

Nia Chigogidze

Matchmaking: Trusted Data Intermediaries and the Connection of Private Data to Public Needs

Master Thesis

at the Chair for Information Systems and Information Management
(Westfälische Wilhelms-Universität, Münster)

Supervisor: Prof. Dr. Dr. h.c. Dr. h.c. Jörg Becker
Tutor: Sebastian Halsbenning

Presented by: Nia Chigogidze
Schlossplatz 2
48149 Münster
+49 251 8338100
nchigogi@uni-muenster.de

Date of Submission: 2021-08-04

Content

Figures

Tables

Abbreviation

CDRC	Consumer Data Research Centre
DDI	Dalberg Data Insights
GDPR	General Data Protection Regulation
ICT	Information Communication Technology
ILVO	Flanders Research Institute for Agriculture, Fisheries and Food
MSP	Multi-Sided Platforms
UBDC	Urban Big Data Centre

1 Introduction

In today's ever-evolving world, public administrations everywhere face increasingly complex, interconnected and widespread social problems (Head & Alford, 2015). In response to these issues, the need for joint action and improved collaboration across sectors has gained prominence and wide attention as an essential element for effective response to these challenges (Klievink, Van der Voort & Veeneman, 2018). This thesis addresses one type of such cross-sectoral initiatives aimed at tackling societal issues through the means of private data driven insights, trusted data intermediaries (Susha, Janssen & Verhulst, 2017a).

Section 1.1. of this chapter will provide background information on the relevance of private data for public needs and an overview of the key characteristics and challenges of private-public data sharing. Section 1.2. will give an introduction to data intermediaries as initiatives aimed at supporting private-public data exchange. It will also introduce trusted data intermediaries as a specific type of such initiative and give reasoning as to why it is particularly deserving of study. Section 1.3 will identify research needs around trusted data intermediaries based on relevant literature. Lastly, section 1.4. will present the direction of inquiry of this thesis, the concrete research questions it will address and the structure of the work.

1.1 Private Data for the Public Good

The term "data revolution", referring to the explosion in the scale of data production, collection and dissemination, has steadfastly become a viral concept and with it so has an awareness of the multitude of ways in which this newfound abundance of data could stand to fundamentally alter all aspects of daily life (Susha, Pardo, Janssen, Adler, Verhulst et al., 2018). These already exponential data resources are expected to double every three years through the steady inflow of information from ever growing digital platforms, wireless sensors and mobile phone use (Henke et al., 2016).

In parallel to the rising amount of data, the capabilities for its processing have also continuously been increasing, with higher data storage capacities and progressively sophisticated analytical algorithms (ibid). These advancing data capabilities have enabled sectors from across the global economy to derive increased value and develop new forms of competition through innovation, improved service efficiency and productivity gains (Manyika et al., 2011). These effects have led to data-driven innovation being reported as one of the principal drivers of competitiveness and job growth in the global market (European Commission, 2018a).

This transformational power of data-driven change does not exclude the public sector. **Access to more representative and up-to-date data holds the potential to give policymakers more accurate and holisitc understanding of societal issues and empower them to solve them** (Verhulst, Young, Winowatan & Zahuranec, 2019). Moreover, improved access to relevant and accurate data can enable governments to adopt more agile and evidence-based decision-making processes, bringing them closer to citizen needs (Susha, Janssen & Verhulst, 2017a). In particular, it has been widely argued that data can be a valuable source in areas such as community economic development, disaster management and citizen-driven decision making (Yoon, Copeland & McNally, 2018). Recognising this potential value, many governments have begun opening their datasets and taking measures to encourage data use and reuse in supporting the development of innovations, services and economic growth (Klievink, Van der Voort & Veeneman, 2018).

However, much of the **most valuable and comprehensive data best suited for helping public sphere decision-makers achieve such results lies in the private sector** (Verhulst et al., 2019). This is due to the large amount of data that citizens in their capacity as consumers provide to private sector companies daily in the forms of financial transaction records, call logs and social media activities (Gil-Garcia, Chengalur-Smith & Duchess, 2007). Further enhanced by the modern-day proliferated use of apps, platforms and sensors, more and more of real-time and accurate data on how people and societies behave is currently privately owned (Verhulst & Sangokoya, 2015). Consequently, much of the data with the potential to transform how critical societal issues are tackled resides out of reach for relevant decision-makers (Susha et al., 2018).

Thus, it has become increasingly clear that in order to address some of society's most pressing and complex challenges, collaboration and coordination in data sharing and analysis among different actors, among them non-governmental actors such as private companies, is vital (Susha, Grönlund & Van Tulder, 2019). In response to this rationale, their developed a **collaborative approach to data-driven innovation,** a practice aimed at creating new public value through a combination of data and expertise from various sources (Klievink, Van der Voort & Veeneman, 2018).

The positive effects of multi-stakeholder data partnerships have on several occasions been demonstrated in practice, in cases where private company data has helped government agents more quickly and effectively react to societal crises, plan urban projects and manage migration flows (Alemanno, 2020; Verhulst & Sangokoya, 2015). The most recent prominent examples have been the multitude of private-public data-sharing initiatives developed for tackling the COVID-19 pandemic, centred around

infection and contact tracing, as well as the development of a deeper understanding of the pandemic's economic and societal effects (Data Collaboratives Explorer). These examples showcase that the benefits that the public sector stands to gain from access to private data are wide-ranging.

Private data access can also signal significant financial savings for governments through avoided spending for the collection of data that is already existent in the private sector (Martens & Duch-Brown, 2020). These savings can prove to be especially beneficial for developing nations, which due to a lack of resources, face a particular shortage of access to datasets containing insights into the best courses of action for their developmental and societal needs (Susha, Janssen & Verhulst, 2017a).

Though beneficial, such **cross-sectoral data sharing and cooperation is often challenging to execute in practice due to a myriad of institutional and technological barriers** (Groves & Neufeldi, 2017). The most common barriers are the concerns over data confidentiality and potential risks to a company's commercial interest tied to the data being shared (European Commission, 2018b). These concerns stem, in part, from the difference in values, organisational cultures, incentives and goals of the wide set of different actors that such data-sharing collaborations have to account for (Klievink, Van der Voort & Veeneman, 2018). Other types of barriers relate to the legal regulation of data sharing, possible malicious breaches of data privacy, and liability issues for data leakage or misuse (Groves & Neufeldi, 2017). Moreover, barriers can also come in the form of differing levels of data literacy, mechanisms for data management and lack of human and financial resources for the effective use of data, as the processing of data can be both an expensive and a complex process (European Commission, 2018b; Van Schalkwyke, Cañares, Chattapadhyay & Andrason, 2016).

The supply and demand for data across government and the private sector are also widely dispersed, leading to issues of their effective matching (Verhulst et al., 2019). Mismatches in public sector needs and available private sector data could lead to compromised data sharing relationships, discouraging future cooperation (Susha, Janssen & Verhulst, 2017b). Additional challenges for any data-sharing partnership might stem from the ad hoc nature of most sharing agreements, which can suffer from a lack of clearly defined roles and structure (Robin, Klein & Jütting, 2016).

To help to address these challenges and support easier exchange of data and cooperation on social issues across sectors, there has been a recent upsurge in a new type of practice and organisation, referred to as data collaboration and data collaboratives (Verhulst et al., 2019).

1.2 Data Matchmakers - Trusted Data Intermediaries

Data collaboratives are a general term that refers to initiatives aimed at fostering collaboration among private, public and academic sectors around data sharing and analysis for the public good (Susha, Janssen & Verhulst, 2017a). They represent initiatives that, at their essence, are driven by three core elements: **i.** recognition of complex and pressing societal issues, often referred to as "wicked" problems (Susha & Gil-Garcia, 2019), **ii.** data-driven innovation, and **iii.** the belief that public governance stands to benefit from collaboration with non-governmental actors (Klievink, Van der Voort & Veeneman, 2018).

Data collaboratives as a concept encompass many different organisational forms, the choice of which can be influenced by the type of data being shared, diversity of data providers and relevant policy issues, among other factors (Susha, Janssen & Verhulst, 2017b). One such form of data collaboratives is the trusted data intermediary (Verhulst et al., 2019). **Trusted data intermediaries** are third-party actors that support collaboration between private sector data providers and other non-private actors (ibid).

According to the typology developed by Verhulst et al. (2019), two main features that characterise trusted data intermediaries are: their restricted accessibility, meaning that only pre-selected and authorised partners can access the corporate data being shared; and the independent usability of data, meaning that private sector data holders have limited involvement over how the data they shared is used, allowing it to be independently applied for varying purposes.

Several factors make **trusted data intermediaries a particularly relevant and exciting form of data collaboratives for study.** Firstly, they make up 20% of all data collaboratives, putting them in the top three most widespread forms of data sharing initiatives (ibid). Secondly, the characteristics of restricted accessibility and independent usability of data make them arguably the most optimal mechanism for connecting private data to public issues, as they mitigate the risks of data sharing for private actors by limiting the access pool to only pre-screened and trustworthy entities, while simultaneously allowing for the multi-purpose use of available data, thus increasing the potential public value that can be derived from it. Lastly, through the nature of their form and operations, trusted data intermediaries are capable of going beyond the simple facilitation of connections among actors to providing additional support in the forms of processing data, democratising data, limiting the cost of data transactions and helping their partner actors in building their data capacities (Janssen & Zuiderwijk, 2014; Yoon, Copeland & McNally, 2018).

1.3 Knowledge Gap - The Linking Practices of Trusted Data Intermediaries

As data collaboratives in general and trusted data intermediaries in particular are a comparatively novel concept, there is a lack of empirical work examining them (Klievink, Van der Voort & Veeneman, 2018; Susha, Grönlund & Van Tulder, 2019). Moreover, whilst the number of such initiatives in practice has been increasing, academic research has not yet developed a systematised insight into this type of partnership (Susha, Grönlund & Van Tulder, 2019).

In particular, more research is needed around the topic of how data collaboratives can be formed around shared interests and capacities to address particular problems and what tools they might use to more effectively identify relevant data sources for particular problems (Susha et al., 2018). Current literature provides limited insight on how information sharing is achieved in practice, in particular in the area of voluntary private-sector data sharing for public value (Susha, Rukanova, Gil-Garcia, Hua Tan & Gasco, 2019). Moreover, it is also necessary to understand the coordination problems data collaboratives might experience in practice and what type of resources are available to address them (Susha, Janssen & Verhulst, 2017b). While necessary for any organisational effort, such collaboration mechanisms are particularly relevant for data intermediaries that have to deal with an exceptionally diverse set of stakeholders, underlining the importance of gaining more extensive insight in this area (Susha & Gil-Garcia, 2019). The need for more knowledge around the applied mechanisms that constitute success factors for data collaboratives in facilitating data sharing has also been expressed by practitioners (Verhulst, Susha & Kostura, 2016).

Of the several issues identified as key topics for future research for advancing data collaboratives practices as a novel research direction by Susha et al. (2018), one has been the area of collaborative capabilities. Namely, a research agenda should focus on the capabilities required to launch a collaborative and keep it operating effectively (ibid). Some of the critical questions that should be addressed in this line of inquiry include what capabilities are required of a data collaborative to provide timely access to private data and how these capabilities vary depending on different contexts (ibid).

Lack of knowledge is even more prominent when it comes to the specificities of trusted data intermediaries as a form of data collaboratives. There is a lack of awareness around what resources trusted data intermediaries utilise and what mechanism they leverage to coordinate network actors (Flipsen, 2019). More inquiry is recommended into the elements of data intermediaries' qualifications and social affiliations and how these characteristics enable them to enter the field of data exchange and play an intermediary role (Van Schalkwyk et al., 2016).

To ensure their effective functioning, trusted data intermediaries require more knowledge on how they can facilitate the communication of data demands to the right supplier and effectively fulfil their matchmaking role, which in turn requires the identification and study of current and emerging practices in this area (Verhulst, Susha & Kosturia, 2016).

The majority of above-discussed elements of areas of current knowledge gaps can be grouped under the general title of **linking practices,** which according to Van Schalkwyk et al. (2016), can be defined as the collection of all the measures, mechanisms and capabilities that data intermediaries rely on in the process of connecting actors and facilitating the flow of data between them.

