium of partners, here it is important to demonstrate the benefits of implementing the infrastructure as something no partner could achieve alone" (AMdEX-7) and "SURFSara is exploring open source software for the AMdEX infrastructure" (AMdEX-7). Further, AMdEX will likely outsource the execution of operational processes, for instance by issuing tenders for the supply of infrastructure, connectivity, auditing and other services: "AMdEX taken kunnen (deels) uitbesteed worden op basis van bijvoorbeeld 5 jaarlijkse tenders. Voor leveren van infrastructuur, connectiviteit, auditing, en andere diensten" (AMdEX-5). The work group for regulation and compliance added that once the AMdEX is operational, an independent third party, judge or arbitrator should oversee its governance and regulation: "Van grootste belang bij de handhaving is dat dit geschied door een onafhankelijke partij. Dit kan een overheidsorgaan zijn, een onafhankelijke derde of een onafhankelijke partij vanuit de exchange" (AMdEX-6). Governance of the markets can be separated in (the regulation of) generic market rules and of (bilateral) rules and agreements specific to data transactions. Generic market rules will concern rules for both entering and exiting the market, transparent regulation processes, as well as terms of use or codes of conduct: "Er wordt door de deelnemers aangekaart dat er regels moeten worden opgesteld over zowel de toegang tot de exchange en marktplaats als de exit daarvan. Daartoe behoren ook regels i) in geval een lid plots verdwijnt van de exchange dan wel marktplaats na het verkrijgen van data en ii) gedragsregels van de leden: hoe behoort een goed lid zich te gedragen, verbod op monopoliseren, etc" (AMdEX-6). Besides generic market rules, a data contributor will be able to make bilateral agreements for data use: "Verder willen deelnemers graag zelf in controle zijn en dus zelf regels (lees: afspraken) kunnen maken van degene(n) met wie zijn data willen delen" (AMdEX-6). These bilateral agreements would be software-enforced: "Key is that ultimately an agreement is able to automatically execute a data science workflow as a transaction in compliance with conditions that are subsequently monitored and enforced" (AMdEX-4). However, they will also be enforced by the independent third party, judge or arbitrator described above. In addition, the AMdEX will likely allow data contributors to enforce their data specific rules and agreements: "Deelnemers willen daarnaast ook zelf acties kunnen ondernemen indien regels niet worden nageleefd. Hierbij gaat het hen dan om partijafspraken, niet om de algemene regels van de exchange" (AMdEX-6).

Regarding the financial arrangements, the workshop focused on the future AMdEX cost-revenue model. Firstly, it was mentioned that the AMdEX is eligible for funding by the Dutch government which will cover 50 percent of development costs: “Up to 50% of the AMdEX development costs will be covered by the Dutch government” (AMdEX-7). To cover the remainder of costs, workshop attendees were asked to think about potential income models. From the discussion that followed, attendees propose two alternative income models. First, partners would invest in the development of the AMdEX (either monetary or by becoming involved in the development process) in exchange for the use of the resulting AMdEX products and services. Second, AMdEX could partner with scientific or educational organisations to make the development of the AMdEX part of their projects: “The AMdEX could receive finances and be further developed as part of larger research projects, for instance in the form of a Lab or testing ground (”proeftuin”).” (AMdEX-7). In the operational phase, income would come from data contributors that pay for the products and services offered. Price models possibly include additional fees for services like secure payment features for data transactions: “Deelnemers zouden bereid zijn daar meer voor te betalen” (AMdEX-6). Regarding the profit motive of the AMdEX, workshop participants emphasized that the party managing the AMdEX should not be commercial or have a profit-motive: “Daarbij wordt nog genoemd dat er (met dit doel) geen commerciële partij achter mag zitten” (AMdEX-6). Instead, the AMdEX should take the form of an association or foundation.

5.5. City Innovation Platform

The business model of *City Innovation Platform* (CIP) was created based on information from online documentation, and an interview with a case representative (see Table 4.4). The information obtained from the interview (CIP-9) corroborates findings from online documentation, while adding in-depth information to it. For instance, the online documentation indicated that data contributors and data users can be anyone, e.g. “Met het ‘City Innovation Platform’ bieden we nu een oplossing voor alle soorten data, ongeacht de aanbieder of vrager” (CIP-2). In the interview it became clear that while the previous finding is correct, current customers come from specific segments: “Op dit moment zijn de vragers naar data vooral journalisten, burgers, consultancy-bedrijven en aanbieders zelf” and “Voor het grootste gedeelte bestaan aanbieders van data uit gemeenten. Daarnaast maken provincies en commerciele partijen (bijvoorbeeld energie leveranciers) gebruik van het City Innovation Platform” (CIP-9). As such, data contributors are mainly governmental institutions like municipalities or provinces, and commercial data suppliers (e.g. energy corporations). Data users currently concern journalists, civilians, consultancy firms and data contributors themselves. Data users are offered products to access and find high quality data. These products concern data indexing, search engine, APIs, data visualisation and apps. For instance, when a data user types a keyword in the search engine, they get suggestions for which data sets are available, because the data has been indexed. Further, some data sets are provided with visualisations. For example, municipal data is visualised to to ease data reuse by civilians. For the data user who is a developer, access to open data can be gained through standardized API calls. Also, an app was created to enable civilians to report issues to their municipality: “Ook maken we zelf apps, bijvoorbeeld de ”Slim Melden” app, waarin burgers zelf eenvouding een melding kunnen maken bij de gemeente”. Currently, developers can gain access to data through API calls. However, in the interview it was mentioned that developers will be offered a self-service developersportal in the future, to ease data reuse by developers: “Voor ontwikkelaars willen we toewerken naar een self-service developersportal, waar ze zelf makkelijk data kunnen hergebruiken” (CIP-9). The value for data users lies in eased data reuse. Further, for developers of data products based on open data published on the City Innovation Platform, showcasing their products will be offered: “Voor ontwikkelaars die data

hergebruiken in hun toepassingen willen we het mogelijk gaan maken om dit werk te etaleren op het platform, zodat anderen zien wat de mogelijkheden zijn met de data" (CIP-9). For data contributors, several modules of services are offered. These modules comprise of products and services that offer data organizing, storage/hosting, indexing, standardization, publication and security. The interview corroborated these modules as offerings to data contributors: "- Data ordenen en intern inzichtelijk en toegankelijk maken met de Datacatalogus. - Webinars, community building - Customer support - Standaardiseren, combineren en het opslaan van data - Data beveiligen en privacy waarborgen - Data inzichtelijk maken met dashboards en visualisaties - Open data indexeren en publiceren" (CIP-9). Further, in the interview services such as customer support, webinars and community building activities came forth. The online documentation discusses value for data contributors in the form of enabling creation of innovative solutions and developing new insights for improved data-driven decision-making: "Met ons City Innovation Platform kunnen we open data, sensorgegevens en externe databronnen combineren tot waardevolle nieuwe inzichten voor betere sturingsmogelijkheden" (CIP-2). However, in the interview the case representative emphasized that the intended value of City Innovation Platform is to enable data contributors to get a grip on sharing data ("Grip krijgen op het delen van data" (CIP-9)). Although these data points conflict, they are both interpreted as kinds of value created for data contributors. Because, new insights for improved data-driven decision-making can be the result of eased data reuse. From the information above we learnt that data contributors can simultaneously be data users, and that value for data users lies in eased reuse of data. Ergo, for the data contributor who reuses it's own data, value in the form of new insights for improved data-driven decision-making as well as value from getting grip on sharing data can be created.

The core resource of CIP is the infrastructure and technical solutions that enables data logistics between data contributors and data users. The interview added that open data is stored in a cloud. This data is accessible through a download tool and through API's. Online documentation adds that these APIs are standardized and open-source: "Het City Innovation Platform omarmt deze standaarden door o.a. gebruik te maken van de FIWARE NGSI API, de OGC Sensorthings API, FIWARE data modellen en de TM Forum API-standaarden" (CIP-1). Besides the infrastructure, core resources concern data storage systems specified to the type of data being stored and staff whose activities concern data science, customer support and creating service level agreements.

For the development and provision of products and services, CIP partners with other organisations. An example provided in the interview was that a sensor supplier delivers and installs sensors in a city, after which CIP standardizes, processes and publishes this data: "Bijvoorbeeld, een leveranciers voorziet een gemeente van sensoren om luchtkwaliteit te meten, waarna het City Innovation Platform deze data standaardiseert en als open data publiceert" (CIP-9). CIP also enlists partners to mediate the consulting and advice sides of projects, as media partner, as well for the development of visualisations, dashboards and web applications and scalable and interoperable technology. Regarding governance, little information was offered by online documentation, mentioning licenses, access rights and price models: "Een Data Markt, om alle soorten data beschikbaar te stellen, conform licentie- en prijsmodellen en toegangsrechten die de aanbieder van de data bepaalt" (CIP-1). In the interview, it became clear that the these licenses, access rights and price models are recorded in service level agreements. These are contracts between parties, i.e. between CIP and a data contributor or between a data contributor and data user. The service level agreements also may include procedures and accountability for potential conflicts, as well as price models.

Online documentation did not specify the financial arrangements utilised by the City Innovation Platform. Contrary, the interview revealed that Civity, the organisation that manages the City Innova-

tion Platform, is a business with a for-profit motive. In addition, it was explained that data contributors pay for the products and services offered through a monthly subscription. The amount charged in the subscription depends on they type of data and the (combination of) products and services. The composition of the price is recorded in the service level agreement between CIP and the data contributor. For data users, CIP does not impose prices for products and services concerning open data. For instance, there is no charge for making an API call or downloading data: "Op dit moment is het maken van een API call gratis, net zoals het downloaden van open data" (CIP-9). For non-open data, the data contributors determines the price (model). CIP may receive a fee or commission of these transactions, but this would be specified in the corresponding service level agreement.

5.6. Dataverse

For the *Dataverse* case, information was obtained from online documentation (see Table 4.4). While getting familiar with the online documentation, it became apparent that the case is referred to in several manners. This is demonstrated by the phrase "The Institute for Quantitative Social Science (IQSS) collaborates with the Harvard University Library and Harvard University Information Technology organization to make the installation of the Harvard Dataverse openly available to researchers and data collectors worldwide from all disciplines, to deposit data" (DV-1), which pertains to the Harvard Dataverse. This phrase originates from the Dataverse Project homepage and hyperlinks to the Harvard Dataverse page. As such, the software utilized by an entity is called a Dataverse, e.g. the Harvard Dataverse. The entity that enables this software is referred to as the Dataverse Project. Both references explain the case (i.e. how the Dataverse works). Therefore, phrases that pertain to the Dataverse of a specific entity (e.g. Harvard Dataverse and DataverseNL) were treated as the case subject as well.

From the content analysis it was inferred that Dataverse's target segments consist of scholars and researchers from "all disciplines" (DV-3). Data contributors comprise of individual scholars such as researchers, teachers and data authors, but also research projects and institutions. For instance, journals, publishers, data distributors, scholarly organizations, granting agencies, research centers, and archivists. Data contributors are offered a software product (the Dataverse repository) to post and disclose data sets along with a persistent identifier (e.g. DOI or Handle) on their own website. Further, the Dataverse allows the data contributor to manage the data, as well as control how data is accessed and by whom. Last, support services are offered: "In addition to search capabilities, each dataverse also offers detailed customizable choices to dataverse owners for providing a hierarchical organization of available information" (DV-2). The intended value of the products and services for data contributors lies in simplicity of data management and data sharing, and an increase in academic credit and web visibility: "This method of delivering services makes available sophisticated technology in a way that is extremely easy to use by a wide range of people and organizations" (DV-2). In addition, the publication of data sets with a persistent identifier enhances data citation thus increasing academic credit: "The benefits are not only simplicity; journals, authors, archives, and data producers will each receive substantially more credit in terms of visibility on the Web and formal scholarly citations than they do now for the data they make available and articles they publish" (DV-2). Similar to data contributors, data user target segments comprise of (future) researchers, journal editors and others. Data users are offered products for searching, finding and downloading data and performing data analysis. Also, the ability to validate the existence of data, verify original data and checking for updates of data: "Journal copy editors and future researchers can validate the existence of a specific data set that cannot be changed, even if they do not have permission to access it, or if they have permission but do not have

the skills to extract and analyze the data" (DV-2). The products and services create value for data users in the form of access to data and analysis of this data: "the benefits can be classified in terms of data access and statistical analysis" (DV-2).

The core resources of Dataverse comprise of web application software with virtual hosting technology: "The technology used by the Dataverse Network project includes Web application software, which is a program that can be used by anyone with access to a Web browser and without any specialized software installed on one's own computer" (DV-1) and "The Dataverse Network project makes extensive use of a more sophisticated version of virtual hosting technology" (DV-2). These technologies ensure that the Dataverse can be used by any data contributor and data user. In addition, the web application software includes search features, statistical software and metadata harvesting facilities necessary for the provision of products and services. Additional resources relate to the Dataverse team, who offer customer support and are responsible for the development, design and technical support under lead of principal investigators.

The Dataverse team is housed at Harvard's Institute for Quantitative Social Science (IQSS). Development of the software is led by IQSS, but is in collaboration with members of the open-source community who are enabled to contribute code through Git(hub): "The Dataverse Project uses git for version control and GitHub for hosting" (DV-3). Further, the IQSS works with the Harvard Open Data Assistance Program to provide user support services. Lastly, partners offer financial support. In online documentation, it became apparent that contents related to governance pertain to specific entities' Dataverses. However, they are broadly similar. To illustrate, the phrase "The following terms and conditions govern all use of the DataverseNL website (https://dataverse.nl/) (the Site) and the services available on or at the Site (taken together, the Service)" (DV-3) originates from the DataverseNL web pages. DataverseNL is the name for a consortium of 11 Dutch universities. Similarly, the phrase "The following terms and conditions govern all use of the Harvard Dataverse application (the Site) and the services available on or at the Site (the Site and the services taken together, the Service) by you and any third parties who use your account" (DV-3) was found on a web page of the Harvard Dataverse. Due to the similarity over the different data contributors, I derive that terms and conditions, "Terms of Use", apply as the main governance mechanisms in Dataverses. The Terms of Use record the rules that apply to use of the products and services, as well as procedures in case of a breach of these rules. This is demonstrated by "Before you contribute Content to the Dataverse, you must ensure that the Content meets our Terms of Use. You will be held legally and financially responsible for damages to any of the participants in DataverseNL if Content you contribute violates these Terms of Use" (DV-3). The Terms of Use also refer to (custom) data usage license agreements: "Data Usage License Agreement– The license agreement between a Depositor and a Downloader governing the limits and restrictions (or lack thereof) of how the downloaded User Submissions can be used" (DV-3). By default, data sets are assigned the Creative Commons Zero ("CC0") Public Domain Dedication Waiver as data usage license agreement. However, the data contributor can choose to customize this agreement: "Data depositors can opt-out of using the CC0 waiver for their datasets, if needed" (DV-3). Lastly, there are Community Norms. Contrary to the Terms of Use and data use agreements, the community norms are not legally binding: "Please be advised that the Community Norms are not a binding contractual agreement, and that downloading datasets from Dataverse does not create a legal obligation to follow these policies" (DV-3). The Community Norms describe the citation requirements that enable value for data contributors in the form of increased academic credit: "Our Community Norms as well as good scientific practices expect that proper credit is given via citation".

Dataverse's income mainly originates from funding parental organisation Harvard, as well as from

foundations. The products and services are free of charge: “Dataverse Network software is open source, free, and available to all” (DV-2) and “We maintain an email based support service that’s free of charge” (DV-3). Although not mentioned in the documentation, costs are likely to include operational costs for development and maintenance of software and costs of staff.

Table 5.1: Case results per business models dimension.

	Value proposition		Value architecture	Value network	Value finance
	Data contributors	**Data users**			
SSO	Target-segment: private companies. Product-service: data storage/hosting and customer support. Intended-value-element: improved decision-making, public value creation	Target-segment: independent researchers. Product-service: customer support, publication services, funding. Intended-value-element: eased data access and publication	Core-resource: dedicated team of specialists and data storage systems	Governance: agreements on rules, review processes and audits. Role: funding and governance partners	Cost-revenue-structure: non-profit, funds supplied and granted by network actors. Pricing-method: free
Vivli	Target-segment: clinical trial researchers. Product-service: data storage/hosting, processing, managing, publishing, access control and customer support. Intended-value-element: increased data discoverability, efficiency and security in sharing, data aggregation and analysis.	Target-segment: research community. Product-service: data search, request, aggregation and analysis tools and customer support. Intended-value-element: increased data findability, simplicity in requesting, accessing and analysing data	Core-resource: digital platform with data storage system, search tools and research environment for data users and staff	Governance: harmonized agreements, request review processes. Role: funding, resource provision, service provision partners	Cost-revenue-structure: non-profit, funded by partners. Pricing-method: product/services charged
HDX	Target-segment: organisations. Product-service: data storage/hosting, processing, analyzing, publishing, access control. Intended-value-element: increased data discoverability	Target-segment: humanitarian-aid workers. Product-service: data search, request, aggregation and analysis tools and customer support. Intended-value-element: increased data findability	Core-resource: digital platform with data hubs based on open-source software, staff	Governance: terms of use, rules, processes for evaluation. Role: service design and testing partners	Cost-revenue-structure: non-profit, income from managing organisation. Pricing-method: free

	Value proposition		Value architecture	Value network	Value finance
	Data contributors	**Data users**			
AMdEX	Target-segment: any. Product-service: infrastructure for data sharing, custom access control. Intended-value-element: sovereignty, compliance, trusted data transactions, new opportunities for cooperation, cost efficiency	Target-segment: consumers or data product providers. Product-service: neutral infrastructure and market rules. Intended-value-element: trusted data transactions, new opportunities for cooperation	Core-resource: digital infrastructure interoperable with custom governance-enforcing algorithms designed by data contributor	Governance: generic market rules and conditions, audit and conflict resolvement processes, bilateral data exchange agreements. Role: partners for development and provision of products/services	Cost-revenue-structure: non-profit, government funding, partner investments in exchange for use of the AMdEX. Pricing-method: products/service related fees for data contributors
CIP	Target-segment: any. Product-service: data storage, standardization, publication, access control, customer support. Intended-value-element: grip on sharing data, data-driven decision making, enabling innovation	Target-segment: any. Product-service: data indexing, search engine, APIs, data visualization and apps. Intended-value-element: eased data reuse	Core-resource: standardized open digital infrastructure, data storage system	Governance: service level agreements. Role: partners for resource and service development and provision	Cost-revenue-structure: for-profit. Pricing-method: monthly subscription for data contributors based on products/service configuration, prices for use of commercial data determined by data contributor
DV	Target-segment: research-affiliated entities. Product-service: data storage/hosting, processing, managing, publishing, access control, customer support. Intended-value-element: increased discoverability, simplified data management/sharing, increased academic credit	Target-segment: (future) researchers. Product-service: data search, validation, authentication, updates, analysis. Intended-value-element: increased data findability, simplicity in requesting, accessing and analysing and citing data	Core-resource: software installation, hosting technology, metadata harvesting and statistical analysis tools	Governance: terms of use, data usage license agreements, requirements, norms and licensing. Role: resource development, service provision partners	Cost-revenue-structure: funded by parent organisation and partners. Pricing-method: free

6

Discussion of results

The research objective of this thesis is to create a better understanding of the Trusted Data Intermediary as a coordination mechanism in data collaboratives, by describing business model archetypes of Trusted Data Intermediaries. In order to achieve this objective, the business model findings presented in the previous chapter need to be grounded in the current literature and synthesized to answer the main research question.

The following sections will address these needs. Section 6.1 to 6.4 will discuss the patterns found for kinds of value, resources, governance mechanisms and financial arrangements and how they relate to the literature on intermediaries. Figure 6.1 at the end of this section summarizes which patterns were identified in which cases. From this discussion, section 6.5 moves on to discuss similarities and differences found between the cases and answer the main research question *"What are business models for value creation by Trusted Data Intermediary for resp. data contributors and data users?"*.

6.1. Kinds of value created by Trusted Data Intermediaries

The first sub research question comprises *"What kinds of value do Trusted Data Intermediaries offer data contributors and data users?"*. For this question, the value proposition dimension was studied. The findings largely correspond to the findings of Yoon et al. (2019) and van Schalkwyk et al. (2016) for data intermediaries in the open data domain (see section 2.2.2). However, instead of merely describing a role or the intended-value-element, this discussion is novel in the sense that it connects the findings with concepts in value creation literature.

6.1.1. Value for data contributors

For data contributors, I distinguish four patterns in the kinds of value created by Trusted Data Intermediaries. To begin, Trusted Data Intermediaries create value as **simplified data management**, by offering solutions for organising and managing data. This increases the efficiency of data management.

This is demonstrated by the cases *Vivli* and *Dataverse*. For *Vivli*, data management created through products for anonymization and mapping of data. *Dataverse* offers data contributors functionalities of organizing their own dataverse. Although not explicated as intended value, the *City Innovation Platform* also offers products for simplifying data management (i.e. Datacatalogus). Similarly, *Humanitarian Data Exchange* offers data contributors checks for eliminating errors in data. Such platform features are similar to those offered by *Vivli*, so I argue that *Humanitarian Data Exchange* also creates value in simplified data management. The second pattern concerns **control** or sovereignty in sharing data. This value is created by access control products. While, the cases *City Innovation Platform*, *Amsterdam Data Exchange* and *Dataverse* make the kind of value explicit, others also explicitly offer access control products (e.g. *Vivli* and *Humanitarian Data Exchange*). Third, value in the form of **increased discoverability** is created. This pattern comes forth in the cases *Humanitarian Data Exchange* and *Vivli*. Last, two cases demonstrate value of the kind **improved decision-making**. Namely, *Social Science One* and *City Innovation Platform*.

From section 1.1 we should recall that data collaboratives are challenged by lacking discoverability and privacy issues. In section 2.2.2, no concept was found that describes abatement of barriers or the creation of collaboration possibilities. However, the concept of appropriated value may be interpreted such that abatement of barriers falls within. On the other hand, with a broad interpretation the concepts of interaction value and synergistic value can apply as well. For instance, one could argue that without the collaboration with the Trusted Data Intermediary, a data contributor would not be able to achieve the same level of exposure to data users (synergistic value). Otherwise, in the process of collaboration with the Trusted Data Intermediary, value is created from data being discoverable (interaction value). One could even argue that transferred resource value applies in the cases where technical resources like data storage systems are deployed as a service for data storage. Therefore, classifying the kinds of value from increased discoverability of data and enhanced efficiency and security in data sharing is difficult. Nonetheless, I choose to apply the concept of appropriated value, to explicate the fact that data contributors capture value from the activities in the collaboration. For data contributors in Trusted Data Intermediaries, appropriated value thus refers to increased discoverability of data, enhanced efficiency and security in data sharing, and eased data processing and management. Next, "improved decision-making" can only occur when the data contributor has learnt from contributing the data. This can be perceived as knowledge adoption by the data contributor, from knowledge created in the process of interacting with others in the value network. Thus, I argue that improved decision-making is value of the kinds interaction value as described by Austin and Seitanidi (2012). In addition, *Social Science One* proposed value for data contributors in the form of enabling public value creation by private companies. However, combined with the aim of creating value by improving decision-making, I would argue that the Trusted Data Intermediary in this case thereby also offers the enhancement of competitiveness. Thus, the Trusted Data Intermediary does not enable public value creation, rather the kind of value offered here is shared value.

6.1.2. Value for data users

For data users, three patterns are distinguishable. To begin, value from **eased reuse of data** is created. *City Innovation Platform* is the only case that explicitly addresses this as intended value. However, other cases enable this value creation as well. The cases *Vivli*, *Dataverse* and *Humanitarian Data Exchange* offer (network partner's) data analysis tools, statistical software or visualisations that enable data reuse. On a higher level of abstraction, *Amsterdam Data Exchange* aims to create value through new opportunities for cooperation and increased trust in data transactions. I would argue that these are

enablers for eased data reuse by collaborations. Although not explicated, *Social Science One* also eases reuse of data. This follows from the value they create for data contributors: "and deployment of the scientific community to help advance societal good with previously inaccessible information" (SSO-2). This sentence implies private company data reuse by the scientific community. As such, all six cases create value in eased reuse of data. Secondly, Trusted Data Intermediaries create value in the form of **increased findability** of data. The intention of creating this value is emphasized by *Humanitarian Data Exchange* and *Vivli*. The underlying products concern search features. As such, *City Innovation Platform* and *Dataverse* create value in findability of data as well. Thirdly, value in the form of **eased data access (requesting)** is created. This is demonstrated by *Social Science One*, *Vivli* and *Dataverse*. Also, *Humanitarian Data Exchange* eases the requesting of data. Although this is not explicitly stated as an intended value, this shows from the access request products for data users.

Data reuse implicates that resources (i.e. data) are being transferred between entities. As such, I would argue that this kind of value can be stipulated as transferred resource value. However, other concepts can also apply, depending on the driving elements. To illustrate, eased data reuse from new opportunities for cooperation can be stipulated as synergistic value. Because, such opportunities enable the achievement of outcomes one would not be able to achieve alone. For the latter two kinds of value (increased findability and data accessibility), academic concepts lack. Both can be interpreted as abatement of barriers. The lack of (low-cost) accessibility of data has been described as a barrier in section 1.1. Findability is not explicitly mentioned as a barrier. However, findability and discoverability are terms that can be interchanged, and lacking discoverability has been mentioned as a barrier. As such, the same problem arises as for data contributors: there is no concept that adequately describes the abatement of barriers for data users. For consistency, I choose to apply the concept appropriated value to capture value from data findability and accessibility.