1.4 Research Objectives and Structure: Linking Practices, Capitals and Commonalities Among Trusted Data Intermediaries

This thesis aims to better the understanding of the linking practices of trusted data intermediaries by investigating and describing their related behaviour and activities. Therefore, this thesis will address the following question: What are the liking practices employed by trusted data intermediaries in the process of facilitating connections and the flow of data among private and public actors?

In order to better examine and differentiate among varying linking practices employed by trusted data intermediaries, the inquiry needs to be guided by a theoretical framework for their classification and grouping. Therefore, for the purposes of this thesis and for the reasons outlined in Chapter 3, the theory of different types of organisational capital proposed by Bourdieu (1986) will be utilised as the appropriate framework. As such, in order to accommodate for this direction of inquiry, the sub question to this first research question of this work will be: What type of capital do the identified linking measures undertaken by trusted data intermediaries fall under?

Moreover, to provide more generalisable takeaways regarding the linking practices of trusted data intermediaries, instead of only individual and select cases, comparisons should be made on whether any similarities or dominant patterns of linking practices emerge among trusted data intermediaries. Hence, the last question posed by this thesis will be: How do linking practices concur and differ across trusted data intermediaries?

An overview of all the research questions of the given thesis can be seen in Table 1.

Table 1: Thesis Research Questions

	Research Questions
RQ 1:	What are the liking practices employed by trusted data intermediaries in the process of facilitating connections and the flow of data among private and public actors?
Sub-RQ1:	What type of capital do the identified linking measures undertaken by trusted data intermediaries fall under?
RQ 2:	How do linking practices concur and differ across trusted data intermediaries?

To address the outlined questions and contribute to the filling of the identified research gap the given work will pursue the following approach: First, it will investigate the extent of the existent knowledge of intermediary linking practices through a review of relevant literature; Second, it will choose a theoretical framework suitable to the systematisation of data intermediary linking practices; Third, it will adopt the methodological approach of a descriptive multiple-case study and gather data on selected cases through a combination of primary and secondary data; Lastly, it will conduct a within-case and cross-case analysis, to identify both individual and common linking measure employed withing studied data intermediaries. The summary of the structure can be seen in Figure 1.

Figure 1: Overview of the Thesis Structure

1	Introduction	Background information on trusted data intermediaries, their relevance to the public sector and the research need. Outline thesis questions and the method of inquiry.	
2	Literature Review	Overview of the different concepts of data intermediaries and justification of the working definition of Trusted Data Intermediary. Review of academic knowledge on data intermediary linking practices systematized according to thematic areas.	
3	Theoretical Framework	Presentation of Bourdieu's Theory of Capital and its use in the organisational and data intermediation contexts	
4	Methodology	Description of the choses method of a multiple-case study research approach. Overview of the case selection criteria and selected cases. Presentation of chosen data collection methods – interviews and online documentation analysis.	
5	Findings	Discussion of within-case observations of linking measures and their classification on the basis of the theoretical framework.	Answer to RQ 1 and Sub-Question 1
6	Discussion	Cross-case analysis of findings and identification of common and diverging patterns.	Answer to RQ 2
7	Limitations & Future Research	Summary of the limitations suffered by the research and recommendations for future research.	

2 Literature Review

This chapter presents a review of literature on trusted data intermediaries. Section 2.1. gives an overview of different concepts and terminology used in the literature on organisations facilitating data sharing or exchange. It concludes by presenting and justifying the working definition of trusted data intermediaries. In Section 2.2. dimensions of linking measures employed by data intermediaries are discussed. Finally, the chapter concludes with the identification of the gap in current literature.

To ensure that the review encompassed all relevant literature, several academic databases were consulted, namely: Google Scholar, JSTOR, Limo (the database linked with KU Leuven), SpringerLink, Scopus and Web of Science. Due to the fragmented nature of relevant literature, several keywords were searched. The initially searched keywords were: "data collaborative", "trusted data intermediary", "data intermediary", "business-to-government data sharing". From the articles that were found as a result of this quarry, the related keywords of "infomediary", "data broker" and "data philanthropy" were discovered and also subsequently searched. Supplementary research was conducted by examining the bibliography of authors most prominent in the field.

The search results were then filtered according to relevance. Works related to business-to-business data exchange, inter-governmental data exchange or too sector-specific works were eliminated to narrow down the scope of the literature. The overview of the remaining literature consequently analysed can be seen in **Appendix A**.

2.1 Intermediaries in Data Collaboration

There is no one clear definition for what constitutes a data intermediary in literature. Instead, there are numerous concepts that describe a similar phenomenon of data sharing facilitation or its particular aspects (Flipsen, 2019), with stakeholders often labelling comparable initiatives with different terms (Susha, Grönlund & Van Tulder, 2019).

To address this terminological ambiguity, this section will identify and present an overview of different definitions of intermediation in the context of data exchange. Finally, the section will conclude by presenting and justifying the chosen working definition of a data intermediary.

2.1.1 Concepts of Data Intermediaries

The concept of intermediaries was first discussed within the context of information communication as a supporting organisation that linked different agents and aided the transfer of information among them (Yoon, Copeland & McNally, 2018). It later went on

to be explored in different organisational contexts from business to e-government (ibid). Thus, over time, intermediaries have been viewed as bridging organisations, brokers, agents or support organisations, with the unifying factor being that intermediaries are entities located in between two or more parties (Van Schalkwyk et al., 2016).

Bailey and Bakos (1997) were one of the first to develop the notion of **electronic intermediaries** concerned with information technologies and, specifically, data exchange. According to them, electronic intermediaries can be characterised by their performance of four main roles: aggregation of information, matching of suppliers and customers, creation of trust and provision of marketing information to suppliers (ibid).

Related to the roles of electronic intermediaries as seen by Bailey & Bakos (1997), is the concept of **information intermediaries** put forward by Womack (2002), which also emphasises the information aggregation role of intermediaries (Flipsen, 2019). Womack (2002) defined information intermediaries as "*any system that mediates between the producers and consumers of information"* (p.133), but more specifically, the ones that "*collect, organise, and distribute information to their clients*" (p.129). Furthermore, it emphasised that a vital aspect of an information intermediary was to offer distinct added value to the client that the latter would not have been able to access independently (ibid).

A variation on the concept of information intermediaries, the term of **infomediary** has been used to describe the process of matching demand and supply through the aggregation of desired information and provision infrastructure that together help build trust between the supplier and the user (Sein & Furuholt, 2012). Hagel and Rayport (1997) define infomediaries as "*a broker or agent who helps users to manage the vast amount of information and safeguard their privacy*" (as cited in Janssen & Zuiderwijk (2014), p. 695). Furthermore, Janssen & Zuiderwijk (2014) developed a sub-categorisation of infomediaries based on the two variables of the level of access to data (whether the data provided is raw or predefined) and level of dialogue (the degree to which the end-users can interact with the information). Thus, infomediary is a well-established term, which depending on the field under consideration, may possess different meanings and refer to individuals, entities or systems (Magalhaes, Roseira & Strover, 2013).

An example of this is its alternative conceptualisation within the field of open government adopted by Magalhaes, Roseira & Strover (2013), who define infomediaries as actors who provide products or services to citizens or other third parties on the basis of public sector data. This work acknowledges that this more recently developed understanding of the term creates an overlap with other concepts related to deriving public value from open or government data, such as open data services or civil start-ups (ibid).

Similar research also focusing specifically on open data by Van Schalkwyk et al. (2016) proposes a definition of an **open data intermediary,** as a middleman in the data supply chain incorporating an open dataset, the use of which would not have been possible without its involvement. This definition once again considers the likelihood that the data would have been accessed without an intermediaries' participation, reflecting the definition of information intermediary and added value by Womack (2002).

This definition of open data intermediary also links to previous research by the same author group that defined open data intermediaries as keystone species playing a vital enabling and value-adding role in the data ecosystem, as the creators of services, platforms and tools for the ecosystem actors (Van Schalkwyk, Willmers & McNaughton, 2016). Furthermore, a literature review of intermediation in the context of e-government conducted by Sorrentino & Niehaves (2010) came to a similar finding that intermediation in this field consists of: linking users (one of which is always a public sector actor), facilitating access and adding human skill or knowledge for the processing of data. These definitions demonstrate the consistent emphasis on the elements of added value and facilitation as vital to the concept of government-related data intermediation.

In opposition to the sharing of open or government data, we find corporate data sharing, an emerging practice in which companies share anonymised data with third parties for them to analyse and utilise for the public good, which has been referred to as **corporate data philanthropy** (Verhulst & Sangokoya, 2014). According to George, Yan & Leidner (2020), data philanthropy can refer to a range of activities conducted by a corporate firm, from different levels of data sharing to more elaborate cooperation and contribution of data science expertise and skills. Data philanthropy may involve one or more entities such as corporations, governments or non-profits who come together around a shared objective of tackling societal challenges (ibid).

Verhulst and Sangokoya (2014) developed a taxonomy examining emerging practices in such corporate data philanthropy. In this process, they identified **trusted intermediaries** as one of the forms of corporate data sharing. However, their proposed definition of such intermediaries was quite broad and was described as companies sharing their data with a limited number of partners for data analysis and modelling or other value chain activities (ibid).

Related to the definition put forward by Janssen & Zuiderwijk (2014) and Magalhaes, Roseira & Strover (2013) is the concept of **Multi-Sided Platforms (MSP)** proposed by Klievink, Bharosa & Tan (2016), which are public-private platforms, positioned between two or more entities that create value by enabling direct communication between different customers (as referenced in Flipsen, 2019). Similarly, to previously concepts of

infomediaries, MSP's are characterized by being located in between actors, however, this definition puts value creation, rather than competition, as at the bases of collaboration (Klievink, Bharosa & Tan, 2016). This relates the concept of MSP's to the definition of data intermediaries as keystone species proposed by Van Schalkwyk, Willmers & McNaughton (2016).

Related to both the notions of corporate data philanthropy and intermediaries' position among multiple actors is the concept of **data brokers**. Analogously to other discussed terms, there is no authoritative definition of data brokerage, and terminology varies on a regional basis and can include "information resellers," "data vendors" or "information brokers" (Rieke, Yu, Robinson & Van Hoboken, 2016). Building upon several international institutional definitions Rieke et al. (2016) define data brokers as *"a company or business unit that earns its primary revenue by supplying data or inferences about people gathered mainly from sources other than the data subjects themselves*" (p.4). Notably, according to this understanding, the customer segment of data brokers can be wide-ranging and include both private and public sector entities (ibid). This definition corresponds closely with the one put forward by Organisation on Economic Co-operation and Development (2013), according to which data brokers are "*firms that gather and merge aggregated information on individuals that are then sold for various uses"* (p.16). In addition, brokers can differ according to their target segment, degree of specialisation and the level of complexity of the information they offer (ibid).

Alternatively, Crain (2018) defines data brokers as "*companies that specialise in the collection and exchange of personal information and are usually associated with large-scale, "big data" operations*" (p.90). Crain (2018) states that though data brokers primarily source information through acquisition and second-hand trawling, they also obtain information directly from consumers to varying degrees. This definition shifts the notion away from the essential element of secondary sourcing as proposed by Rieke et al. (2016) and brings in a new notion of the scale of the data operation.

According to one of the most recent definitions, data intermediaries present a concrete example of a more significant phenomenon of cross-sectoral informational collaboration referred to as **data collaboratives** (Verhulst et al., 2019). In a widely accepted definition by Susha, Janssen & Verhulst (2017a), 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*" (p.2691). This definition is quite broad and emphasises the collaborative aspect of the parties' relationships, which takes them beyond simple agreements to share data and shifts the central focus on jointly deriving value from data (ibid). Though most collaboratives focus on creating value through private sector

data, they could also be working with data from a wide variety of sources such as open data or data from non-profit or academic stakeholders (ibid).

The concept of data collaboratives is related to that of **cross-sectoral partnerships** as put forth by Selsky & Parker (2005), which are "*cross-sector projects formed explicitly to address social issues and causes that actively engage the partners on an ongoing basis*" (as referenced in Susha et al., 2019, p.228). Although this notion was initially developed to encompass all types of collaboration, it has successfully been applied to the study of specifically information sharing based cross-sectoral collaboration (Susha et al., 2019). Similar to data collaboratives, this definition also focuses on the cross-sectoral nature of collaboration and the creation of value to address some social need (ibid). The definition of data collaboratives is also closely connected to that of **data-driven social partnerships**, which are "*a collaboration between actors in one or more sectors to leverage data from different parties, at any stage of its lifecycle, for public benefit in policy or science*" (Susha, Grönlund & Van Tulder, 2019, p. 119). This definition was designed to span over different research disciplines and is an overarching concept that unifies many of the above-discussed terms, such as data collaboratives and data philanthropy (ibid).