6.2. Key resources for value creation by Trusted Data Intermediaries

The second sub research question is *"What are key resources for value creation by Trusted Data Intermediaries?"*. For this question, the value architecture dimension is studied.

From the comparison of the cases' business models, I find the following patterns. Firstly, five of the cases operate on some sort of digital infrastructure. *Amsterdam Data Exchange* and *City Innovation Platform* emphasize that products and services are offered over an underlying **standardized digital infrastructure**. The same applies to the cases *Dataverse*, *Humanitarian Data Exchange* and *Vivli*, although differently implemented. *Dataverse* uses a digital infrastructure in the form of **installable software** that can be downloaded by customers. While, *Humanitarian Data Exchange* and *Vivli* implement the digital infrastructure as a **platform**. Secondly, **data storage systems** are demonstrated by the cases *Social Science One*, *City Innovation Platform*, *Vivli*. Further, *Dataverse* explicates data storage as virtual hosting technology. This technology hosts data archives stipulated as Dataverses: "A Dataverse repository is the software installation, which then hosts multiple virtual archives called Dataverses. Each dataverse contains datasets, and each dataset contains descriptive metadata and data files (including documentation and code that accompany the data)" (DV-1). Similar, *Humanitarian Data Exchange* offers data storage through its platform: "HDX is based on CKAN, an open-source data management system for powering data hubs and data portals" (HDX-3). Further, two cases (*Dataverse* and *Humanitarian Data Exchange*) demonstrate a core-resource based on open-source software. This

entails that any entity may contribute to the development of the software. For these cases, this has implications for the value network. Contrary to the other cases, *Amsterdam Data Exchange* does not centralize data storage. Rather, data is stored by the data contributors themselves and linked to the generic infrastructure through custom made algorithms. Thirdly, *Social Science One*, *Humanitarian Data Exchange*, *Vivli* and *City Innovation Platform* explicate the core-resource in the form of **staff** that enable the provision of products and services. In particular, *Social Science One* emphasizes the resource of staff as a core-resource, because of their prominent function as a trusted third party between data contributor and data user. All four cases deploy staff for the provision of services and execution of governance. Although not explicated, the email-based customer support service offered by *Dataverse* implies staff for the provision for services here as well. Fourth, *Humanitarian Data Exchange* and *City Innovation Platform* also deploy staff for the development of products and services. For example, for the creation of visualizations and performing data science activities. Last, some cases supplement the digital infrastructure or storage systems with **data analytics features**. For example, the *Vivli* platform offers a research environment with statistical tools. Similar, *Dataverse* offers tools for metadata harvesting and statistical analysis. Also, *Humanitarian Data Exchange* offers tools for analyzing data for errors.

The findings from the business models largely contradict the literature. Literature on data intermediaries identifies resources to be primarily of financial and human nature (see section 2.2.3). Contrary to this literature, technical resources are found to be most prominent. With regards to offering the products that enable creating value in the form of eased, efficient and secure sharing and exchange of data, it is clear that all Trusted Data Intermediaries are dependent to IT systems. Foremost, digital infrastructures, platform, data storage systems, software, and tools are deployed by the Trusted Data Intermediaries for the offering of products and services. It is reasonable to expect this in Trusted Data Intermediary business models, since the main focus lies with the enabling of data transactions and data can only be transferred through IT systems. The results of this study concur with Amit and Zott (2001)'s argument that in electronic markets information-based resources are more important to value creation than other types of resources and capabilities. However, they contradict the literature on data intermediaries. An explanation for this may be that resources were not a focus point in the literature. For instance, van Schalkwyk et al. (2016) and Yoon et al. (2019) studied value creation by data intermediaries for communities, emphasizing their roles rather than operational mechanisms. Likewise, literature identifies financial resources, yet this is not corroborated by the results. There are two possible explanations for this. On the one hand, it is possible that the financial resources are not present, because the Trusted Data Intermediaries simply do not have or use own financial resources. On the other hand, the lack of financial resources identified in this dimension may be a consequence of the chosen theoretical framework. The theoretical framework explicates finances are a separate dimension, making it an umbrella concept for the elements of *pricing-method* and *cost-revenue-structure*. The results of the multiple-case studies showed that the cost-revenue-structure is primarily based on funds from partners. As such, one could argue that these funds constitute the financial resources of the Trusted Data Intermediaries, thus that the funds are substitute resources. Due to the chosen theoretical framework, these resources do not come up in the value architecture dimension, rather in the value network and value finance dimensions. Another type of resource identified for Trusted Data Intermediaries does concur with the literature. This type of resource refers to the staff that develop or provide products and services, and execute governance. Especially in the case of *Social Science One*, staff deployed as resources for providing services and overseeing governance are the core resource. In this case, the staff are specialized for the social science domain. Thus, human resources are identified in the results.

6.3. Key governance mechanisms for value creation by Trusted Data Intermediaries

Third, the sub research question *"What are key governance mechanisms for value creation by Trusted Data Intermediaries?"* has been asked. To answer this question, I looked at the value network of Trusted Data Intermediaries.

Regarding governance mechanisms, three patterns are observed from the cases' business models. To begin, **agreements** are found in all cases. Different types of agreements can be identified. *City Innovation Platform* indicated only to use service level agreements specific to a certain data contributor. While, *Amsterdam Data Exchange* enables bilateral agreements between data contributors and data users. *Social Science One* makes separate agreements with data contributors on the rules for the staff, e.g. regarding the scope of the project or which data may (not) be accessed. Also, data users agree to rules, standards and processes. Similar, *Dataverse* and *Vivli* leverages agreements that pertain to data contributors or data users. For *Vivli* these agreements are harmonized so that additional agreements between the parties are unnecessary. For *Dataverse* these are data usage license agreements. These agreements eliminate the need for bilateral agreements between the parties. Besides bilateral agreements, *Dataverse* makes all customers agree to Terms of Use. This is also the case for *Humanitarian Data Exchange*. A second governance mechanism pertains to rules imposed on customers and network partners. Cases can implement **rules in agreements**, e.g. *Social Science One*, *Vivli* and *Amsterdam Data Exchange*. For *Humanitarian Data Exchange* and *Dataverse*, rules are implemented in the Terms of Use. *Dataverse* also implements rules for licensing in data usage license agreements. In addition, *Amsterdam Data Exchange* implements **software-enforced rules**. Thirdly, review processes are recurrent in the business models. These can pertain to **data request review processes**, or **reviews of contributed data**. For instance, *Vivli* has processes in place for the reviews of data requests. *Humanitarian Data Exchange* reviews contributed data. Alternatively, cases make use of **audits by independent partners**. For example, *Social Science One*'s processes involve reviews of the data itself, the review of data users, as well as the prospected data use and the performance of human resources. For the latter process, an independent partner is responsible. Similar, in the *Amsterdam Data Exchange*, review processes take the form of audits and reviews by an independent arbitrator.

The identified governance mechanisms corroborate literature, and even offer more insights into the exact manners of implementation. For instance, the cases demonstrate that agreements may vary from standardized to custom. In addition, rules are enforced in different ways, e.g. copyright laws, self-determined, software-enforced or enforced by regulation processes. What stands out, is that *Social Science One* uses regulation processes executed by partners to provide self-feedback. This is done as a way to increase the trust of network partners in the Trusted Data Intermediary. Similarly, *Amsterdam Data Exchange* emphasized the importance of rules and an independent arbitrator to ensure trust in data transactions. These mechanisms is what Stadtler and Probst (2012) describe as bonding social capital: norms that create trust, communication and information sharing.

6.4. Key financial arrangements for value creation by Trusted Data Intermediaries

The last sub research question is formulated as *"What are financial arrangements of Trusted Data Intermediaries?"*. For this question, the value finance dimensions of business models were assessed.

Several patterns are observed over the different elements in the finance dimension of the cases' business models. To begin, the profit motive of most cases is **non-profit** oriented. The exception is *City Innovation Platform*, which explicated a for-profit motive. *Dataverse* did not explicate a profit motive, but given that its managing organisation is an educational institution, I assume the profit motive is less commercial than is the case for *City Innovation Platform*. A second pattern is that some cases **charge customers for products and services**. The cases *Vivli*, *City Innovation Platform* and *Amsterdam Data Exchange* incur a pricing method, although they implement them in different manners. For instance, *Vivli* charges both data users and data contributors, while *City Innovation Platform* charges only data contributors with a monthly subscription tailored to the chosen (combination of) products and services. While, *Amsterdam Data Exchange* demands investments from network partners, prior to being added to the value network. The cases *Humanitarian Data Exchange*, *Dataverse* and *Social Science One* offer products and services free of charge. However, *Social Science One* explicated that they may consider a pricing method for data contributors in future projects. Similar, *City Innovation Platform* may in the future develop a price model the use of the API products by data users. Thirdly, the cases without pricing methods (e.g. *Social Science One*, *Humanitarian Data Exchange* and *Dataverse*) fully rely on funding from other entities. Specifically, the latter two receive this **funding through their managing (parent) organisation**. Fourth, cases that have income from pricing models can also receive funding. This is demonstrated by *Amsterdam Data Exchange* and *Vivli* who receive **funds directly from network partners**.

Similar to the literature described in section 2.2.5, the studied business models indicate that a non-profit motive and income from customer charges are not mutually exclusive. Furthermore, the income can simultaneously come from customer charges as well as funds from network partners. This is demonstrated by *Vivli*, which is a non-profit organisation with both a pricing model and income from funding. However, most cases receive funding from other entities, e.g. network partners or parent organisations. Because these financial arrangements relate to partner resources, I describe these financial arrangements as substitute financial resources. This term accommodates financial resources that were identified in the literature but not in the value architecture dimensions. Moreover, the term indicates the dependency to supporting partners when it comes to the financing of Trusted Data Intermediaries' business models.

6.5. Trusted Data Intermediary business models

The main research question of this thesis is "What are business models for value creation by Trusted Data Intermediaries for resp. data contributors and data users?" An answer to this question should create a better understanding of how Trusted Data Intermediaries operate. In order to answer the research question and reach my objective, I will explain the similarities and differences for each pattern identified in the previous section by the means of case characteristics.

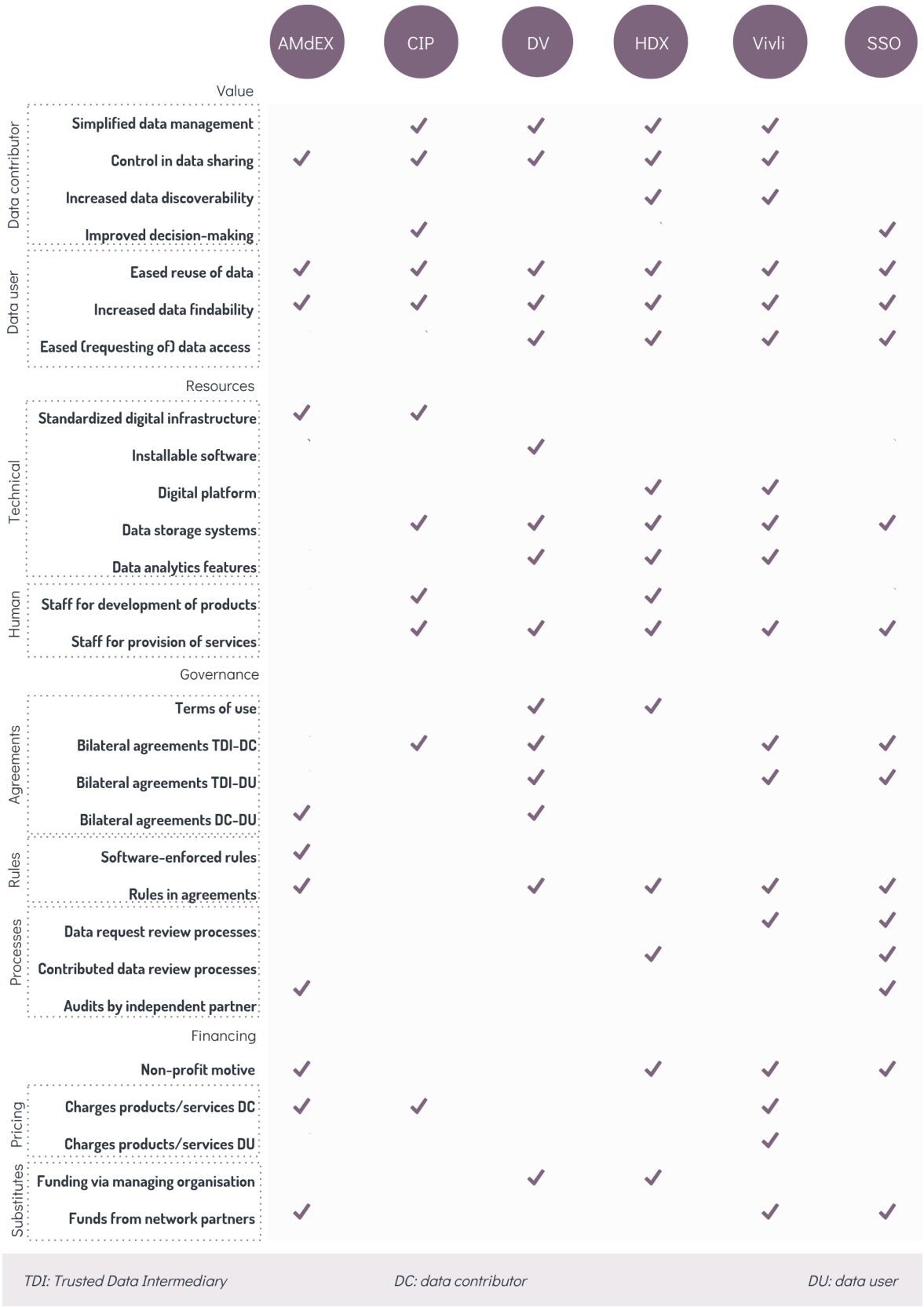


Figure 6.1: Visual representation of findings per case.

6.5.1. Characterization of cases

In section 4.4.2, tentative characteristics such as non-profit motive and a focus on specific data domains were suggested. Furthermore, in section 2.1.1 Janssen and Zuiderwijk (2014) suggested that intermediaries can be characterized by specialization. These characteristics are corroborated and supplemented by the findings described earlier. First of all, the cases *Amsterdam Data Exchange*, *Humanitarian Data Exchange*, *Vivli* and *Social Science One* explicated their non-profit motive. Further, *City Innovation Platform* explicated a for-profit motive. *Dataverse* is the only case that did not explicate a profit motive. As described earlier (section 6.4), I assume that *Dataverse*'s profit motive is less commercially oriented than that of *City Innovation Platform*, given that it's parenting organisation is an educational institution. Secondly, the cases differ in specificity of target segments. For instance, *Amsterdam Data Exchange* and *City Innovation Platform* indicated not to focus on specific target groups. *Dataverse* is focused on scholarly data, but does not focus on specific research areas. Similar, *Humanitarian Data Exchange* does not focus on data from specific sources, but requires data to be relevant for humanitarian crises. Contrary to *Dataverse*, *Social Science One* and *Vivli* focus on specific research areas. *Social Science One* focuses on private sector data that is relevant for a specific research area (here: social sciences). *Vivli* is most specific, since it focuses solely on data from clinical trial researches. Janssen and Zuiderwijk (2014) referred to this as the characteristic of specialization. Thirdly, an emergent characteristic from the analysis is the (de)centralized data storage. *Amsterdam Data Exchange* has emphasized that its infrastructure does not centrally stores data in a cloud: "Independence: decentralized market, so there is no centralized cloud solution" (AMdEX-7). Contrary, *City Innovation Platform* places emphasis on standardized and centralized data storage in a cloud: "Open data cloud, toegankelijk via download tools en standardized APIs" (CIP-9). *Social Science One* develops data storage systems for projects, so it decentralizes the company data that has been made accessible to data users: "This has involved an enormous amount of effort: creating entirely new systems for securely storing the data, incorporating cutting-edge differential privacy and other techniques, and building a dedicated team of specialists" (SSO-3). *Dataverse* works with virtual hosting technology, which implies centralized data storage: "A Dataverse repository is the software installation, which then hosts multiple virtual archives called Dataverses" (DV-1). In a similar fashion, *Vivli* and *Humanitarian Data Exchange* offer hosting services through their digital platform in the form of data hubs or repositories: "Humanitarian Data Exchange is based on CKAN, an open-source data management system for powering data hubs and data portals" (HDX-2) and "The Vivli platform includes an independent data repository, in-depth search engine and a secure research environment" (Vivli-3). In Figure 6.2 I have placed the cases on the spectra of the three identified characteristics.

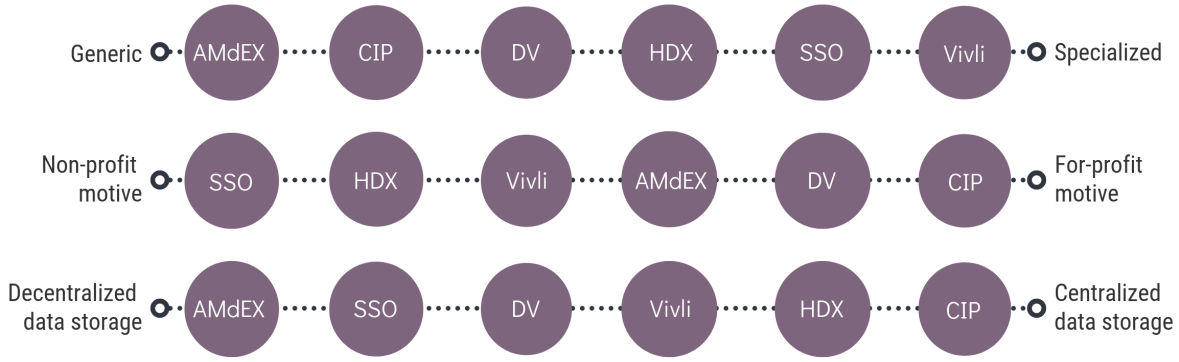


Figure 6.2: Case characteristics.

6.5.2. Differences and similarities explained

Value

Above all, it can be observed that all cases create value for data users in the form of eased reuse of data and increased data findability. Secondly, value for data contributors from control in data sharing is found for all cases but *Social Science One*. From section 6.2 we know that *Social Science One* emphasizes human resources over technical resources. In other cases, data access control is mostly offered through products provided by technical resources such as a digital infrastructure. Further, from the within-case analysis in section 5.1, we know that the staff has the decision-making power on which research proposals are awarded access to the data. This contradicts that control in data sharing lies with the data contributor. As such, the difference between *Social Science One* and other cases on this matter can be explained by the governance structure and value architecture of *Social Science One*. Thirdly, it is observed that value from improved decision-making is intended by the cases *Social Science One* and *City Innovation Platform*. This is interesting, because they are opposites of each other regarding profit motive. *City Innovation Platform* is the only case that explicated a commercial intent with the business model. While, *Social Science One* emphasizes its non-profit motive. This is especially important for *Social Science One*, since it needs to show private sector data contributors that collaboration with *Social Science One* would not create a competitive disadvantage. Fourth, four cases (*Vivli*, *Humanitarian Data Exchange*, *Dataverse* and *Social Science One*) additionally create value for data users in the form of eased (requesting of) data access. The cases that do not create this value (*Amsterdam Data Exchange* and *City Innovation Platform*) both do not have specific target segments. As such, a possible explanation for the difference is that specific data user segments may encounter barriers in accessing sector-relevant data. From section 1.1, we know that data accessibility is a common barrier. It is possible that the Trusted Data Intermediaries that specialize for specific target segments of data users, address these needs for sector-relevant data access. Fifth, it stands out that the cases that create value in the form of increased data discoverability (i.e. *Humanitarian Data Exchange* and *Vivli*) both operate on a digital platform. The platforms act as an interface for the data hosted by the Trusted Data Intermediaries. Further, both cases utilize search features and data analytics features. Such features are also offered by the cases *City Innovation Platform* and *Dataverse*, but they implement different types of digital infrastructures. As such, a possible explanation may be that the central storage of data, accessible through a digital platform is the discriminator for this form of value for data contributors. Last, intended value for data contributors in the form of simplified data management is created by *City Innovation Platform*, *Dataverse*, *Humanitarian Data Exchange* and *Vivli*. From the previous section we know that these cases all make use of centralized data storage solutions, opposed to the others. As such, the centrality of data storage may explain why this kind of value is created by these cases only.

Resources

Regarding the key resources, most noteworthy is that all cases utilize technical resources. Secondly, it should be observed that *Amsterdam Data Exchange* is the only case that does not utilize data storage systems. This can be attributed to their proposition of a decentralized data market. As described in section 5.4, the decentralization of data storage is part of the intended value in the form of control in data sharing. A third observation the use of a digital infrastructure. As indicated before, *Social Science One* is the only case that does not utilize a digital infrastructure (i.e. software, platform or standardized infrastructure) for the provision of products and services. Rather, *Social Science One* relies on staff and decentralized data storage systems for this purpose. This can be explained by the fact that *Social*

Science One hand-picks the data users that will obtain access to the contributed data. Therefore, *Social Science One* lacks the need for a digital infrastructure accessible by a large population of data users. Fourth, *Amsterdam Data Exchange* is the only case that does not utilize human resources. It relies solely on the utilization of a standardized digital infrastructure. This difference can be explained by the network dimension of the *Amsterdam Data Exchange*. Since audits by an independent partner is part of the governance model, one could argue that *Amsterdam Data Exchange* utilizes substitute human resources. In a similar fashion, most cases that do not explicate staff for the development of products (i.e. *Amsterdam Data Exchange*, *Dataverse*, *Vivli*) utilize substitute human resources from network partners. Last, it should be noted that resources like data analytics features have shown not to be a prerequisite for the creation of value in the form of eased reuse of data.

Governance

All cases that emphasize value in the form of control in data sharing leverage agreements. Contrary, not all cases that create this kind of value leverage rules. The exception being *City Innovation Platform*. *City Innovation Platform* is the only case with an explicit for-profit motive. This commercial orientation may relate to the organisation being less inclined to impose rules on customers. Similar, *City Innovation Platform* does not have any review processes in place. Neither does *Dataverse*. For *Dataverse* a possible explanation may be limited financial resources, as it does not receive additional funding and it does not impose a pricing model to cover the costs of review processes. Instead, *Dataverse* is the only case that leverages (rules in) agreements with both customer segments, as well as between customers. Second, only *Amsterdam Data Exchange* enforces rules through software. Other cases that utilize a digital infrastructure do not emphasize this form of enforcing rules. This difference can be attributed to the fact that the *Amsterdam Data Exchange* is still being developed. Alternatively, the discriminator may be the decentralized infrastructure. *Amsterdam Data Exchange* will enable data contributors to create custom algorithms for connecting to the generic infrastructure over which data can be exchanged. Such a path is not extant in the other cases with a digital infrastructure, since in those cases data is stored centrally. Third, of the cases that enforce review processes, only *Social Science One* and *Vivli* leverage processes for reviewing data requests. These cases are most specifically oriented to sectors. However, a more logical explanation would be that the discriminator is the type of data handled. We know that *Social Science One* gains access to private sector data that could pose a threat to the company's competitive advantage should data be accessed by unauthorized users. Similar, *Vivli* handles clinical trial data. Thus, review processes for access requests to this data may be a precaution in safeguarding what the data will be used for. Fourth, most cases demonstrate bilateral agreements with the data contributors. The exceptions, *Humanitarian Data Exchange* and *Amsterdam Data Exchange* did not explicate these agreements. However, from their respective within-case analyses in chapter 5, we can derive that such agreements are implemented in another form. For *Humanitarian Data Exchange*, the chosen implementation are the Terms of Use that data contributors agree to. For *Amsterdam Data Exchange*, generic market rules have been emphasized. Although not specified, it is likely that prior to the use products and services, data contributors will have to agree to these rules in some form, for example by agreeing to codes of conduct or Terms of Use. Fifth, the cases that leverage bilateral agreements with data users (i.e. *Vivli*, *Social Science One* and *Dataverse*) have explicated the intended value for data users in the form of eased (requesting of) data access. Therefore, it is possible that the Trusted Data Intermediaries have chosen to implement these agreements as instruments for the creation of this kinds of value. Sixth, the two cases that leverage governance mechanisms in the form of audits by independent partners, *Amsterdam Data Exchange* and *Social Science One*, both decentralize data storage. This is interesting, as these are also the only cases to emphasize trust in

the network in the form of norms that create trust, communication and information sharing. Seventh, only *Humanitarian Data Exchange* and *Social Science One* leverage processes to review contributed data. From these cases we know that part of their value proposition is to increase data access, findability and reuse of sector-relevant data. For *Social Science One* this pertains to the sector of social sciences, while for *Humanitarian Data Exchange* it pertains to data relevant for crisis response. Interestingly, *Vivli* did not explicate review processes of contributed data. While, *Vivli* is also sector-specific. It is possible that *Vivli* does leverage such processes, but that this is a part of the registration process of new data contributors. This process was not explicated by the online documentation. However, given the prices imposed on data contributors, as well as the use of harmonized data agreements between *Vivli* and data contributors, it is likely that such a process is implemented. Eighth, *Amsterdam Data Exchange* and *Dataverse* have explicated bilateral agreements between data contributors and data users. Between their business models, no other commonalities are found besides that they both utilize a digital infrastructure. Possibly, the need for bilateral agreements between customer segments can come from the fact that the data is not specific to sectors or industries. A reason for *City Innovation Platform* not to have these agreements could be that currently most data transactions on the *City Innovation Platform* concern open data. For non-open data, *City Innovation Platform* indicated that data contributors are enabled to impose a pricing model for data users themselves. Such a pricing model implies an agreement between the seller and buyer. So, although not explicated, *City Innovation Platform* may enable an agreement between data contributors and data users.