Data collaboratives encompass various organisational forms employing differing ways of sharing, accessing and using data (Susha, Janssen, & Verhulst, 2017a). According to the typology developed by Verhulst et al. (2019), such data collaboratives are capable of being differentiated based on two main variables: engagement and accessibility. The first variable of engagement refers to the degree to which any potential use of the contributed data is determined through a collaboration of data contributors and users. This can vary from independent use when data is used with a high degree of autonomy with little control from the side of the data contributor, cooperative use when the utilisation of the data is regulated jointly by the data contributor and user and determined use when data can only be handled for the specific purposes designated by data holders. The second variable of accessibility refers to the conditionality of accessing the private sector data by external parties and could be either open, meaning that there are few restrictions on which actors can access and use the data, or restricted, in which case only authorised users are allowed access to private data.

According to Verhulst et al. (2019), one of the models of data collaboratives that emerge on the spectrum of these two variables are trusted intermediaries. They define **trusted intermediaries** as third-party actors that "*support collaboration between private-sector data providers and data users from the public sector, civil society, or academia*" (ibid.,

p. 11). In relation to these two variables, they are characterised by independent usage of the contributed private sector data and restricted access to said data.

Under this overarching definition, Verhulst et al. (2019) also define two more specific types of trusted intermediaries: **data brokers and third-party data analytics projects**. Data brokers are third party actors who facilitate the purpose and time-bound connections around the sharing of data, collaborating closely with both the demand and the user side while limiting the direct interactions users have with data doners. This definition differs significantly from the previously discussed definition of data brokers proposed by Rieke et al. (2016), as it does not focus on the aspect of revenue nor the strict requirement of external sourcing of the data.

Third-party analytics projects, while also being third-party actors connecting data doners and users, focus more heavily on data analytics and share derived insights but not the underlying data with users (Verhulst et al., 2019).

Table 2: Concepts and Definitions of Data Intermediaries

Concept	Definition
Electronic Intermediary	Actors who aggregate information, match suppliers and customers and create trust in markets where the exchange of goods and services is partially or fully automated by information technology (Bailey and Bakos, 1997).
Information Intermediary	Any system that mediates between the producers and consumers of information through the collection, organisation and distribution of information (Womack, 2002).
Infomediary	Actors who match demand and supply through the aggregation of desired information and provision of institutional infrastructure, that help build trust between the supplier and the user (Sein & Furuholt, 2012). An agent helps users manage the vast amount of information and safeguard their privacy (Hagel and Rayport, 1997). Actors who provide products or services to citizens or other third parties based on public sector data (Magalhaes, Roseira & Strover, 2013).
Open Data Intermediary	An entity positioned between two actors in a data supply chain, which incorporates an open data set the use of which might otherwise have not been facilitated had it not been for the involvement of the intermediary (Van Schalkwyk et al., 2016).
Trusted Intermediary Corporate Data Philanthropy	Companies sharing their data with a limited number of partners through a third actor (Verhulst & Sangokoya, 2014).
Multi-Sided Platform	Public-private platforms, positioned between two or more entities that create value by enabling direct communication between different customers (Klievink, Bharosa & Tan, 2016).

Trusted Intermediary	Third-party actors supporting collaboration between private-sector data providers and data users from the public sector, civil society, or academia (Verhulst et al., 2019).
Data Broker	A company or business unit that earns its primary revenue by supplying data or inferences about people gathered mainly from sources other than the data subjects themselves (Rieke et al., 2016). Companies that specialise in the collection and exchange of personal information, usually associated with "big data" operations, that can be sourced directly from the consumer (Crain, 2018). Third parties which facilitate connections and match the supply of data holders with the demand for it in a purpose and time-bound manner (Verhulst et al., 2019).
Third-party Analytics Project	Actors which access private-sector data, conduct targeted analysis, and share insights, but not the underlying data, with public or civil sector partners (Verhulst et al., 2019)

2.1.2 Working Definition of Trusted Data Intermediary

As shown in Table 2, there are many concurring and overlapping conceptualisations of intermediation in data sharing. Therefore, it is important for the working definition of the trusted data intermediary to include the most widely accepted core elements of the concept, but also to put particular focus on the elements that are the most relevant for this particular research.

The core elements that were the most commonly present across different definitions of data intermediaries were the following: (i) intermediaries are middlemen located between at least two actors (Van Schalkwyk et al., 2016; Klievink, Bharosa & Tan, 2016; Womack, 2002; Verhulst et al., 2019), (ii) their functions include some form of data collection, structuring or analysis (Bailey and Bakos, 1997; Womack, 2002; Sein & Furuholt, 2012; Verhulst et al., 2019) and (iii) they aim to offer some type of added, most commonly social or public, value (Klievink, Bharosa & Tan, 2016; George, Yan & Leidner, 2020; Verhulst et al., 2019). In addition, the element of trust also emerges as a common theme with several definitions making direct or supplementary reference to a data intermediaries' function of building and establishing trust among different actors on the market (Bailey and Bakos, 1997; Sein & Furuholt, 2012; Van Schalkwyk, Willmers & McNaughton, 2016).

A notion that was not as widely referenced across literature but presents itself as a relevant element of the definition of data intermediaries in the context of this work is the cross-sectoral aspect of the collaboration and data exchange facilitated by data

intermediaries. We see this theme emerge most dominantly in the definition of a trusted intermediary as a form of a data collaborative put forth by Susha, Janssen & Verhulst (2017a), Verhulst et al. (2019) and in the definition of cross-sectoral partnerships by Selsky & Parker (2005), which emphasises the role of data intermediaries in linking actors from public, private and academic sectors.

The element of a data intermediary having to produce economic value or revenue was also presented as part of several definitions (Rieke et al., 2016; Susha, Grönlund & Van Tulder, 2019), however, it did not emerge as a dominant enough element to warrant being included in the working definition, especially taking into account that in practice data intermediaries have been observed to exist as non-profit organisations (Van Schalkwyk et al., 2016). The same can be said about the inclusion of the element of mandatory use of open datasets within the definition of a data intermediary (Magalhaes, Roseira & Strover, 2013; Van Schalkwyk et al., 2016), as this is a specific characteristic of intermediaries in the open data field, which are outside of the scope of the given work focused on data sharing between public and private sectors.

Moreover, the working definition of trusted data intermediaries should aim to account for as many forms of data intermediaries as possible. For this reason, the sub-class specification of intermediaries into data brokers and third-party analytics projects proposed by Verhulst et al. (2019) will be disregarded. This classification, though useful for understanding the emerging practices of data sharing on the market, would be unnecessarily limiting for the purposes of this paper as, in practice, intermediaries can combine elements characteristic of both data brokers and analytics projects in their work.

Taking all of the above mentioned into consideration, the following will be the working definition of trusted data intermediaries used throughout this thesis: ***Trusted data intermediaries are agents that enable cross-sectoral (and particularly private-public) data sharing, trust building and collaboration, most often through some form of data collection and processing, aimed at the generation of added value for partners and redressal of public needs.***

2.2 Linking Measures of Data Intermediaries

Not unlike the concept of data intermediaries themselves, the issue of measures undertaken by data intermediaries to effectively link private and public users has no clear conceptual definition in literature. According to Van Schalkwyk et al. (2016), **linking measures refer to the process by which "*data intermediaries connect actors and facilitate the flow of data*" (p. 3)**. One could find various aspects of this process being discussed in the literature on issues of data privacy (Eckartz, Hofman & Veenstra, 2014),

data sharing risk (Martens & Duch-Brown, 2020), the value proposition of data sharing (Klievink, Van der Voort & Veeneman, 2018) or, in select cases, more comprehensively in both the technical and social aspects of data collaboration (Susha, Janssen & Verhulst, 2017b; Van Schalkwyk et al., 2016; Flipsen, 2019; Susha, 2020). This section will discuss the linking measures around the dominant themes that emerge across literature.

2.2.1 The Intermediary Model

Susha, Janssen & Verhulst (2017b) present the intermediary model in itself as a distinct linking measure. They observed that coordination issues in a data collaborative can become more complex when there is a multiplicity of actors in both data providers and data users (ibid). They identified the transfer of knowledge between parties through the use of a third-party intermediary as a mechanism employed by data collaboratives for addressing the problem of matching potential data providers and data users. According to Susha, Janssen & Verhulst (2017b), an intermediary can positively mitigate complexity and match user needs by coordinating what contributors are willing to provide with the users' needs and intentions but did not further examine the details of this process.

Similarly, Susha, Janssen & Verhulst (2017a) also saw the intermediary model as one of the "*mechanisms to match the supply and demand in data collaboratives*" (p. 2695), which they refer to as facilitation. According to them, facilitation in data exchange can either be self-facilitated by having direct contact between the contributor and the user of the data or intermediary-facilitated. In the latter case, the intermediary can either serve data-related functions, whereby it provides pre-gathered data or technological support for data sharing or organisational functions, where the intermediary plays a coordinating role securing the participation of both sides (ibid).

Martens & Duch-Brown (2020) also present the establishment of intermediaries as one of the top ways of managing the challenges of business to government data transferring. According to them, data sharing comes with considerable "*ex-ante costs related to finding a suitable data sharing partner, negotiating a contractual arrangement, re-formatting and cleaning the data*" (p.16). Technological and institutional arrangements enabled by introducing trusted intermediaries can lower such transaction costs and facilitate more voluntary transactions among a broader range of suppliers and users (ibid).

Klievink, Van der Voort & Veeneman (2018) likewise note the role of a trusted intermediary in establishing the ground rules, offering initial trust and presenting the advantages of collaboration as crucial in establishing data exchange.

Van Schalkwyk et al. (2016) built upon the role of intermediation in data linking practices by emphasising the aspect of collaboration among data intermediaries. Their research found that the flow of data among actors is often facilitated not by a single intermediary agent but by a chain of intermediaries with complementary capabilities and resources (ibid). Such connected functioning of multiple intermediaries increases the probability that maximum value is derived from the exchange of data between the provider and the user (ibid). The practical examples they present draw an association with the different functions of data intermediaries as put forward by Susha, Janssen & Verhulst (2017a), whereby intermediaries with data-related functions and intermediaries with organisational functions can cooperate to ensure the best outcomes.

2.2.2 Social Connections and Networks

Several authors underline the importance of network connections for the facilitation of data collaboration. Existing networks and reputation can be one of the chief ways in which connections between data providers and users can be formed (Susha, Janssen & Verhulst, 2017b). Moreover, public relations and the broadening of social networks can also be one of the principal incentives for data sharing that intermediaries' can offer to their partners (Susha & Gil-Garcia, 2019).

For data intermediaries, networks can consist of data contributors, data users and other entities such as governments, businesses, social institutions, experts or even citizens (Flipsen, 2019). These network actors can fulfil either enabling or supporting roles for the data intermediary (ibid).

A well-developed network, and the associated higher level of social capital, can help data intermediaries be more attune to the needs of their targeted segments and increase the likelihood of the end-use of data (Van Schalkwyk et al., 2016). In some instances, social networks can additionally help intermediaries gain access to new user groups (ibid). Van Schalkwyk et al. (2016) even stipulate, that social connections and capital are the chief ways intermediaries facilitate transactions of data between those who have it and those who need it, with all other methods being supplementary. This seems to reflect the position of practitioners, who report that successful data sharing projects are customarily built on personal relationships and credibility (Susha, 2020).

This importance of social connections as a linking practice is also reiterated by Klievink, Van der Voort & Veeneman (2018). In their work, they outlined several key factors relevant to the understanding of how data collaboratives can effectively carry out an exchange of information, with one of the foremost among them being the existence of previous relationships or collaborations, which was seen as a strongly positive influence

on the creation of trust. They found that in cases where the disclosure of information was a part of continued collaboration with a data intermediary, donors were more open and willing to collaborate, as they already had channels of communication in place (ibid). Klievink, Van der Voort & Veeneman (2018) see this relation as a cycle by which collaboration and trust continuously reinforced each other and strengthened the collaborative's capacity to create valuable exchange over time. Similar conclusions were reached in the research conducted by Susha (2020), which found that "*prior collaboration, reputation, credibility, community support, personal relationships*" were all crucial in the creation of a network effect and the success of data collaboratives (p.15).