Financing

All cases but *City Innovation Platform* receive substitute financial resources. This can be attributed to the for-profit motive of the case organisation, as well as the lack of a parent organisation. Secondly, only *Vivli* charges data users. Although *City Innovation Platform* has indicated they may consider a pricing model for their API products, as this is currently not implemented. *Vivli* offers data users an extensive user environment for conducting research. Thus, the income from the pricing models may be needed to cover the developmental and operational costs of the extensive products. However, *Vivli* receives substitute financial resources, so it is not completely reliant on the pricing models. A second explanation can relate to imposing a threshold for data users. *Vivli* enables transactions of data from clinical trials. Thus, the pricing model for data users can be instrumental in protecting the purpose of the data being used for advancements in medical science. This explanation presupposes that governance mechanisms are internalized in the financial arrangements of *Vivli*. Thirdly, the cases that utilize a digital infrastructure and do not receive funds via a parenting organisation (i.e. *Amsterdam Data Exchange*, *City Innovation Platform* and *Vivli*), all charge data contributors for the use of products and services. Interesting to note is that *Social Science One* has indicated they may ask funding by data contributors in the future. Contrary to the other cases, *Social Science One* does not operate on a digital infrastructure. Rather, their core resources concern staff. From the within-case analysis of *Social Science One* (section 5.1), we know that the staff (i.e. the commission) is compensated for their work.

6.5.3. Trusted Data Intermediary business model archetypes

In the previous sections, patterns in Trusted Data Intermediary business models have been identified. Moreover, the presence of these patterns have been explained in the light of characteristics of the studied Trusted Data Intermediaries. From this discussion it became apparent that some patterns

occur regardless of the Trusted Data Intermediary's characteristics. For instance, all studied Trusted Data Intermediaries create value for data users in the form of eased reuse of data. In addition, it became apparent that some business model patterns can be related to case characteristics. To illustrate, cases not specialized to -specific target segments (i.e. *Amsterdam Data Exchange*, *City Innovation Platform* and *Dataverse*) leverage bilateral agreements between customers as governance mechanisms. In Figure 6.3, the proposed relations between business model patterns and characteristics have been visualized.

In the next paragraphs, I propose two Trusted Data Intermediary archetypes based on these relations. First of all, the five patterns identified in all cases constitute both archetypes. Both archetypes utilize technical resources, (substitute) human resources and leverage agreements to create value for data users in the forms of eased reuse of data and increased data findability. Based on the specialization of the Trusted Data Intermediary, the archetypes constitute different elements. First, I discuss the business model archetype for Trusted Data Intermediaries that do not specialize to specific target segment. This archetype will be referred to as the *Generic Trusted Data Intermediary business model archetype*. In addition, I discuss an archetype for Trusted Data Intermediaries specialized to specific target segments, which I refer to as the *Specialized Trusted Data Intermediary business model archetype*. Both archetypes are supplemented with conditional constituents identified in the discussion in section 6.5.2. As well, since the characteristics *data storage centrality*, *profit motive* an *specialization* do not exclude each other, the implications of these additional characteristics for the business model archetypes are discussed. Figure 6.4 offers a visual representation of the business model archetypes and the possible variations identified in this thesis.

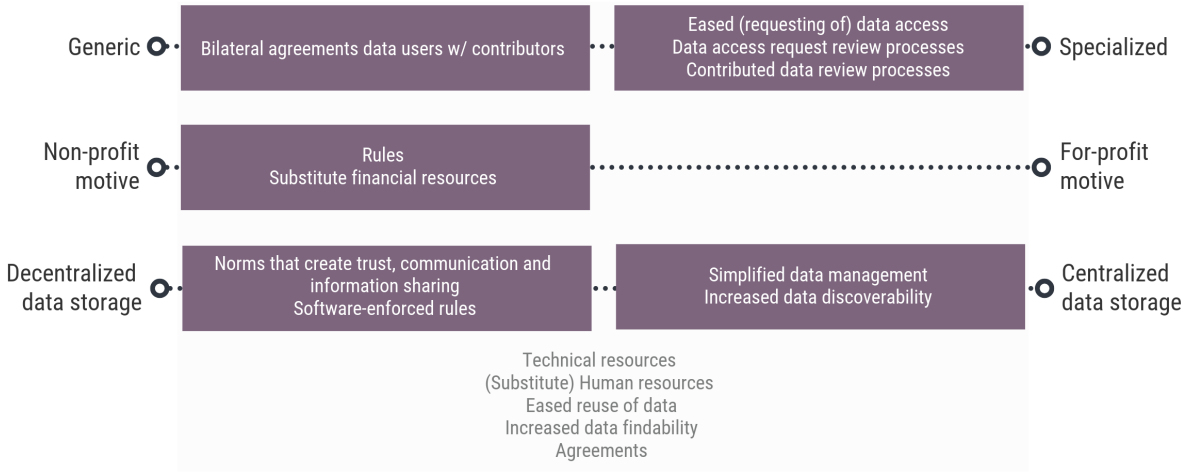


Figure 6.3: Proposed relations between business model patterns and Trusted Data Intermediary characteristics.

Generic Trusted Data Intermediary business model archetype

Trusted Data Intermediaries create value for data users in the forms of eased reuse of data and increased data findability. Should the resources concern centralized data storage, then value for data contributors in the forms of simplified data management and increased data discoverability may be created. To create value, technical resources and (substitute) human resources are utilized. Human resources from network partners can be leveraged (external staff), or internal staff can be deployed for the provision or development of products and services. Technical resources may comprise of a digital infrastructure or data storage system. Trusted Data Intermediaries that utilize a digital infrastructure may create value for data contributors in the form of control in data sharing. This kind of value may be supported

with governance mechanisms in the form of agreements. In particular, bilateral agreements between data users and data contributors may be leveraged. Intermediaries may receive substitute financial resources from either a parent organisation or network partners, or it may charge customers for the use of products and services. Trusted Data Intermediaries that utilize a digital infrastructure and lack substitute financial resources from a parenting organisation may charge data contributors and data users for the use of products and services.

Specialized Trusted Data Intermediary business model archetype

Trusted Data Intermediaries specialized to specific target segments create value for data users in the form of eased (requesting of) data access, eased reuse of data and increased data findability. Specialized Trusted Data Intermediaries that utilize a centralized data storage system can additionally create value for data contributors in the forms of simplified data management and increased data discoverability. To create value, technical resources and (substitute) human resources are utilized. Human resources from network partners can be leveraged (external staff), or internal staff can be deployed for the provision or development of products and services. Technical resources may comprise of a digital infrastructure or data storage system. Trusted Data Intermediaries that utilize a digital infrastructure may create value for data contributors in the form of control in data sharing. Instrumental to value creation are governance mechanisms in the form agreements and of review processes. Such processes can either be executed by the Trusted Data Intermediary or by a network partner. Both the contributed data, as well as the data access requests may be reviewed. Especially for Trusted Data Intermediaries that handle specific, potentially sensitive data, data access review processes may be instrumental to safeguard the purpose of the data exchange. Another potential safeguarding instrument is to make bilateral agreements with data users. Trusted Data Intermediaries that operate on a non-profit motive may also implement rules to govern the network of partners, data users and data contributors. Should the Trusted Data Intermediary utilize a decentralized system for data storage, it may enforce rules through software. To cover operational costs, the Trusted Data Intermediary may opt for charging customers for the products and services offered. In particular for Trusted Data Intermediaries with a non-profit motive, substitute financial resources from a parent organisation or network partners are relevant.

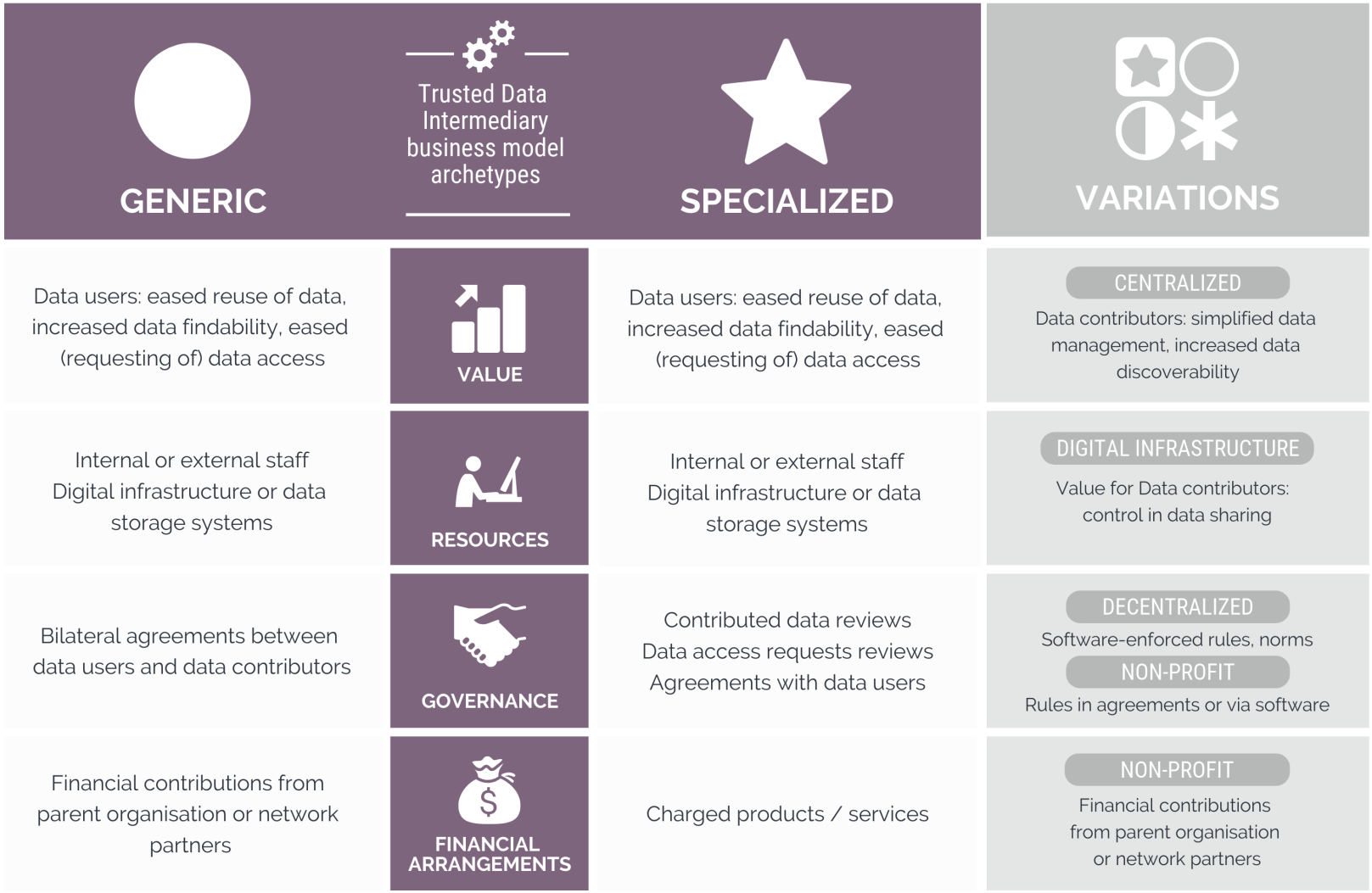


Figure 6.4: Trusted Data Intermediary business model archetypes.

7

Reflection

In this section, I reflect on findings on Trusted Data Intermediaries' business models in the light of the chosen working definition for Trusted Data Intermediaries as presented in section 2.1.2. Further, I reflect on the implications of the choice for the Unified Business Model framework to describe business models. Lastly, I reflect on the methodological limitations encountered in this thesis and their implications for the results' reliability and validity.