The importance of continued and prior collaboration was comparably emphasised by Robin, Klein & Jütting (2016), who stated that since obtaining data sharing agreements from private partners is a lengthy process, investing in existing relationships and building up upon collaborations is of vital importance. Such long-term relationships demonstrate a firm value when data is being exchanged to direct emergency response efforts, as it shortens the time necessary to link relevant actors and exchange data, which in such cases can be of curtail importance (Winowatan, Zahuranec, Young & Verhulst, 2020). Existing collaborations can also facilitate further linking through endorsements by previous high-profile partners or broader industry support (Susha, 2020). This has been demonstrated in practice through instances where support in new partner organisations was garnered due to the visible engagement of high-level leadership of other organisations with the data-sharing initiative (Winowatan, Young, & Verhulst, 2019).

2.2.3 Technical Resources and Expertise

According to a study of data intermediaries conducted by Van Schalkwyk et al. (2016), technical capabilities of collecting, analysing and organising data were the linking measures most predominantly used by intermediaries in the process of facilitating data exchange. Similar findings were produced by the research of Flipsen (2019), who found that technical resources and expertise were some of the most used capabilities by trusted intermediaries in their value proposition to their users. These technical resources took different forms, among them analytical data features, storage systems, platforms and software that simplified the use and reuse of data, but crucially every studied data intermediary utilised some form of technical feature as part of their value proposition (ibid). Both these studies indicate an emphasis on capabilities of storing, processing and analysing data as central to the notion of data intermediaries that were previously reflected as elements in the definition of data intermediaries (Sein & Furuholt, 2012; Verhulst et al., 2019), hinting at the importance of technical capabilities as a linking measure.

Access to research and insights from data science expertise seems to be one of the leading incentives for cooperation put forward to partners by data collaboratives (Susha & Gil-Garcia, 2019). In particular, data intermediaries can offer additional value to their users through aggregated analysis of different data sets from various sources, which can be greater than would have been derived from analysing any of the data sets separately (Martens & Duch-Brown, 2020). Although this type of aggregated data analysis can in design be targeted towards addressing a social issue, private firms nonetheless stand to acquire new capacity and knowledge through seeing and understanding their contributed data in a way that would not have been possible on its own (Robin, Klein & Jütting, 2016).

Technical interventions can also help address concerns about data privacy. In such case, the linking action can take the form of "*anonymisation by filtering or aggregation by combining data sets into a single record*" (p.8) or of setting up technical access control mechanisms with identity management (Eckartz, Hofman & Veenstra, 2014). In the latter case, technical infrastructure capabilities ensure the privacy and security of data in terms of its protection from unauthorised or malicious access (Groves & Neufeldi, 2017).

There are also many technical interventions available that increase the value and utility of data on the user end, such as data structuring, publication of data under well-established data standards or p through existing vocabularies and ontologies (Eckartz, Hofman & Veenstra, 2014).

Moreover, technical measures present one of the chief ways in which intermediaries can address potential insufficiency of data quality through the extension of the dataset to improve its completeness, addition of metadata and context or data visualisation (ibid). This is an important aspect since, as reported by practitioners, data quality is a factor that significantly influences the success of a data collaborative (Susha, 2020).

Most importantly, technical capabilities allow intermediaries to produce problem-specific and actionable data insight, which allow partners to reflect on the effectiveness of the collaboration (Susha & Gil-Garcia, 2019). If the collaboration does not produce impacts in line with the targeted ones, this may deter existing partners and new ones from pursuing further cooperation (ibid). Thus, the technical ability to derive conclusions from data and ensure that it is exchanged in a way that allows for the fulfilment of set out goals is a vital linking measure for data intermediaries (ibid).

2.2.4 Governance Mechanisms

As an essential factor in inter-organisational collaboration and information sharing, governance can significantly affect the success of data collaboratives (Susha & Gil-

Garcia, 2019). Governance in a data-sharing initiative can be defined as the set of coordinating and monitoring activities that help clarify roles and responsibilities within the collaboration, create trust and improve coordination (ibid).

A common issue that needs to be addressed through adopting appropriate governance measures is that of trust. By partially giving up control over their data assets, data contributors open themselves up to the possibility of their information being used to develop new products or services that might be in opposition to their interests (Klievink, Van der Voort & Veeneman, 2018). This perceived vulnerability is amplified if the purposes and circle of potential data users are ambiguous (Verhulst et al., 2019).

Mitigation of the fear of data use for unnegotiated purposes through contractual agreements or reliance on reputation is a valuable means of facilitating data exchange available to data collaboratives (Klievink, Van der Voort & Veeneman, 2018). Susha & Gil-Garcia (2019) even go as far as to say that identifying and mitigating such risks is a central point in designing data collaboratives. In practice, this means that the type and level of control over the data offered to contributing partners is often a deciding factor for effective collaboration (Klievink, Van der Voort & Veeneman, 2018).

The mitigation of these risks can take the form of allowing suppliers to negotiate specific conditions of access that reflect the cost and risk characteristics of the particular data in question (Martens & Duch-Brown, 2020). In addition, allowing licensing mechanisms to reflect the policies of the data owners can be a valuable means of building trust (Eckartz, Hofman & Veenstra, 2014). This concurs with the conclusion made by Susha & Gil-Garcia (2019) that initiatives such as data collaboratives "*might require collaboration mechanisms that are more flexible and dynamic and adapt to specific projects, even when the partners are the same*" (p. 2894). According to Martens & Duch-Brown (2020), instances where such opportunities for individual negotiation were not accounted for and intermediaries imposed standard contracts and equal terms for all parties have led to the failure of the initiatives to scale up.

However, provisional templates for data sharing also present a valuable linking tool for data intermediaries, as they provide partners inexperienced in data sharing with a regulatory system that guarantees them legal certainty while simultaneously simplifying their research and organisational efforts (Groves & Neufeldi, 2017). Moreover, although flexible contracts are a facilitating tool on the side of the supplier, more stringent contracts on the consumer side that prescribing detailed rules for data access, usage, licensing and privacy procedures have also been demonstrated to be an enabler for collaborations (Winowatan, Young & Verhulst, 2020a). Such practices can be accompanied by the additional governance measure of conducting pre-contractual screenings through

requiring persons wishing to access private datasets to submit proposals detailing the context, goal, and timeline for the analysis they wish to conduct (ibid).

According to Klievink, Van der Voort & Veeneman (2018), the perceived vulnerabilities of data suppliers rises if the object of exchange is real-time, unprocessed, raw data. To mitigate this particular risk, trusted intermediaries can take it upon themselves to process the data in-house and transfer only the derived insights or services to the end-user (Martens & Duch-Brown, 2020). Indeed, this approach suggested by Martens & Duch-Brown (2020) seems to be already implemented in practice, as showcased in the data collaborative taxonomy of Verhulst et al. (2019), which distinguished third-party analytics projects as a type of trusted data intermediary that shares insights, but not the underlying data, with public or civil sector partners.

An alternative or additional governance mechanism assuring partners of data privacy employed is the adoption of internationally recognised privacy guidelines developed by reputable actors that are already familiar and elicit trust in their potential partners (Winowatan et al., 2020; Winowatan, Young & Verhulst, 2020a).

2.2.5 Economic Incentives and Financial Resources

Trusted data intermediaries need to develop mechanisms for limiting the gap between private costs and public benefits of data sharing (Martens & Duch-Brown, 2020). As suggested by Eckartz, Hofman & Veenstra (2014), some measures that can account for the economic constraints of private-public data exchange include: "*i) sharing the costs of opening up data, ii) defining a pricing structure for data set usage and iii) setting up a separate organisation to govern and maintain the data and its usage*" (p.10). The later measure potentially envisions intermediaries as appropriate mechanisms that can utilise the preceding first two economic measures to effectively link private and public actors.

The availability of financial resources is also recognised as a crucial linking measure for data collaboratives by Susha & Gil-Garcia (2019). Although many data collaboratives operate on a non-profit basis, they also bear significant costs for acquiring the resources in the form of both data and data expertise (ibid). This makes the question of funding of central importance in a data intermediary's ability to operate and bring actors together (Susha & Gil-Garcia, 2019). According to Flipsen (2019), means of gathering such financial resources may vary from charging customers for products and services to receiving funding from a parent organisation or network partners. Moreover, trusted intermediaries might receive income from a combination of various sources, such as both through partnership sponsoring and customer charges (ibid). In such cases, having an inflow of income from customer charges does not exclude the possibility that the

intermediary operates on a non-profit basis (ibid). Thus, having possession of a diversified income stream to boost financial resources to re-invest in future collaborations can be a valuable linking measure at the disposal of trusted data intermediaries.

Furthermore, advantage can stem not only from having a good source of income flow but also from knowledge of best forms of income generation. As demonstrated by a data collaborative case study by Winowatan, Young & Verhulst (2019), particular financial structuring of data partnerships, such as joint-partner funding, can serve to minimise competitive concerns among actors and assist the development of trust.

2.2.6 Governmental Action

Throughout literature, governmental support is also recognised as a potentially valuable facilitator of connection between data partners, capable of providing a safe environment by setting standard rules or incentives (Klievink, Van der Voort & Veeneman, 2018). According to Susha et al. (2019), the governmental measures for cross-sectoral data exchange can vary in levels of intensity and degree of involvement. They can take the form of active participation, when a government is directly involved in the process of negotiating mutually beneficial agreements, to more passive involvement, when the government serves a supplementary supportive role and stimulates data sharing through relevant policies and incentives (ibid). Governments can often take such measures as part of long running projects aimed at developing business-to-government information exchange (Klievink, Janssen & Tan, 2012). The degree of governmental facilitation can depend on the particularities of the data-sharing initiatives (Susha et al., 2019).

According to Janssen & Zuiderwijk (2014), governments can also play an important enabling and supporting role in helping intermediaries connect data to users through funding, organising stimulating activities such as hackathons or connecting different actors in the data ecosystem. Moreover, governments can serve a supporting function of providing intermediaries access to government-owned information to complement and contextualise their analysis of private data and lead to more valuable conclusions, which indeed transpires in practice (Winowatan, Young & Verhulst, 2020a, 2020b).

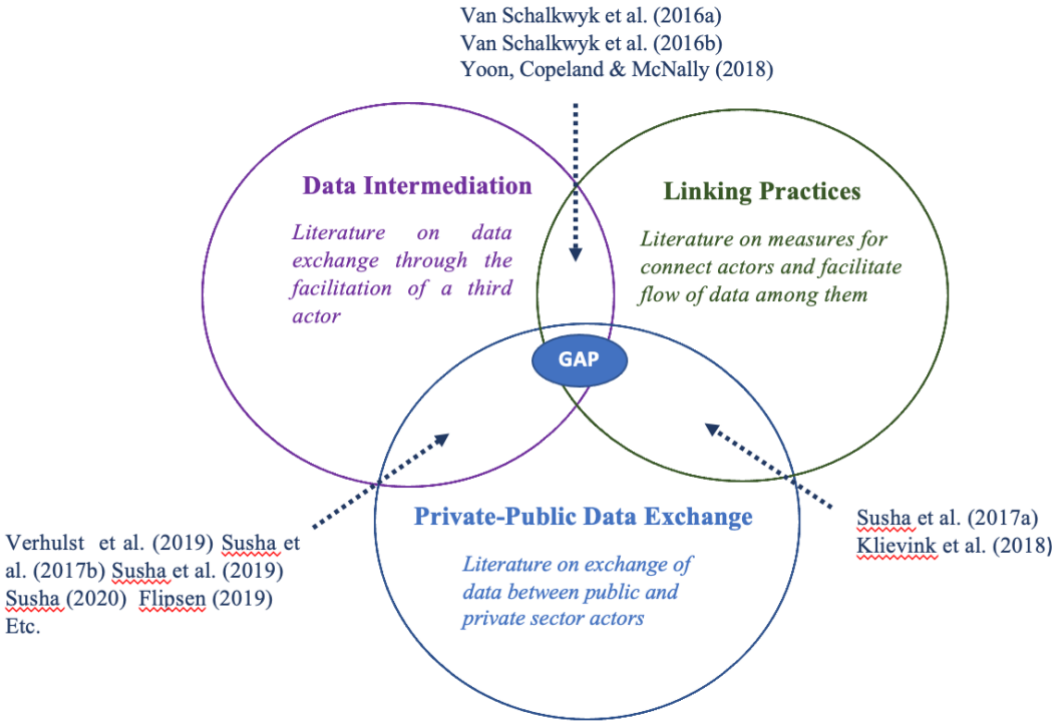
2.3 Gap in Literature

It can be concluded that the literature on the linking practices employed by data intermediaries connecting actors across sectors lacks systematisation in two chief ways. Firstly, the terminological ambiguity and variety cause the relevant research to be spread out across different segments of literature. Secondly, the understanding of linking practices in itself has rarely been approached holistically, with literature instead focusing

on different aspects of the operation of data intermediaries, with some of them falling within the bounds of linking practices.

There is an evident lack of systematic study on data intermediaries in the field of private-public data exchange. This lack of literature is further exasperated by the factor that a lot of older literature examining public-private partnerships and inter-organisational collaboration cannot account for the recent developments in the data ecosystem and the proliferation of data and Information Communication Technology (ICT), making them less relevant in the modern-day context (Susha, 2020). The most well-developed segment of research around this topic was found in the literature on data collaboratives. However, even in this domain, literature was wanting. Part of this can be attributed to the fact that data collaboratives as a concept are relatively new to the field, due to which there is a lack of connected empirical work (Klievink, Van der Voort & Veeneman, 2018; Susha, Grönlund & Van Tulder, 2019).

Figure 2: Gap in Current Literature



In particular, there is a gap in the literature of work covering the intersection of data intermediation, private-public data exchange and linking practices. Works that address private-public data exchange and linking practices (Susha, Janssen, & Verhulst 2017b; Klievink, Van der Voort & Veeneman, 2018) take a more general approach overviewing a wide range of collaboration forms and lack focus on the specificities of the intermediation model. On the other hand, works that focus on the linking practices of data

intermediaries tend to centre on open or governmental data (Van Schalkwyk et al., 2016; Van Schalkwyk, Willmers & McNaughton, 2016; Yoon, Copeland & McNally, 2018) and thus lack relevant insight on the particularities of facilitating private-public data exchange. In other cases, observations on intermediaries are hypothetical or based on literature reviews and not results of empirical research (Eckartz, Hofman & Veenstra, 2014) or, as mentioned earlier, fail to account for the current context of the data revolution (Susha, 2020). The most complete, and the only work to deal with all three elements of this intersection, is Flipsen (2019), which examines the business models of data intermediaries facilitating private-public data exchange and touches upon the subject of linking measures through discussion on the value propositions of data intermediaries.

3 Theoretical Framework

This chapter presents the theoretical framework used as the basis for this work, which is an adapted model of Bourdieu's theory of capital. According to this model developed by Pierre Bourdieu, the social world is composed of a multidimensional structure in which actors are positioned and interact with each other on the basis of three relational parameters: field, habitus and capital (Van Schalkwyk et al., 2016). This theory proposed by Bourdieu represents a distinct overarching framework, which has been proven to be applicable across diverse topics with considerable success (Atkinson, 2019) among them in the areas of organisational analysis (Emirbayer & Johnson, 2008) and digital communication technologies (Ignatow & Robinson, 2017).

Section 3.1 of this chapter will overview Bourdieu's concepts of capital and the related notions of field and habitus. Section 3.2. will discuss Bourdieu's concepts within the organisational context. Lastly, section 3.3. will discuss the relation of Bourdieu's concepts to intermediary organisations and data intermediaries in particular.

3.1 Field, Habitus and Capital

Bourdieu and Wacquant (1992) define **field** as *"a network, or a configuration of objective relations between positions"* (p.97). In other words, a field represents *"a series of institutions, rules, rituals, conventions, categories, designation, appointments and titles which … produce and authorise certain discourse and activities* (Webb, Schirato & Danaher, 2002, p.21). Put simply, a field can be seen as a collection of formal and informal norms that govern a particular social sphere of activity (Edgerton & Roberts, 2014). These series of norms, in turn, condition and constrain the behaviour of individuals within it and shape their motivations (Ignatow & Robinson, 2017).

Crucial in Bourdieu's understanding of the concept of field is the rejection of absolute or unifying space (Hilgers & Mangez, 2015). Instead, there are various fields within the social world, with all fields being relational and characterised by specificities of their constituents and the given activity of the field (ibid.) Thus fields are autonomous domains with specific customs and characteristic activities that *"respond to rules of functioning and institutions that are specific to it and which define the relations among the agents"* (ibid, p. 5).

Moreover, a field is also a hierarchical structure of dominant and subdominant positions (Albert & Kleinman, 2011). The relative standing of actors within the field is determined by their possession of capital (Atkinson, 2019). However, the position of actors is not static, they are constantly contending for higher recognition, which designates that fields

are almost always dynamic, another of its essential characteristics (ibid). As such, a field can also be seen as a battlefield of agents competing for a position within it, waged through the use of capital available to the agents (Van Schalkwyk et al., 2016). The actors can compete to preserve or increase the value of the capital they possess or redefine the rules that govern the field (Bourdieu & Wacquant, 1992).

As such, Bourdieu's notion of field can be perceived in two chief ways, firstly, as a specialised environment that constrains and shapes the existence, behaviour and interaction of the actors within it and secondly, as a conflict zone among these actors for positions within the field.

In opposition to the field's broad scope, **habitus** is a characteristic of a specific entity operating within the field and represents a link between the social structure and an individual's agency (Barrett & Martina, 2012). Bourdieu conceptualises habitus as a stable apparatus of learned attitudes, perceptions and behaviours towards one's long-term possibilities, acquired through social and close-network interactions (ibid). The formation of habitus is also closely related to the resource of knowledge that an entity has access to, often influenced by economics and social networks (Huang, 2019).

The concept of habitus can be perceived as a theoretical bridge between the notions of field and capital and represents the mechanism by which agents endowed with certain types of capital adopt strategies to propel and position themselves within the field (Bourdieu & Wacquant, 1992).

The notion of **capital** represents the set of resources and powers usable by entities to collect profits and execute their developmental opportunities (Ragnedda, 2018). The most overarching and unifying definition presents capital as:

"a collection of goods and skills, of knowledge and acknowledgment, belonging to an individual or to a group that he or she can mobilise to develop influence, gain power, or bargain other elements of this collection" (Neveu, 2018, p. 347).

According to Harker, Mahar & Wilkes (1990) for Bourdieu, capital: *"acts as a social relation within a system of exchange, and the term is extended 'to all the goods, material and symbolic, without distinction, that present themselves as rare and worthy of being sought after in a particular social formation" (p.13).* It is a durable commodity capable of taking many forms that requires time and effort to acquire but, in turn, carries the potential of producing profit and benefit for the agent (Bourdieu, 1986).

Thus, we can see that Bourdieu's definition of capital is quite broad and includes both material items and intangible assets, both of which are equally capable of possessing

symbolic value that makes them thought after possessions (Harker, Mahar & Wilkes, 1990). However, this wide definition can be broken down into two key elements: firstly, that capital must have some type of value and secondly, that capital must be accumulable and convertible into other forms of capital.

Capital can be valuable to an entity either due to its volume or its type and uniqueness (Atkinson, 2019). However, when assessing the significance of any capital, both material and immaterial, it has to be kept in mind that *"capital does not have intrinsic value in terms of being advantageous but only makes sense in relation to a field in which it is employed"* (Van Schalkwyk, 2016, p.10). Cumulativeness of capital means that firstly, it takes time to accumulate it and secondly, that once it has been accumulated, it tends to persist in its being (Zang, 2010). As mentioned previously, the amount of capital accumulated by an entity determines its position and amount of power within a field (Webb, Schirato & Danaher, 2002). For these purposes of field positioning, capital can be exchanged for other types of capital, converted to have higher symbolic value and serve as a basis for domination in a field (Harker, Mahar & Wilkes, 1990).

Bourdieu categorises capital into three main forms: economic capital, social capital, cultural capital and an additional form - symbolic capital. However, recent literature has distinguished **other supplementary types of capital**, such as technical capital (Van Schalkwyk, 2016). Bourdieu himself accounted for the probability that in addition to the primary forms of capital, each specific field can generate its own unique forms of capital dictated by the field's internal logic (Bourdieu, 1986; Albert & Kleinman, 2011).

Economic capital is the collection of the monetary and material resources that an entity possesses (Albert & Kleinman, 2011). These resources are constructed by the general elements of production such as land, factories, jobs and inheritance (Huang, 2019), as well as immaterial objects of direct monetary value such as intellectual property rights (Pret, Shaw & Drakopoulou Dodd, 2016).

Thus, in the most straightforward definition, economic capital refers to the ownership of money and property (Atkinson, 2019). However, it may encompass anything that can be *"immediately and directly convertible into money and may be institutionalised in the form of property rights"* (Bourdieu, 1986 p.16).

An important characteristic of economic capital is that of all the forms of capital, it is the one that most easily allows for access to goods and services, whilst other forms of capital generally require some level of conversion in this process (Bourdieu, 1986). Moreover, it can be converted at a higher speed to both other types of economic capital

as well as to other forms of capital, such as cultural capital (Neveu, 2018). This attribute has led Bourdieu (1986) to describe economic capital as the capital at the root of all other types of capital. However, it should be noted that this simplicity of conversion is only comparative, with economic capital also requiring time and effort for a changeover in the most outcome maximising ways, which in some instances, such as during the conversion of economic capital into social capital, can even be an arduous and long-term process (Neveu, 2018).

Social capital is the capital that stems from membership in a particular group (Bourdieu, 1986). It is the "*aggregate of the actual or potential resources which are linked to possession of a durable network of more or less institutionalised relationships of mutual acquaintance and recognition*" (ibid, p.21). This definition incorporates several key characteristics of social capital. Firstly, by "potential resources", Bourdieu refers to the benefits one receives through membership in a particular group, which acts as a collectively-owned capital that the group members preserve through the feelings of trust and obligation (Prell, 2006). Secondly, the notion of "institutionalised relationships" alludes to these social relationships' long-term and stable nature, whereby social capital is continuously reinforced and maintained through ongoing complex sets of interactions (ibid).

Put in a different way, social capital is capital of connections with agents that possess high degrees of economic or cultural capital that can be drawn on for personal benefit or serve as a proxy for an entity's affiliation with specific titles or prestigious groups (Atkinson, 2019). Bourdieu (1986) refers to this notion as a "credential" bestowed onto an entity by the group, which the entity can use to collect additional capital.

More recently, the notion of networking has been incorporated into the concept of social capital (Ferri, Deakins & Whittam, 2009). It has emerged as a means to emphasise the networking approach to social capital, referring to the fact that the trust and reciprocity characteristic to social capital emerge and are maintained through social networks (Prell, 2006), which are defined as the system of formal and informal relationships accumulated by entities through interactions in the workplace, local associations or other meeting places (Harper, 2002).

According to Bourdieu (1986), the volume of social capital possessed by an entity is determined firstly by the size of its network of connections which he is capable of effectively mobilising to his support and secondly, by the volume of capital possessed in their own right by each of such members of his network. Mobilisation in this context means a network member's willingness or sense of obligation to offer support upon

request and the resources that they subsequently make available (Ferri, Deakins & Whittam, 2009).

The profits that can be derived through such mobilisation comprise a wide range of "*services accrued from useful relationships, and symbolic profits, such as those derived from association with a rare, prestigious group*" (Bourdieu, 1986, p.22). In certain conditions, the collection of social obligation can even be converted into economic capital by granting institutionalised title (Bourdieu, 1986) or more directly through access to information or exclusive funding (Ferri, Deakins & Whittam, 2009). However, this relation between social and economic capital is not always stable or simple to convert (Huang, 2019).

Examples of what specific and practical forms social capital can take in practice (see Table 3) can include: access to individuals who can positively influence an entity's activities, such as business advisors, venture capitalists, customers and suppliers (Ferri, Deakins & Whittam, 2009); access to privileged knowledge and information through group membership (ibid); access to the right connections that allows for the facilitation of fruitful collaborations fitting with the entity's goals (Albert & Kleinman, 2011); the existence of a trust-based relationship with other entities, in which cooperation can take place without formalised or less formalised agreements (Prell, 2006).