7.1. Reflection on definition for Trusted Data Intermediaries

Due to a lack of a definition in academic literature, I chose the following working definition for Trusted Data Intermediaries in section 2.1.2: *Trusted Data Intermediaries are entities entrusted with enabling data transactions between data contributors and data users across sectors, thereby generating value for collaborators, and enabling the creation of public value.*

From the results, we see that not all Trusted Data Intermediaries collect and aggregate data from data contributors and disperse this from one place. Nor are Trusted Data Intermediaries by default non-profit oriented. Trusted Data Intermediaries facilitate public-private relations, as well as relations within resp. the private sector and public sector. The transactions enabled by Trusted Data Intermediaries concern data sets, information on data (metadata), or products and services based on data. Transactions are fully automated by information technology. For this purpose, Trusted Data Intermediaries create platforms, digital infrastructures or data storage systems.

These findings overlap with Digital Civil Society (2017)'s definition for Trusted Data Intermediaries. However, the results also contradict this definition. For instance, Trusted Data Intermediaries may have a for-profit motive. Further, data is not always collected, aggregated and opened. As well, the relationship between data contributors and data users is not always negotiated by the Trusted Data Intermediary. For the other concepts and definitions in Table 2.2, similar issues arise. There is no definition present that covers all these aspects, thus there is no single definition applicable that covers all findings. Because of this, I argue that the chosen working definition holds, and that this definition

should be used in future research on intermediary entities in data collaboratives.

7.2. Reflection on the chosen theoretical framework

The business model ontology by Al-Debei and Avison (2010) was chosen and adapted to form the theoretical framework. Based on this framework, the coding scheme was created. As such, the choices made in the construction of the theoretical framework have implications for the research results.

The first implication relates to the interdependence and interrelation between the business model dimensions. I already pointed out that the theoretical framework has implications on the results in the discussion of results. In sub section 6.2, the stipulation of financial resources versus financial arrangements has been discussed. Further, in section 6.5, I discussed the internalization of governance mechanisms as financial arrangements. These examples illustrate how business model dimensions are interdependent and interrelated. This implies that the theoretical framework is vulnerable in respect to reliability of the results. However, to counter this, case study protocol in Appendix A was created that entails the operationalization of the theoretical framework, as well as the coding logic and analytical constructs.

The second implication relates to the scope of the theoretical framework. The original framework by Al-Debei and Avison (2010) contains more elements per dimension than were adopted in this thesis. The only dimensions that address all of the original elements are the value proposition dimension and the value finance dimension. Regarding the value architecture dimension, the elements regarding value-configuration and core-competency were discarded. I deem this not problematic. In their argument, Al-Debei and Avison (2010) build upon the work of Hedman and Kalling (2003, as referred to in Al-Debei and Avison, 2010). The argument rests upon the notion that configuration of resources demonstrates the competitiveness of the value proposition. Related to value-configuration is core-competency: repeatable patterns in value-configuration can create core competencies (Al-Debei and Avison, 2010), thus influencing the competitive advantage. For the purpose of this thesis, the competitive advantage of the business models was left out of scope. On the other hand, in chapter 5, examples are provided of how technical resources and human resources were configured in the provision of products and services. As such, the value configuration and core competencies are addressed implicitly. Regarding the value network dimension, the chosen operationalization deviates the most from the original business model ontology. The biggest difference is the chosen conceptualization of "governance". As described in section 3.2, the notion of governance as suggested by Al-Debei and Avison (2010) would not sufficiently answer the research question regarding key governance mechanisms. So, I chose the conceptualization of governance according to transaction cost theory (Williamson, 1998). Because of this, the dominance of network actors is not addressed by the current case business models. While, this could be an interesting point of research, because it would emphasize the role of the Trusted Data Intermediary in the network. As such, I find this an important avenue for future research to address. Furthermore, the elements regarding actors, relationships, communication flows, channels and network modes were left out. While actors and relationships were touched upon in a similar fashion as was done for value-configuration, the other elements were not represented by the results. For the purpose of answering the research question on key governance mechanisms, this has little implications as these elements are out of scope for the question. However, these elements could still be important to Trusted Data Intermediaries and their business models. Especially given the premise that Trusted Data Intermediaries can serve as a coordination mechanisms for overcoming barriers in data collaboratives, I'd argue that future research should investigate the network of Trusted Data Intermediaries more elaborately.

7.3. Reflection on methodological limitations

The chosen methodology, described in chapter 4, has several implications for the results. Section 4.5 addressed the precautions taken to cope with the methodological limitations. Here, I reflect on these precautions and discuss their impact on parsimony, logical coherence and testability of the research results. Concerning logical coherence and testability, reliability and validity of the research is discussed.

7.3.1. Parsimony of results

In the choice for a multiple-case study, Yin (2003) and Eisenhardt (1989) warned for a lack of parsimony in the representation of results. Especially given the qualitative, exploratory nature, the credibility of the findings is dependent to the line of argumentation. Here, an author may be tempted to over explain, thus creating large volumes of results.

Initially, I preferred a graphical representation of the cases' business models. However, I felt this did not do justice to the interdependence and interrelations between the four different business model dimensions. Therefore, I have chosen to represent the business models textually. The textual business models are structured as answers for the content analysis questions provided in section 4.3. A drawback of the textual representation is the resulting volume of text. To ensure parsimony, I limited the text for each business model to cover no more than 1,5 page. This had implications for my argumentation, because it required prioritizing which elements would be exemplified. I aimed to provide examples the various sources of evidence for a case. However, I emphasized examples from the content analysis of the information obtained through direct observations and the interview. Although online documentation is subject to change over time, in principle it is better accessible than the two aforementioned sources. Therefore, this choice was made.

Regarding the parsimony of the identified business model archetypes, I chose to emphasize the identified patterns in sections 6.1 to 6.4 in bold. I used the same concepts to describe the two Trusted Data Intermediary business model archetypes. Lastly, I graphically represent the two archetypes and possible variations to simplify communication on the business model constituents.

7.3.2. Reliability of the results

I deem the research design *reliable*, given the provision of a specific data sources, a case study protocol and case study database. However, the single-coder design, my lack of experience and training, as well as the bi-lingual data can impede on the reliability. To deal with these impediments, specifically for the content analysis, I have explicated the coding rules and analytical constructs, thereby attempting to increase the reliability of the research. The implications of this research design are that results may be replicated by future research, thus increasing reliability.

In reflection of the application of the research design, I learnt that evaluating the coded references regularly was necessary for the consistency in applying the coding rules. Also, it was clear that some data documents were more easily coded than others. This can be attributed to either the purpose of the text (e.g. advertising for potential customers), the grammatical tense (e.g. products that could be developed in the future) and the language (e.g. Dutch or English). For these texts, it was even more imperative to review the coded references regularly. In the analysis, I found it a challenge to do an

analysis of qualitative nature. As I am trained in quantitative analysis, I struggled for a while with achieving the level of analysis required to create the business model archetypes. Because within-case analysis and cross-case analysis can be applied in numerous different ways, I found myself rewriting sections multiple times. While this was time-consuming, it increased my familiarity with the cases as stand-alone entities and in comparison to each other. This enabled me to find new, indirect patterns and see emergent characteristics. In addition, the results held, despite a different reporting structure. As such, I feel that the results are sufficiently reliable.

7.3.3. Validity of the results

Regarding validity, the research design is *validatable in principle*, given that the research design can be adopted in a quantitative approach. With respect to external validity, the theory can only be externalized towards cases with certain characteristics, i.e. *specialization*, *profit motive* and *data storage centrality*. Because the research design was based on theoretical sampling rather than a replication logic, the external validity of the results has not been maximized. Future research should focus on external validity and test the proposed relationship between business model elements and characteristics. Because, as for all proposed relations it is possible that hidden variables impact the relation, creating interaction effects that instigate possibly false conclusions.

Similar, construct validity in the design has not been maximized, as it lacks multiple sources of evidence half of the cases. While the chain of evidence is safeguarded, the implementation of the web crawler strategy can create bias in the data sources that were selected. Furthermore, as not all cases were available for direct observations or interviews, triangulation of data did not occur. Future research may seek to strengthen the chain and sources of evidence.

Semantic and functional validity of this research design is higher than comparable researches (e.g. Janssen and Zuiderwijk (2014) and Ranerup et al. (2016)). Uncertainties surrounding semantic validity are reduced by the chosen theoretical framework. The operationalization of the business model elements specific to Trusted Data Intermediaries explicates the semantics of the elements. In addition, the explication of coding rules and analytical constructs supports the functional validity of the content analysis.

8

Conclusion

An integral part of achieving the Sustainable Development Goals is to encourage and promote effective public-private partnerships. However, such partnerships -especially those that focus on data collection, sharing or processing to address societal challenges- face barriers that impede on their effectiveness. These initiatives are stipulated as “data collaboratives”. Research has shown that barriers in matching data contribution and data use can be countered by intermediary entities. These entities - stipulated as Trusted Data Intermediaries- enable data contributors and data users to participate in or form a data collaborative, by creating value for them. So far, these entities have received little to no academic attention. Therefore, until know it remained unknown how these entities operate and how they create value for data contributors and data users.

This thesis set out with the objective to create a better understanding of the Trusted Data Intermediary as a coordination mechanism in data collaboratives, by describing Trusted Data Intermediary business model archetypes. A theoretical framework was developed that addresses the value proposition, architecture, network and finance dimensions of Trusted Data Intermediary business models. From an exploratory case study of six cases, it was found that differences and similarities between cases’ business models may be explained by Trusted Data Intermediary characteristics. From this discussion, relations between the characteristics and business model elements were proposed. Based on these relations, two business model archetypes have been developed: the **Generic Trusted Data Intermediary business model archetype** and the **Specialized Trusted Data Intermediary business model archetype**.

These archetypes serve as the first step in understanding which types of Trusted Data Intermediaries there are, and how they operate. By gaining this understanding, and an understanding of alternative Trusted Data Intermediary business models, the position of Trusted Data Intermediary’s between data contributors and data users has become more clear.

8.1. Contribution and recommendations to the academic field

This thesis was motivated by the lacking academic knowledge on intermediaries as a coordination mechanism for data collaboratives. Its findings contribute to academic knowledge on these types of intermediary entities operate in several ways. Firstly, a concept and definition for intermediary entities in data collaboratives -Trusted Data Intermediaries- is coined: *Trusted Data Intermediaries are entities entrusted with enabling data transactions between data contributors and data users across sectors, thereby generating value for collaborators, and enabling the creation of public value.* Further, this thesis systematically reviewed the state of the art on intermediary entities. Such an overview of the state of the art was lacking before. In addition, this research design demonstrated how web crawling can be used for retrieving online documentation, thereby integrating a quantitative method with a qualitative research approach. Most importantly, initial theory on Trusted Data Intermediaries is developed by proposing relations between business model components and Trusted Data Intermediary characteristics on specialization, profit motive and centrality. The findings provide support for the statements on Trusted Data Intermediaries of the Digital Civil Society (2017), thereby validating their findings on an academic basis.

While contributions to the academic field were made, knowledge gaps remain that can be addressed in future research endeavours. To begin, this research took an exploratory approach for describing an initial theory on Trusted Data Intermediary business models. The time constraints and lack of data availability has limited the options for extensive theory development. As such, future research may focus on testing the proposed relations. For instance, it can be researched whether the archetypes still hold when more cases are researched. As the field of data collaboratives will develop further, more candidate cases are likely to arise. In addition, future research can opt for researching the impact of different characteristics to the business model archetypes. The research design used in this thesis can be extended such that the theory resultant from this research serves as the propositions on which replication logic may be applied in future research. Further, the research design of this thesis can be adopted and applied to accommodate both qualitative and quantitative data. In the choice of the method for designing the multiple-case study, I opted for a design that can accommodate quantitative data. While the research area is novel and little data is currently available, in the future this may change, enabling quantitative data analysis in the multiple-case study. Thus, while future researchers have their work cut out regarding Trusted Data Intermediaries, this thesis serves as a starting point.

8.2. Contribution and recommendations to practice

Besides an academic contribution, this thesis contributes to practice. The business model archetypes can serve as a tool for Trusted Data Intermediaries to understand, analyze, communicate and manage their strategic choices. Firstly, the theoretical framework developed for this thesis may serve as a template for decision-makers in Trusted Data Intermediaries, so they can map out their own business model. Alternatively, decision-makers may learn from and be inspired by the six business models from the case study. This entails that the recognition of resources, governance mechanisms and financial arrangements may lead them to understand their value proposition better, and see possibilities to improve. For decision-makers in data collaboratives, the contribution of this research is to show how Trusted Data Intermediaries match data contributors with data users. This sheds light on the potential role of a Trusted Data Intermediary as a coordination mechanism in a data collaborative, and their potential value to addressing societal challenges.