Cultural capital is a type of capital that is "*made of knowledge and know-how, of the skills and analytical tools that allow one to manage and produce social relations, cultural products, and technical devices*" (Neveu, 2018, p.351). In a slightly different definition by Webb, Schirato & Danaher (2002), cultural capital is interpreted as "*a form of value associated with culturally authorised tastes, consumption patterns, attributes, skills and awards"* (p. X), which brings in an additional element of public "authorisation" of value.

At the centre of both definitions, and underlining its importance as a source of institutionalised cultural capital, is academic qualification or association with academic circles and practices (Bourdieu, 1986). More broadly, it is the knowledge that lies at the heart of cultural capital (Webb, Schirato & Danaher, 2002). The value of acquiring such knowledge lies in the fact that it enables an entity to express and communicate concepts and abstract ideas in a way that is easily understandable to other actors within the same system (Atkinson, 2019). Therefore, the essence of cultural capital can be summarised as the "*mastery of specific systems of symbols and signs which have come to be valued in the social order*" (ibid, p.107).

Habits and skills which facilitate and simplify the continued acquisition of education and knowledge are also essential parts of cultural capital (Prieur, & Savage, 2011). Thus,

the focus of cultural capital is not only the end product in the form of the acquired knowledge or qualification, with the means of consuming cultural goods in themselves being embodied forms of cultural capital (Emmison & Frow, 1998).

An additional important aspect of cultural capital is the notion of evaluation, often linked to the knowledge field through entrance competitions and certifications (Bourdieu, 1986; Prieur & Savage, 2011). It can take the form of recognised certificates or licenses, giving cultural capital a more easily comparable and exchangeable disposition (Huang, 2019).

Symbolic capital is the transfiguration of these three forms of basic, and other additional types, of capital and represents the recognition or prestige that an agent enjoys by possessing these capitals (Neveu, 2018). This type of capital thus comes to exist "*through the double process of accumulation of another capital and the recognition of the extraordinary meaning of this accumulation by a public possessing the interpretive tools and the socialisation necessary for making sense of the situation*." (ibid, p.362). Through the second half of that definition, we can see that symbolic capital is closely tied and dependent upon the social space within which it exists. Expression of symbolic capital, such as prestige and reputation, do not carry value in and of themselves, but depend on the societal belief that a particular agent possesses these qualities (Webb, Schirato & Danaher, 2002). Therefore, once the value of any type of particular capital is recognised within the field, it can become a source of symbolic capital (Pret, Shaw & Drakopoulou Dodd, 2016). However, despite this perceptive quality of symbolic capital, it can still be objectified to a certain degree in the form of awards and formal recognitions (ibid).

The value of symbolic capital lies in the fact that it can generate legitimacy and trust for those who possess it (ibid). However, symbolic capital goes beyond simple attributes of visibility and prestige. Through the legitimacy that is gained through possession of symbolic capital, an agent gains a "stand out" position on the field, which equips them with the power to dictate new rules, organisational structures and definitions of new phenomena which they can impose on other field participants (De Clercq & Voronov 2009).

According to Zang (2010), **technical capital** is the structural relationship between technologies and other actors, which carries the potential of reproduction or innovation through technological means. It is made up of several elements: a) the technologies that are owned, can be used by or are designed for a particular entity (however, this should not be equated with having economic accessibility to certain technological products) b) authorisation for the use of particular technological artefacts in ways that is not available

to other actors and c) the characteristics of usage of technology that entities impose upon other (Romele, 2020). Finally, it is the accumulation of digital competencies and technologies, which cover both an entity's knowledge and skill in regards to digital information, communication, safety, problem-solving and access to external and tangible digital resources (Ragnedda, 2018).

This type of capital can be accumulated by establishing and maintaining relations with technologies (Zang, 2010). The more technical capital an entity possesses, the more it will be recognised as an authority and be able to more freely operate within a technologically mediated social world which will, at the same time, become increasingly tailored to the agent (Romele, 2020).

Overall, although the notion of different types of capitals is the theoretical aspect of Bourdieu's work most central to this thesis, for better comprehension of the concepts of capital as well as its contextualisation, it is essential to discuss the related concepts of field and habitus, as the three are closely linked and form a unified conceptual model. The direct connection between field, habitus and each form of capital is expressed in that that the value given to capital is influenced by the social and cultural characteristics of the habitus and field (Harker, Mahar & Wilkes, 1990), while the nature and availability of capital affect and influence the structure of the field and the habitus of its actors (Van Schalkwyk, 2016).

Table 3: Exemplars of Different Types of Capital

Economic Capital	• Possession of income, savings, intellectual property and tangible business assets (Pret, Shaw & Drakopoulou Dodd, 2016); • Possession of assets capable of being immediately or directly converted into money or of being institutionalised in the form of property rights (Bourdieu, 1986); • Possession of the general elements of production and could be things such as land, factories, jobs and inheritance (Huang, 2019); • Possession of assets usable for the hiring of human resources and expertise (Albert & Kleinman, 2011); • Availability of means to pay organisations for access to the desired data (Van Schalkwyk et al., 2016).
Cultural Capital	• Holding valued dispositions, cultural goods, skills and education (Pret, Shaw & Drakopoulou Dodd, 2016); • Having extensive knowledge and mastery of information, concepts, trends, theories and history in the relevant field (Atkinson, 2019); • Holding high academic qualifications or affiliations (Bourdieu, 1986); • Holding leading positions in competitions (Bourdieu, 1986); • Possessing habits or skills which facilitate and simplify the continued acquisition of education and knowledge (Prieur & Savage, 2011); • Holding of certifications (Neveu, 2018).

Social Capital	• Having memberships in societies, relations, networks and alliances (Pret, Shaw & Drakopoulou Dodd, 2016). • Having an extensive network of connections which one is capable of mobilising for their assistance and benefit (Bourdieu, 1986). • Having a network of connections made up of agents who in turn themselves possess high levels of capital (Bourdieu, 1986). • Having interconnections with other companies (Atkinson, 2019). • Having the right connections that allow for the facilitation of collaboration with reputable experts (Albert & Kleinman, 2011). • Having the image of trustworthiness, resulting in other agents being more open to cooperation, less formalised cooperation and reciprocity in resource sharing (Prell, 2006). • Possessing privileged access to knowledge resources by virtue of network connections (Ferri, Deakins & Whittam, 2009). • Having network mediated access to individuals capable of playing an influential role in the agent's success, for example, business advisors, venture capitalists, customers or suppliers (Ferri, Deakins & Whittam, 2009). • Having access to third-party organisations that provide support and funding through the virtue of the network (Ferri, Deakins & Whittam, 2009) • Being well informed about relevant local, organisational, civil, club activities and possessing an ability to influence them through the virtue of being part of a network (Harper, 2002). • Having the capability to speed up the time frame for certain processes/activities by calling on personal connections (Huang, 2019). • Membership in professional unions or guilds (Pret, Shaw & Drakopoulou Dodd, 2016).
Symbolic Capital	• Holding of awards, trophies, publicity, reputation, prestige (Pret, Shaw & Drakopoulou Dodd, 2016); • Being the subject of advertisements and publicity in reputable sources (Pret, Shaw & Drakopoulou Dodd, 2016); • Having membership of an esteemed group or organisation (Atkinson, 2019); • Having a reputation for being a dominant actor in a particular field (Atkinson, 2019); • Possessing a status of professionalism (professional knowledge, skill, experience and ethics) (Noordegraaf & Schinkel, 2011); • Having a reputation for conducting high-quality research work (Van Schalkwyk et al., 2016); • Having the ability to impose definitions of phenomena on other field participants, establish new field rules and dictate best new practices (De Clercq & Voronov, 2009); • Having a stand out position in comparison to other actors within a field (De Clercq & Voronov, 2009); • Having a reputation of being actively involved in charity and philanthropy (Pret, Shaw & Drakopoulou Dodd, 2016); • Having a long history of industry experience (Pret, Shaw & Drakopoulou Dodd, 2016); • Possessing customer or client loyalty despite the availability of alternatives on the market (Pret, Shaw & Drakopoulou Dodd, 2016).

Technical Capital	• Possession of technologies, particularly those designed for or those that can be used for an entity's particular processes and objectives (Romele, 2020). • Having authorisation for the use of particular technological artefacts in ways that is not available to other actors (Romele, 2020). • Possessing knowledge and skill in regards to digital information, communication, safety and problem-solving (Ragnedda, 2018). • Possessing the abilities of data extraction, reorganisation, anonymisation, validation, treatment, formatting and processing (Van Schalkwyk et al., 2016). • Having the knowledge and skill to facilitate the data publication in useful formats (Van Schalkwyk et al., 2016). • Having the knowledge and skill to develop applications for interpreting large datasets (Van Schalkwyk et al., 2016).

3.2 Bourdieu's Concepts in an Organisational Context

The Bourdieuian concept of capital, and its roles and effects, have been widely applied and studied within not only the individual but also the organisational context. Although formal organisations were rarely the primary focus of his sociology, Bourdieu himself analysed the field of firms and their ability to possess unique types of capital and habitus (Atkinson, 2019; Hallett & Gougherty, 2018).

The organisational field refers to organisations that operate within a recognised area of institutional life, e.g., car manufacturing, book publishing (Emirbayer & Johnson, 2008). Notably, the field includes all entities that play some type of role in the activity under question (ibid). Similar to the individual context, the organisational field is also characterised by hierarchy (Khanchel & Kahla, 2013). Organisations in a more dominant position will aim to cement their position through maintained stability and resulting ability to influence the rules of the field for new entries (ibid).

Organisations also possess a form of habitus. Similar to how their lived experiences shape individuals' behaviours and habitus, organisational behaviour and action are also influenced by organisational experience (Dobbin, 2008). This is exemplified by the fact that management within an organisation often does not consider all available options for addressing given issues and instead guide themselves on analogue experiences from their past and the resulting organisational habits (ibid). This organisational consciousness can be seen as organisational habitus (ibid). It has also been described as the collection of the institutional beliefs, rules and roles that, when brought together, determine an organisation's perception and evaluation of the organisational field (Vaughan, 2008).

The concept of capital, in particular, has enjoyed widespread use in organisational theory (Emirbayer & Johnson, 2008). Historically, the most connoted form of capital in

an organisation has been economic capital, however, recently other forms have also begun getting more attention in the organisational arena (ibid).

One such has been social capital, which research has demonstrated can contribute in many ways to creating new value for an organisation (Tsai, 2000). In their work, Ferri, Deakins & Whittam (2009) stipulated that social capital in the form of network relations plays an important role in the process of new venture creation and development by providing young organisations with both tangible and intangible resources in the forms of finances, knowledge, information and expert access. Such benefits of social capital can have both short and long-term effects on company growth (ibid). In particular, high levels of social and symbolic capital can serve to make an entity's entrance onto a field more assertive and equip them with the ability to gain legitimacy and, ultimately, other necessary resources (De Clercq & Voronov, 2009). In some instances, the benefits gained through such social capital can be significant enough to compensate for the lack of monetary resources (Ferri, Deakins & Whittam, 2009). A similar intrinsic link between social capital and new venture development was observed by Davidsson and Honig (2003) in the form of the recognition and exploitation of opportunities enabled through external relations in a social structure.

When discussing organisational social capital, it is relevant to consider to what extent such capital can be seen as an aggregation of the social capital of individual members of the organisation (Waldstrøm, 2003). This cross-over is certainly possible, as exemplified through instances when interpersonal relations between individuals lead to the sharing of valuable information or organisational collaborations (ibid). It can thus be stipulated that, as long as the goals of an individual are in tune with that of the organisation, it is likely that an increase in an individual's social capital will increase the social capital of their organisation (ibid).

Similar to social capital, symbolic capital has also been identified as possessing a significant effect on a company's long-term performance (Pret, Shaw & Drakopoulou Dodd, 2016). An important aspect of organisational symbolic capital is the ability that it grants organisations to determine and shift industry perceptions to their benefit, e.g., through dictating what technological arrangements are considered the most innovative and best practices in the field (De Clercq & Voronov, 2009). Such perception setting is most often based on their own technological capabilities and innovations (ibid).

Cultural capital also carries influence for organisations as it presents a means of obtaining legitimacy, which many organisations seek as much as monetary resources (Dobbin, 2008). In addition to its standard components, in an organisational context, the cultural capital can also comprise an entity's capability for knowledge production,

knowledge dissemination or other aspects of its scientific practice (Albert & Kleinman, 2011). Institutionalised cultural capital can also serve organisations in framing their ideological position, which could help them connect and form partnerships with organisations with similar stances and aims (Greenspan, 2014).

In turn, an organisation's economic capital can be directly affected by the amount of the other capitals it possesses, in the form of access to funding opportunities, client bases or even political patronage necessary for the effective execution of its activities (ibid).

3.3 Bourdieu's Concepts, Intermediation and Data Intermediaries

Compared to the sphere of general organisational analysis, theory on Bourdieu's concepts in relation to concretely intermediary type organisations is relatively scarce. However, some works do link the notion of capital with the building of relationships and governing mechanisms in an inter-organisational context (Tsai, 2000).

For example, social capital has been found as being of high value for agents specialising in the facilitation of linkages between other structural units (ibid). In such instances, social capital expressed in previous network relations and the resulting trust was observed to be positively correlated to the number of new linkages formed by the agent (ibid). Specifically, entities with a higher degree of social capital were more likely to generate linkages quicker and in a more effort efficient manner (ibid).

According to Burt (2001), social capital in an organisational context can be created through a network in which agents broker connections between otherwise disconnected entities. This approach stems from the fact that real markets are imperfect and possess structural holes, whereby specific agents are unaware of the benefits they could offer each other (Burt, 1999). Often this is because agents operate within different flows of information (ibid). Holding a connection forging role in the market can even be seen as a source of social capital in and of itself (Burt, 2001). In this understanding social capital presents itself as "*a function of brokerage opportunities*" (ibid p.4).

The use of Bourdieu's theories of social interactions, and particularly that of capital, in the sphere of data intermediation is further supported by the findings of Van Schalkwyk et al. (2016). Their work examined the practices employed by open data intermediaries in developing countries for connecting data providers and end-users through the lens of the different types of capital.

Using Bourdieu's concepts as a narrative model for data intermediaries, Van Schalkwyk et al. (2016) postulated the following: the environment within which data systems and their transmission take place within countries can be considered a relatively autonomous

field, within which each particular transaction involves at least two agents. The first agent, possessing a particular habitus and capital, might have a comparative advantage due to the possession of a valuable asset in the form of data, the second agent, also possessing a particular habitus and capital, might have a comparatively weaker position due to its lack of material or symbolic resources. The relation between the two agents becomes possible through the involvement of a third agent, the intermediary, who similarly possesses its specific habitus and capital. Through the close alignment with both sides of the transaction, the intermediary agent can facilitate a situation in which the habitus of the first two agents meet, and transfer or conversion of capital among them can effectively take place.

Some of the particular expressions of different types of capital that Van Schalkwyk et al. (2016) identified in data intermediaries (see Table 3) include: data extraction, treatment, formatting and interpretation as forms of technical capital, long-standing relationship with key personnel as a form of social capital and reputation for the production of high-quality research as a form of symbolic capital. Overall, they pointed to the value of different types of capital in matching data supply and use and referenced a discovered overreliance of data intermediaries on technical capital as a potential limitation of their linking efforts (ibid).

Taking the Bourdieuian narrative concept of data intermediation developed by Van Schalkwyk et al. (2016) and applying it to private-public data intermediation that is the subject of this work, the following can be theorised:

The environment of data utilisation for the public good can be considered a comparatively autonomous field, comprised of private organisations, public bodies and intermediary organisations. Although these organisations are different in type, their grouping into one field is justified per the principle that a field includes not only one particular type of organisation but also all entities that play some role in the activity in question (Emirbayer & Johnson, 2008). In the given case, private organisations play a role in the activity through the provision of valuable data, intermediaries through facilitation of its transfer to public sector actors, and public entities through its transformation into a public good.

Additionally, all actors within this field possess their own unique habitus. As organisational habitus is determined by past organisational experiences, internal rules and perceptions (Dobbin, 2008; Vaughan, 2008), the fact that actors within this field possess different backgrounds in terms of their sector and organisational form dictates their diverging habitus. These differences in their respective habitus put them at either a

comparative advantage or disadvantage to each other and affect the process of negotiating data exchange.

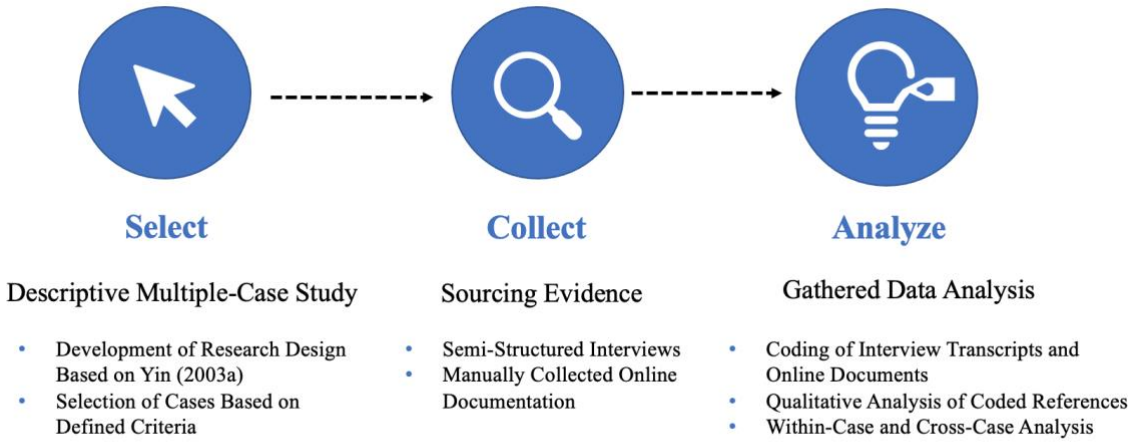
Moreover, as capital is inherently spread unequally among actors within a field (Bourdieu, 1986), each entity will possess its particular volume and type of capital. The difference in the amount of capital possessed by agents, particularly social and cultural capital, may cause them to operate within different information flows, leading to them being unaware of each other and the mutual benefits of data sharing between them (Burt, 1999). In such instances, the interaction between agents can be facilitated through the involvement of a data intermediary agent. The intermediary, through the use of the capital available to it, aligns itself to both the private entities and public entities at either side of the transaction, creating an environment in which the two agents can have a fruitful collaboration and exchange of data (Van Schalkwyk et al., 2016).

Following this theorisation, and based on previous research, it can be concluded that the theory of types of capital can be a valuable framework for understanding specific characteristics of data intermediaries, particularly those connected to their linking practices, unveiling previously indiscernible traits and predicting intermediaries' behaviour and impacts on data ecosystems (Van Schalkwyk et al., 2016).

4 Methodology

A multiple-case study method was chosen to answer the research questions of this thesis presented in Table 1. This chapter gives a detailed account of the design and execution of the selected method. Section 4.1 explains the case study method and justifies its selection for the given research purposes. Section 4.2 explains how data was collected from two sources of evidence, while section 4.3 elaborates on the method for making inferences from the collected data. An overview of the applied research design is reflected in Figure 3.

Figure 3: Overview of the Research Design



4.1 Research Design - Multiple-Case Study

The choice of the adoption of a certain type of research methodology is dependent on the type of the proposed research question (Yin, 2003a). The research questions put forth in Table 1 approach the linking practices of trusted data intermediaries through "what" questions. According to Yin (2003a), "what" questions favour research methods such as archival analysis and surveys. However, as trusted data intermediaries are a comparatively new research area, meaning that the availability of data for archival analysis as well as the number of potential survey respondents is limited (Flipsen, 2019). This recentness and actuality of trusted data intermediaries make them a contemporary event, in which case, according to Yin (2003a), the case study is a preferred method of examination.

A case study is "*an empirical inquiry that investigates a contemporary phenomenon within its real-life context*" (Yin, 2003a, p.13). It is a research method that allows the gaining of a holistic understanding of the characteristics of particular organisational and

managerial processes (ibid) and the complex interactions among organisations, technologies and people (Dube & Pare, 2003). The linking practices employed by trusted data intermediaries, into which this work aims to gain insight, represents a particular example of current organisational processes characterised by interaction complexity, once again indicating case study as an appropriate method of inquiry.

There can be several categories of case studies, among them descriptive. A descriptive case study presents an account of a phenomenon within its real-life context (Yin, 2003b). It makes careful observations of the what, where and when of the phenomenon of interest for the research and documents it in detail (Bhattacherjee, 2012). In addition, it can be used to illustrate aspects of the phenomenon thought to be representative of the typical within the field of inquiry (Schell, 1992).

Moreover, all types of case study research can comprise of either single and multiple case studies (Yin, 2003a). According to Yin (2003a), multiple-case studies are generally preferable to single-case studies, as they allow for the possibility of replication and external generalisability of findings. It is a suitable research design when studying real-life events that show numerous sources of evidence (Zaidah, 2007).

This work will take on the method of a descriptive multiple-case study. This approach is appropriate for this work in light of the research questions it puts forward. The answer to the main research question and sub-question 1, formulated as ‘what’ questions, would entail a detailed relation of the linking practices employed by trusted data intermediaries, making it suitable for a descriptive study. Moreover, since multi-case studies are generally preferable to single-case studies, and given that in the particular instance, there are numerous sources of evidence in the form of different trusted data intermediaries operating on the market, for the validity and generalisability of the outcomes, the choice of multiple-case design is also justified.

To overcome potential limitations of the case study method and enhance the reliability of any findings, this work will be conducted following the proposed case-study research design approach and recommendations of Yin (2003a), which is one of the most widely cited and leading methodologies in case study research (Flipsen, 2019; Yazan, 2015). This design approach includes providing an overview of the case selection, detailing data collection procedures, presenting interview formats and questions and formatting the resulting information, all of which will be discussed at length in the following sections of this chapter.

4.2 Case Selection

According to Yin (2003a), case selection in a multiple-case study design should follow the logic of replication. This means that each case included in the multiple-case study must be carefully selected to predict either the same results or contrasting results for predictable reasons. Both types of replication logic first require establishing a foundation that will detail the reasons for why the phenomenon under question is likely to be found to be varying or consistent across cases (ibid).

This case selection for the given work will follow the logic of replication, meaning that the results of the case studies, that is, the linking practices employed by trusted data intermediaries, are expected to be comparably similar. The basis for the adoption of this logic is the presumption that due to the shared area of activity, characteristics and goals of trusted data intermediaries, they will employ similar practices in linking private and public partners in the process of data sharing.

Additionally, to allow for the generalisability of case study findings and enable cross-case analysis, the selected cases should be diverse (Eisenhardt, 1989; Flipsen, 2019). In the given instance, diversity in trusted data intermediaries could be expressed in terms of the geographic area of their activity, policy areas of activity and profit-generating models. However, considering the specificities of the theoretical framework presented in Chapter 3, specifically the notion of field, the selected cases should be limited to one broader geographical area. Due to practical issues of accessibility, Europe was chosen as the geographic area of case selection. Moreover, other practical considerations regarding the availability of information on the cases should be considered (Flipsen, 2019). Lastly, the selected cases should sufficiently cover the scope of the theoretical framework used to analyse the findings (ibid).

Based on these considerations, four case selection criteria, shown in Table 4, were identified. The case selection criteria reflect the best practices for case selection prescribed in literature as discussed above, whilst considering practical considerations, the specificities stemming from the given research questions and theoretical framework.

Table 4: Case Selection Criteria

Case Selection Criteria
1. Cases need to meet the definition of trusted data intermediary set out in Section 2.4.
2. Case information must be available and accessible.
3. The selected cases should be diverse (geographical area of activity within Europe, policy area of activity).

4. The available case information must sufficiently cover the theoretical framework elements set out in Chapter 3.

A preliminary list of potential cases was first compiled through the use of the Data Collaborative Explorer (Data Collaborative Explorer), an extensive online repository of existing data collaboratives. Firstly, only the data collaboratives registered as operating in Europe were selected. Secondly, out of those, only the data collaboratives that were classified as trusted data intermediaries were chosen for further examination. This list was supplemented by collaboratives found through additional manual desk research. Lastly, the websites of potential cases, particularly the description of case organisations activities, were studied to ensure that their activities possessed the elements of the working definition of trusted data intermediary outlined in Section 2.1.2. As the following step, to ensure the availability of sufficient information, only the trusted data intermediaries whose representatives agreed to participate in interviews were selected for the final case study list.