In practice, I urge Trusted Data Intermediaries to understand and communicate what kinds of value they offer to data contributors and data users. Research on data collaboratives has shown that data contributors are more inclined to participate when it is clear to them what value they can derive from it. In the multiple-case study, I found that not all cases succeed in explicating which kinds of value are created. While, Trusted Data Intermediaries have shown to fulfill a role in abating barriers. This abatement should be explicitly addressed per customer segment. As well, any additional benefits should be specified. In identification of the kinds of value created, Trusted Data Intermediaries should be mindful of the interdependence and interrelation of resources, network relations and financial arrangements. The instruments chosen to create value can strengthen or impair on the value creation. A thorough understanding of your business model will therefore aid in reaching your objectives of addressing a grand societal challenge.

A

Case Study Protocol

This appendix provides the case study protocol for a case study with the objective to describe business model archetypes of Trusted Data Intermediaries by answering the research questions provided below. The following operational procedures apply:

1. Data collection procedures
   (a) For each case, establish the entity website (or the "about"-webpage), search for any relevant reports or other documentation
   (b) Go to *https://github.com/mflipsen/Crawler*, download the Crawler program. Unzip the file. Ensure a clean installation of Python version 3.7 or higher.
   (c) Open the files *setup.py* and *run.py* a Python environment of choice. In *setup.py*, enter the case name in quotation marks. Subsequently, open and run *run.py*.
   (d) Utilize other sources of evidence, e.g. direct observations or structured interviews (along the interview questions provided below).
2. Within-case analysis procedures
   (a) Apply content analysis along the coding rules provided below. Store the coded output as a copy, or make use of an environment for qualitative research and its database (for example, NVivo software).
   (b) Interpret the data for each business model element along the analytical constructs provided below. Discuss the case business model constituents by the means of the coded references.
   (c) Develop a descriptive model for each case, structured per dimension and element.
3. Cross-case analysis
   (a) Discuss case characteristics
   (b) Assess differences and similarities between cases that have the same characteristics and between cases that have different characteristics

4. Case study reporting: include
   (a) the audience, research questions and theoretical framework
   (b) the coding rules and analytical constructs
   (c) a brief description of the cases
   (d) list of which data documents were retrieved, their source or reference and access date.
   (e) a descriptive model for each case
   (f) answers to the sub research questions
   (g) discussion of similarities and differences identified in the cross-case analysis
   (h) answers to the main research question in the form of business model archetypes

Research questions.

Research question.	
Main research question	What are business models for value creation by Trusted Data Intermediaries for resp. data contributors and data users?
Sub question 1	What kinds of value do Trusted Data Intermediaries offer data contributors and data users?
Sub question 2	What are key resources for value creation by Trusted Data Intermediaries?
Sub question 3	What are key governance mechanisms for value creation by Trusted Data Intermediaries?
Sub question 4	What are financial arrangements of Trusted Data Intermediaries?

Interview questions.

Dimension	**Element**	**Interview question**
Value proposition	*Target-segment*	1. Do you focus on specific target segments for data users and data contributors?
		2. Who are contributors of data?
		3. Who are the users of data?
	Product-service	4. What products or services are offered to data users?
		5. What products or services are offered to data contributors?
	Intended-value-element	6. What is the added value of these products and/or services to users of data?
		7. What is the added value of these products and/or services to data contributors?
Value architecture	*Core-resource*	8. What resources are used to offer products and/or services?
Value network	*Role*	9. Are all resources self-developed?
		10. Are all products/services self-developed?
		10a. If not, what partners are involved?
	Governance	11. Is there interaction between data contributors and users of data?
		11a. If so, in what way is this interaction coordinated?
		12. How are possible conflicts in data transactions resolved?
Value finance	*Cost-revenue-structure*	13. Does your business have a profit motive?
	Pricing-method	14. Do customers pay for products and/or services?
		14a. If so, how are prices determined?

Coding logic.

If the unit of analysis describes, refers to, or contains…	Then assign the unit to node…
An entity that shares, opens, or provides access to data or variations to these terms	*Target-segment: data contributor*
The offering for or receiving, obtaining, getting or variations to these terms of a product, service, tool or variations to these terms by a data contributor	*Product-service: data contributor*
Utility, gains, benefits, opportunities, possibilities or variations to these terms for, or the enabling or connecting or variations to these terms of a data contributor	*Intended-value-element: data contributor*
An entity that uses, downloads, processes or gains access to data or variations to these terms	*Target-segment: data user*
The offering for or receiving, obtaining, getting or offering or variations to these terms of a product, service, tool or variations to these terms by a data user	*Product-service: data user*
Utility, gains, benefits, opportunities, possibilities or variations to these terms for, or the enabling or connecting or variations to these terms of a data users	*Intended-value-element: data user*
The organizational assets deployed to create value, e.g. financial capital, software, digital infrastructure, staff, processes, systems, trust, norms and values in the network, or variations to these terms	*Core-resource*
The mechanisms or contractual institutions leveraged by the Trusted Data Intermediary to coordinate data contributors, data users and network partners, e.g. contracts, agreements, licenses or variations to these terms	*Governance*
The provision of resources, execution of processes or variations to these terms by an entity which are leveraged (or variations to this term) in the process of value creation for data users and data contributors	*Role*
The arrangements that lead to revenues for the Trusted Data Intermediary from offering products and services or receiving of funds or grants or variations to these terms, and costs incurred from the deployment of resources	*Cost-revenue-structure*
The fees, charges, subscription, costs of use or variations to these terms that are or aren't imposed on data contributors or data users	*Pricing-method*

Analytical constructs.

Content analysis question	Element	Analytical constructs
What are the target segments?	*Target-segment*	Private company / corporation / commercial parties = private-sector data contributor or user Government / municipality / = public-sector data contributor or user Individual / citizen / inhabitant (NL: inwoner) / developer / journalist = civilian data contributor or user Researcher / journal / research center / scholarly organisation = academic data contributor or user
What are the products-services offered to each target segment	*Product-service*	Customer training / support = customer service Research environment and (analytical / aggregation) tools = data analysis product Search engine / search filter / query-building tool / data grid = data search product Data repository / catalog / hosting = data storage product/service Infrastructure / APIs connection = data logistics service Data organisation / management / anonymization / evaluation tools = data management service Request form / following data sets = data access service Metadata / data set statistics / persistent identifier / dashboards / data reports / data visualizations = data informative product Grants / funds = financial product Peer pre-review feedback service = data publishing service Publishing free from pre-approval = data publishing service Data citation / accreditation = data publishing product Messaging tool / option = platform communication product
What is the intended value from offering these products and services?	*Intended-value-element*	Finding new value / creating possibilities / opening doors / room for innovation (NL: ruimte voor innovatieve oplossingen) / development new solutions / shape market = achievement of new possibilities Reduced workload / data aggregation = increased efficiency Enlist support / work together / cross-sectoral = creation of collaboration possibilities Improve business / new insights / better decision-making possibilities (NL: betere sturingsmogelijkheden) / data-driven decision-making (NL: datagedreven sturing) = improved decision-making Produce social good / advance societal good / increase data impact / create solutions for social issue = enabling public value creation Show data availability / data existence / make data discoverable = increased data discoverability Decide what to share / remain ownership / exercise control / have grip on data sharing (NL: grip op data delen) / ensure security / protected interests = increased security in data sharing Gain academic credit = increased visibility Easy to use / simplicity / familiarity / harmonization / reduce time = eased processes Access to data / research independence = increased data accessibility

Content analysis question	Element	Analytical constructs
		Search features / capabilities / query = increased data findability
What are the core resources utilized to offer products and services to target segments?	*Core-resource*	Commission / team = staff Infrastructure / platform / API / = technical system for the data logistics Data management system = technical system for data contributors to manage their data Data storage facilities / repository = technical system where data is stored or hosted Tools / search engine = technology for creating or accessing specific information in the technical systems Software / code = technology that powers products and services on a technical system and delivers this to customers
What are roles of network partners?	*Role*	Partner provides user support / administers / manages / executes processes = partner provides replaceable resources Entity provides funds = partner provides replaceable, possibly indispensable financial resources Partner contribute to software development = partner provides replaceable resources Partner participates in design of products/services = partner provides replaceable resources Partner manages the case organisation = partner is a parent organisation Partner provides hosting / data storage systems = partner provides indispensable, irreplaceable (due to customization) resources Partner powers technical systems = partner provides indispensable resources
What mechanisms or contractual institutions are leveraged to govern the network?	*Governance*	Market rules = rules imposed on data transactions and market participants Data (contributor/ use/ exchange) agreements = governance of data transactions through agreements Waiver / licenses (NL: licenties) = access / data use permissions for data users Norms = regulation of how products/services may be used Terms (of use) / conditions = rules that customers accept or agree to as conditional to the use of products/services Verification / review = series of steps to ensure compliance with rules and/or agreements Target segments can/cannot/must/should/are (not) allowed… = regulation of how products/services may be used Procedures / request forms = implementation of rules and/or agreements
What is the cost-revenue structure related to the provision of products and services to target segments?	*Cost-revenue-structure*	Funds / grants = financial resources from network partners Not-for-profit basis = non-profit motive Commercially oriented = for-profit motive Expenses = costs incurred from offering products/services
What pricing methods are imposed on target segments?	*Pricing-method*	Free = price of products/services = customers do not pay for products and services Commission = price of products/services = customers pay for products and services Cost to share data = price of products/services = customers pay for products and services Contributions = price of products/services = customers pay for products and services

References

Natalia Adler, Ciro Cattuto, Kyriaki Kalimeri, Daniela Paolotti, Michele Tizzoni, Stefaan Verhulst, Elad Yom-Tov, and Andrew Young. How search engine data enhance the understanding of determinants of suicide in India and inform prevention: Observational study. *Journal of Medical Internet Research*, 21(1), 2019. 10.2196/10179.

Mutaz M Al-Debei and David Avison. Developing a unified framework of the business model concept. *European Journal of Information Systems*, 19(3):359–376, 6 2010. 10.1057/ejis.2010.21.

Raphael Amit and Christoph Zott. Value creation in e-business. *Strategic Management Journal*, 22(6-7):493–520, 2001. 10.1002/smj.187.

James E. Austin and Maria May Seitanidi. Collaborative Value Creation. *Nonprofit and Voluntary Sector Quarterly*, 41(6):929–968, 2012. 10.1177/0899764012454685.

Joseph P Bailey and Yannis Bakos. An Exploratory Study of the Emerging Role of Electronic Intermediaries. *International Journal of Electronic Commerce*, 1(3):7–20, 1997.

Pieter Ballon. *Control and Value in Mobile Communications*. PhD thesis, Vrije Universiteit Brussel, 2009. URL https://dx.doi.org/10.2139/ssrn.1331439.

H K Barghava and V Choudhary. Economics of an electronic intermediary with aggregation benefits. *Information Systems Research*, 15(1):22–36, 2004.

Inge Bleijenbergh. Case Selection. In Albert J. Mills, Gabrielle Durepos, and Elden Wiebe, editors, *Encyclopedia of Case Study Research*, pages 61–63. SAGE Publications, Los Angeles, CA, 2010. ISBN 9781412956703.

Harry Bouwman, Henny de Vos, and Timber Haaker. *Mobile Service Innovation and Business Models*. Springer-Verlag Berlin Heidelberg, 2008. ISBN 978-3-540-79238-3.

Harry Bouwman, Mark De Reuver, Sam Solaimani, Dave Daas, Paul Iske, and Bastiaan Walenkamp. Business Models Tooling and a Research Agenda. *25th Bled eConference*, pages 1–28, 2012.

Gisele da Silva Craveiro and Claudio Albano. Open data intermediaries: coproduction in budget transparency. *Transforming Government: People, Process and Policy*, 11(1):119–131, 2017. 10.1108/TG-12-2015-0057.

Digital Civil Society. Trusted Data Intermediaries, 1 2017. URL https://medium.com/the-digital-civil-society-lab/trusted-data-intermediaries-c58426c2c994.

Jan Dul and Tony Hak. *Case Study Methodology in Business Research*. Elsevier Ltd., 2008. ISBN 978-0-7506-8196-4.

Vincent J Duriau, Rhonda K Reger, and Michael D Pfarrer. A Content Analysis of the Content Analysis Literature in Organization Studies. *Organizational Research Methods*, 10(1):5–34, 2007.

Joan C. Durrance, Dana Walker, Maria Souden, and Karen E. Fisher. The role of community-

based, problem-centered information intermediaries in local problem solving. *Proceedings of the American Society for Information Science and Technology*, 43(1):1–17, 2007. 10.1002/meet.14504301181.

Kathleen M Eisenhardt. Building Theories from Case Study Research. *The Academy of Management Review*, 14(4):532–550, 1989. URL https://www.jstor.org/stable/258557.

Omar A. El Sawy and Francis Pereira. Digital Business Models: Review and Synthesis. In *Business Modelling in the Dynamic Digital Space*. Springer Berlin Heidelberg, 2013a. ISBN 978-3-642-31765-1. https://doi.org/10.1007/978-3-642-31765-1{_}2.

Omar A. El Sawy and Francis Pereira. VISOR: A Unified Framework for Business Modeling in the Evolving Digital Space. In *Business modeling in the dynamic digital space; an ecosystem approach*, chapter 3. Springer, 2013b. ISBN 978-3-642-31764-4. 10.1007/978-3-642-31765-1. URL http://link.springer.com/10.1007/978-3-642-31765-1.

A Michael Froomkin. The Essential Role of Trusted Third Parties in Electronic Commerce. *Oregon Law Review*, 75:49–115, 1996.

Mila Gasco-Hernandez, Wenhui Feng, and J. Ramon Gil-Garcia. Providing Public Value through Data Sharing: Understanding Critical Factors of Food Traceability for Local Farms and Institutional Buyers. *Proceedings of the 51st Hawaii International Conference on System Sciences*, 9:2276–2285, 2018. 10.24251/hicss.2018.285.

General Assembly Resolution 70/1. Transforming our world: the 2030 Agenda for Sustainable Development, 10 2015. URL https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E.

Jaap Gordijn. Value Creation from E-Business Models. In Wendy L Curry, editor, *Value Creation from E-Business Models*, chapter 5, pages 98–127. Butterworth-Heinemann, 2004. ISBN 9780750661409.

Roger V Gould and Roberto M Fernandez. Structures of Mediation : A Formal Approach to Brokerage in Transaction Networks. *Sociological Methodology*, 19:89–126, 1989.

Jukka Heikkila, Marikka Heikkila, and Markku Tinnila. The Role of Business Models in Developing Business Networks. In T Saarinen, M Tinnila, and A Tseng, editors, *Electronic Commerce*, pages 221–231. IGI Global, 2005. 10.4018/978-1-59904-943-4.ch020.

Susan C Herring. Web Content Analysis: Expanding the Paradigm. In Jeremy Hunsinger, Lisbeth Klastrup, and Matthew Allen, editors, *International Handbook of Internet Research*, pages 233–249. Springer Netherlands, Dordrecht, 2009. 10.1007/978-1-4020-9789-8{_}14. URL http://link.springer.com/10.1007/978-1-4020-9789-8_14.

Marijn Janssen and Alexander Verbraeck. Evaluating the Information Architecture of an Electronic Intermediary. *Journal of Organizational Computing and Electronic Commerce*, 15(1):35–60, 2005. 10.1207/s15327744joce1501.

Marijn Janssen and Anneke Zuiderwijk. Infomediary Business Models for Connecting Open Data Providers and Users. *Social Science Computer Review*, 32(5):694–711, 2014. 10.1177/0894439314525902.

Thorhildur Jetzek, Michel Avital, and Niels Bjørn-Andersen. Generating Value from Open Government Data. *International Conference on Information Systems*, (July 2016):62–82, 2013.

Alexandre Joyce and Raymond L. Paquin. The triple layered business model canvas: A tool to design more sustainable business models. *Journal of Cleaner Production*, 135:1474–1486, 11 2016. 10.1016/j.jclepro.2016.06.067.

Jannis Kallinikos, Aleksi Aaltonen, and Attila Marton. The Ambivalent Ontology of Digital Artifacts. *MIS Quarterly*, 37(2):357–370, 2013. URL https://www.jstor.org/stable/43825913.

Sung Min Kim and Young Guk Ha. Automated discovery of small business domain knowledge using web crawling and data mining. *2016 International Conference on Big Data and Smart Computing, BigComp 2016*, pages 481–484, 2016. 10.1109/BIGCOMP.2016.7425974.

Gary King. An introduction to the dataverse network as an infrastructure for data sharing. *Sociological Methods and Research*, 36(2):173–199, 2007. 10.1177/0049124107306660.

Gary King and Nate Persily. A New Model for Industry-Academic Partnerships, 1 2019. URL https://gking.harvard.edu/files/gking/files/partnerships.pdf.

Bram Klievink, Nitesh Bharosa, and Yao Hua Tan. The collaborative realization of public values and business goals: Governance and infrastructure of public-private information platforms. *Government Information Quarterly*, 33(1):67–79, 2016. 10.1016/j.giq.2015.12.002. URL http://dx.doi.org/10.1016/j.giq.2015.12.002.

Bram Klievink, Haiko Van Der Voort, and Wijnand Veeneman. Creating value through data collaboratives. *Information Polity*, 23:379–397, 2018. 10.3233/IP-180070.

Carol Kort and Jaap Gordijn. Modeling Strategic Partnerships Using the E3value Ontology: A Field Study in the Banking Industry. In Peter Rittgen, editor, *Handbook of Ontologies for Business Interaction*, chapter 18, pages 310–325. 2008. 10.4018/978-1-59904-660-0.ch018.

Klaus Krippendorff. *Content Analysis: An Introduction to Its Methodology*. SAGE Publications, 2 edition, 2004. ISBN 978-0761915454. URL http://www.uk.sagepub.com/textbooks/Book234903.

Wouter Los. Towards an internationally trusted exchange of data. Technical report, AMdEX, Amsterdam, 1 2018.

Gustavo Magalhaes, Catarina Roseira, and Sharon Strover. Open government data intermediaries : a terminology framework. In *Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance*, 2013. 10.1145/2591888.2591947.

Bill McLaughlin. Types of Interviews Fundamentals of Research CSS505, 2006.

Eunice Mercado-lara and J Ramon Gil-garcia. Open Government and Data Intermediaries : The Case of AidData. In *Proceedings of the 15th Annual International Conference on Digital Government Research*, pages 335–336, 2014. ISBN 9781450329019.

Steve Muylle and Amit Basu. Online support for business processes by electronic intermediaries. *Decision Support Systems*, 45(4):845–857, 2008. ISSN 01679236. 10.1016/j.dss.2008.02.005.

Christian Nielsen and Morten Lund. An Introduction to Business Models. *Ssrn*, (March 2015), 2015. 10.2139/ssrn.2579454.

Aviral Nigam. Web Crawling Algorithms. *International Journal of Computer Science and Artificial Intelligence*, 4(3):63–67, 2014. ISSN 22264450. 10.5963/ijcsai0403001.

Alexander Osterwalder and Yves Pigneur. *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. 2010. ISBN 9782839905800.

Michael Porter and Mark Kramer. Creating Shared Value. *Hardvard Business Review*, (February), 2011. ISSN 00178012. 10.1108/09600039410055963.

Bertrand V Quelin, Ilze Kivleniece, and Sergio Lazzarini. Public-Private Collaboration , Hybridity and Social Value : Towards New Theoretical Perspectives. *Journal of Management Studies*, 54(6):763–792, 2017. 10.1111/joms.12274.

Agneta Ranerup, Helle Zinner Henriksen, and Jonas Hedman. An analysis of business models in Public Service Platforms. *Government Information Quarterly*, 33(1):6–14, 2016. 10 .1016/j.giq.2016.01.010.

Horst W. J. Rittel and Melvin M. Webber. Dilemmas in a general theory of planning. *Policy Sciences*, 4(2):155–169, 6 1973. ISSN 0032-2687. 10.1007/BF01405730. URL http://link.springer.com/10.1007/BF01405730.

N Robin, T Klein, and J Jutting. Public-Private Partnerships for Statistics: Lessons Learned, Future Steps: A focus on the use of non-official data sources for national statistics and public policy. Technical report, OECD Publishing, Paris, 2 2016.

Chandni Saini and Vinay Arora. Information retrieval in web crawling: A survey. In *2016 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2016*, pages 2635–2643. IEEE, 2016. 10.1109/ICACCI.2016.7732456.

Djoko Sigit Sayogo, Jing Zhang, Theresa A. Pardo, Giri K. Tayi, Jana Hrdinova, David F. Andersen, and Luis Felipe Luna-Reyes. Going beyond open data: Challenges and motivations for smart disclosure in ethical consumption. *Journal of Theoretical and Applied Electronic Commerce Research*, 9(2):1–16, 2014. 10.4067/S0718-18762014000200002.

Jason Seawright and John Gerring. Case Selection Techniques in Case Study Research. *Political Research Quarterly*, 61(2):294–308, 2008.

P J Skevington and T P Hart. Trusted third parties in electronic commerce. *BT Technology Journal*, 15(2):39–44, 1997.

Lea Stadtler and Gilbert Probst. How broker organizations can facilitate public-private partnerships for development. *European Management Journal*, 30(1):32–46, 2012. 10.1016/j.emj.2011.10.002. URL http://dx.doi.org/10.1016/j.emj.2011.10.002.

Iryna Susha, Marijn Janssen, and Stefaan Verhulst. Data collaboratives as "bazaars"? *Transforming Government: People, Process and Policy*, 11(1):157–172, 2017a. 10.1108/tg-01-2017-0007.

Iryna Susha, Marijn Janssen, and Stefaan Verhulst. Data Collaboratives as a New Frontier of Cross-Sector Partnerships in the Age of Open Data: Taxonomy Development. In *Proceedings of the 50th Hawaii International Conference on System Sciences*, pages 2691–2700, 2017b. http://hdl.handle.net/10125/41480.

Iryna Susha, Marijn Janssen, Stefaan Verhulst, and Theresa Pardo. Data Collaboratives: How to Create Value from Data for Public Problem Solving? In *Proceedings of the 18th Annual International Conference on Digital Government Research - dg.o '17*, pages 604–606, 2017c. 10.1145/3085228.3085309.

Iryna Susha, Åke Grönlund, and Rob Van Tulder. Data driven social partnerships: Exploring an emergent trend in search of research challenges and questions. *Government*

Information Quarterly, 36(1):112–128, 2019. 10.1016/j.giq.2018.11.002.

Karl Täuscher and Sven M. Laudien. Understanding platform business models: A mixed methods study of marketplaces. *European Management Journal*, 36(3):319–329, 2018. 10.1016/j.emj.2017.06.005.

David J. Teece. Business models, business strategy and innovation. *Long Range Planning*, 43(2-3):172–194, 2010. 10.1016/j.lrp.2009.07.003.

François van Schalkwyk, Michelle Willmers, and Maurice McNaughton. Viscous Open Data: The Roles of Intermediaries in an Open Data Ecosystem. *Information Technology for Development*, 22:68–83, 2016. 10.1080/02681102.2015.1081868.

Mario Vilas. Python-googlesearch, 2018. URL https://python-googlesearch.readthedocs.io/en/latest/.

Vivli. 2018 Progress Report. Technical report, Vivli Center for Global Clinical Research Data, 2018. URL http://www.pageturnpro.com/Vivli-Inc/87942-Vivli-2018-Progress-Report/flex.html.

Robert Philip Weber. *Basic Content Analysis*. SAGE Publications, Newbury Park, CA, 2 edition, 1990. ISBN 0803938632.

Jane Webster and Richard T Watson. Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 26(2):xiii – xxiii, 2002. 10.1.1.104.6570.

Peter Weill and Michael Vitale. Portals, Agents, Auctions, Aggregators and Other Intermediaries. In *Place to space: Migrating to Ebusiness models*, chapter 7, pages 151–182. Harvard Business Press, 2001. ISBN 9781422163436.

Oliver E. Williamson. Transaction-Cost Economics: The Governance of Contractual Relations. *The Journal of Law & Economics*, 22(2):233–261, 1979. URL https://www.jstor.org/stable/725118.

Oliver E. Williamson. Transaction cost economics: How it works; where it is headed. *Economist*, 146(1):23–58, 1998. 10.1023/A:1003263908567.

Ryan Womack. Information intermediaries and optimal information distribution. *Library & Information Science Research*, 24(2):129–155, 2002. URL https://doi.org/10.1016/S0740-8188(02)00109-3.

Bedrettin Yazan. Three Approaches to Case Study Methods in Education: Yin , Merriam , and Stake. *The Qualitative Report*, 20(2):134–152, 2015. URL https://nsuworks.nova.edu/tqr/vol20/iss2/12.

R.K. Yin. Validity and generalization in future case study evaluations. *Evaluation*, 19(3): 321–332, 2013. 10.1177/1356389013497081.

Robert K. Yin. *Case Study Research Design and Methods*, volume 5. SAGE Publications, 3 edition, 2003. ISBN 0761925538.

Ayoung Yoon, Andrea Copeland, and Paula Jo McNally. Empowering communities with data: Role of data intermediaries for communities' data utilization. *Proceedings of the Association for Information Science and Technology*, 55(1):583–592, 2019. 10.1002/pra2.2018.14505501063.

Christoph Zott, Raphael Amit, and Lorenzo Massa. The business model: Recent developments and future research. *Journal of Management*, 37(4):1019–1042, 2011. ISSN 01492063. 10.1177/0149206311406265.