Via this process and based on the selection criteria presented in Table 4, the following 5 cases were selected: *Consumer Data Research Centre, Dalberg Data Insights, DjustConnect, Urban Big Data Centre and Positium.* A brief description of each case can be found in Table 5.

Table 5: Selected Cases and their Brief Descriptions

Case	Case Description
Consumer Data Research Centre (CDRC)	The Consumer Data Research Centre (CDRC) is a multi-institution laboratory that collects, synthesizes, analyses and provides access to consumer-related datasets from around the United Kingdom, led by the University of Leeds and University College London. They closely partner with members from academia, industry and the public sector.
Dalberg Data Insights (DDI)	Dalberg Data Insights (DDI) is part of the Dalberg Global Development Advisors strategy firm, which supports developmental organisations and local governments to analyse data held by private companies using various analytics tools.
DjustConnect	DjustConnect is a Belgian data-sharing platform focusing on the agri-food sector, based at the Flanders Research Institute for Agriculture, Fisheries and Food.
Positium	Positium is an Estonian based company that provides innovative data solutions and analytics using anonymised mobile positioning data from mobile network operators to help decision-makers in governments take informed measures.
Urban Big Data Centre (UBDC)	Urban Big Data Centre (UBDC) is a UK based research centre and data service affiliated with the University of Glasgow, offering expertise, training, data collections, and data tools to promote innovative research methods and the use of big data for the improvement of well-being in cities.

4.3 Data Collection

To answer the posed research questions, reliable data on the selected cases must be gathered. Data for case studies can come from different sources, the primary of them being: documentation, direct observation, participant observations, interviews and physical artefacts (Yin, 2003a). It is typical for case studies to combine several complementary data collection methods (Dube & Pare, 2003). The use of multiple sources of evidence is encouraged as good practice by Yin (2003a) as a means of developing a more holistic view of the studied issue and triangulating data for increased credibility of results.

On these bases, this study employed two complementary data collection methods: interviews and online documentation analysis.

4.3.1 Interviews

The primary source of evidence for this research comes from interviews. According to Yin (2003a), interviews are an essential source of information for case studies and enable targeted and insightful observations focused directly on the selected issues.

Therefore, interview questions and processes should fulfil two requirements: they must follow a prescribed and structured line of inquiry and, at the same time, follow a conversational and flexible manner to better serve the line of inquiry (Yin, 2003a). This can be achieved through the use of the semi-structured interview method. The semi-structured interview method consists of conversationally conferring with one person at a time, employing a mix of open-ended and closed-ended questions and regularly resorting to follow up-questions (Adams, 2015). Semi-structured interviews are particularly advantageous in cases where the line of inquiry deals with an area with a lack of previous research (ibid), as in the case of trusted data intermediaries.

As interview questions can carry bias or be leading (Yin, 2003a), it is crucial to consider and prepare the questions in advance. Moreover, a semi-fixed set of questions is also vital for ensuring consistency of gathered data across cases. Interviews should start with basic questions about the background information of the interviewee (Jacob & Furgerson, 2012). The following substantive questions should be based on the existing body of knowledge around the issues and stem directly from the research question (ibid). The content of the questions should also relate to the elements of the theoretical framework in use for the research (Flipsen, 2019).

It is customary for the semi-structured interview format to have a mix of open-ended and closed-ended questions (Adams, 2015). In particular, close-ended questions can serve

as gateways to more open-ended and in-depth probing (ibid). According to Yin (2003a), it is also appropriate to ask respondents to propose their own insights on certain occurrences, which can be used as a basis for follow up inquiry. Wider reaching and expansive questions are also encouraged as means of gathering meaningful insight and generating a free flow of conversation (Jacob & Furgerson, 2012). However, it is important to be flexible and open to on-the-spot revision of the questioning guide per the flow of the conversation (Adams, 2015; Jacob & Furgerson, 2012).

Bearing these considerations in mind, an interview question guide was developed (**see Appendix C**). Pertaining to the guidance provided in literature, it comprised both open and closed-ended questions and covers background, technical, and substantive issues related to the line of inquiry. The questions were subdivided into four dimensions: introductory questions, which focused on background information; circle of partners, which contained questions on the organisations' list of both public and private partners; collaboration forming, which addressed the practices and measures employed by intermediaries in their processes; and concluding questions, which asked for any additional input and contacts for potential additional interviews.

This set of questions served as a guide during interviews, and the majority of them were addressed in each interview, but following the nature of semi-structured interviews, they were on occasion deviated from depending on the flow of the conversation, and follow-up questions were improvised.

Potential interview subjects were identified through the public information available on the case study organisations online websites. The professional position or job descriptions of the contacted persons included high-level management, project or operations managers, stakeholder relations managers and public relations managers, as these were deemed to be the positions most closely involved in the processes related to the research questions. Invitations for an interview were extended to the corresponding persons via email. In the initial invitation, they were informed of the research background and objectives. Upon their request, they were also provided with the interview questions in advance of the interview in a follow-up communication. Following a confirmation, 30-45 minute long interviews were held with each interviewee in an online video call format.

With the explicit permission of the interviewees, the audio of the interviews was recorded to avoid any recollection inaccuracies. The audio recordings of the interviews were deleted immediately after their transcription. The transcripts of the interviews can be found in Appendix sections F-M.

To increase the reliability of findings, it was striven to obtain more than one interview for each case study organisation identified in Table 5. In the end, a total of eight interviews were conducted. With the exception of UBDC, two respondents were interviewed from each studied case. An overview of the interviews, respondents' positions within the intermediaries and the Appendix section with the corresponding transcript can be seen in Table 6.

Table 6: Overview of Conducted Interviews

Organisation	Professional Role of Interviewee	Appendix Section:
Consumer Data Research Centre (CDRC)	Project Delivery Manager	Appendix F
	High Level Management	Appendix G
Dalberg Data Insights (DDI)	High Level Management	Appendix H
	Senior Project Manager	Appendix I
DjustConnect	Current Project Manager Previous Project Manager	Appendix J
Positium	High Level Management	Appendix K
	Sales Director	Appendix L
Urban Big Data Centre (UBDC)	Senior Data Scientist	Appendix M

4.3.2 Online Documentation

A secondary source of evidence for the case study constitutes online documentation. According to Yin (2003a), documents are beneficial for corroborating and augmenting evidence from other sources, making it a well-suited secondary source to complement the primary information gathered through interviews.

The online documentation used as sources of data consists of: web contents from the case websites, files available for download at said case websites and online articles or reports mentioning the cases. The documentation search and retrieval were conducted manually, primarily through reference to the case websites, and secondly through online search engines utilising the case's names as search words. Only documentation found on sources deemed reputable was used as evidence sources.

To address the possible deterring effect of the dynamic nature of web content and its susceptibility to change, the online documents were downloaded as PDF files, assigned with the source URL and marked with a timestamp of date of access. They were then

given reference codes that are used as means of reference in this work. An overview of all studied online documents containing the document URL, time of access and reference code can be seen in **Appendix D**.

4.4 Data Analysis

The stage of data analysis involves examining, categorising and recombining gathered evidence to address the line of inquiry of the study (Yin, 2003a). In analysing the gathered evidence, this study will take a qualitative approach, motivated by the aim of producing analytical rather than statistical generalisation.

Firstly, the data gathered through both the interviews and the online documentation **was analysed through the use of coding**. Coding is one of the predominant ways of analysing qualitative data (Saldaña, 2013). It refers to the use of short words or phrases for the assigning of representative and summative value to a larger portion of data (ibid). It can be applied to many forms of data, among them interview transcripts, documents and internet websites (ibid). Additionally, it is the method of choice for analysis of data that aims to find repetitive patterns of action within gathered data (ibid). These characteristics make coding a suitable choice of data analysis method for the given research.

Saldaña (2013) recommends to code data at multiple stages - preliminary (provisional) coding before the start of the analysis and additional coding during the analysis process. A preliminary code list should be developed in coordination with the study's conceptual framework and research goals (ibid). Moreover, in addition to main codes further subcodes can be developed for more detailed content analysis (ibid). The use of subcodes is particularly recommended for studies with multiple participants or cases (ibid).

In line with this guidance, the preliminary coding list for the analysis of the interview transcripts and online documents was derived from the theoretical framework presented in chapter 3. Namely, the initial codes were formed from the primary forms of capital as described in Appendix B, with the five primary codes being: Economic Capital, Social Capital, Cultural Capital, Symbolic Capital and Technological Capital. Moreover, based on the conceptualisations related to these notions of capital developed in literature, also summarised in Appendix B, preliminary subcodes were developed. In the process of primary analysis additional subcodes were established and added to the code list. This updated code list then served as a basis for the second round of data analysis. The final list of codes and their definitions refined following the several rounds of reiteration and their respective definitions can be found in **Appendix E.**

Following the coding, for the validation of the case study findings from multiple sources of evidence, **converging lines of inquiry were developed through the process of triangulation** (Yin, 2003a). In the given work, triangulation was only conducted through combining two different sources of data, namely the interviews and online documentation, while the method and theory of analysis remained the same for both sources of data. Thus, this work adopted the approach of data triangulation (Denzin, 1978).

In the process of combining the data, the practical guidance on strategies of data triangulation proposed by Leuffen, Shikano & Walter (2013) was followed, namely that of a weighted average approach. According to this strategy, the content from each source is averaged and weighted by the quality of its information (Leuffen, Shikano & Walter, 2013; Kern, 2018). In the context of the given study, this means that although information from both sources was averaged in the process of analysing the linking practices of data intermediaries, the data stemming from interviews was given a higher weight, as it was deemed more detailed and up-to-date and thus of higher quality than data from online documentation.

Lastly, both within-case (Chapter 5) and cross-case analysis (Chapter 6) were conducted. A particular characteristic of a multiple-case study research design is that it permits both within-case and cross-case analysis (Eisenhardt, 1989). A within-case analysis approaches a case as a stand-alone entity and identifies patterns unique to the particular case (ibid). It also helps develop an initial familiarity with the data that at a later stage, helps facilitate cross-case analysis (ibid). In turn, cross-case analysis, or as referred to by Yin (2003a) cross-case synthesis, is an aggregation of finding across the range of individual studies. The cross-case analysis approach allows for looking at data from multiple perspectives, thus increasing the reliability of any conclusions and the likelihood of producing novel findings (Eisenhardt, 1989).

On the basis of the coding, each case presented in Table 5 was first studied individually. This encompassed analysing which of the gathered data was connected and relevant to the case's linking practices and what types of capital these activities constituted. The analysis at this stage also aided in the creation of a general understanding of the theoretical concepts within their given context of data intermediation.

The findings from individual cases were then compared across all studied intermediates. In particular, findings were compared per the expressions of different types of capital in different cases. The found similarities and differences of these expressions of types of capital are described in Chapter 6.

5 Findings

This chapter will demonstrate the findings gathered on the selected data intermediaries. The findings on each case will be presented in the following structure: background information on the case organisation, its main areas of activities, the linking measures employed by it and lastly, a table summarising the linking measures and their classification into forms of capital as presented in Chapter 3. All findings are supported with the reference code of the corresponding data document as presented in Appendix D or with the Appendix section of the relevant interview as provided in Table 6.

5.1 Consumer Data Research Centre

The Consumer Data Research Centre (CDRC) is a data intermediary organisation bringing together research expertise from four United Kingdom universities: University of Leeds, University College London, University of Liverpool and the University of Oxford (CDRC-1). It was established in 2014 with funding from the Economic and Social Research Council, a non-departmental public body financed by the United Kingdom Government as part of the Big Data Network Initiative (ibid). The Big Data Network Initiative was a project aimed at making data routinely collected by businesses and local government organisations accessible for research and innovative development purposes (CDRC-2). Within the bounds of this project, several research organisations focusing on different topical areas were established, with the CDRC being one of them (ibid).

The focus of the CDRC lies in the consumer data generated in the United Kingdom that holds the potential of delivering valuable insights into human behaviour (CDRC-1). Its objective is to promote the use of consumer data for society benefiting social science research and the maximisation of the innovative potential of consumer-related organisations (CDRC-3). In this aim, they work with public sector partners to provide them with data insights and context on societal issues and industry partners to improve their understanding of consumer behaviour and strategic decision making (CDRC-1).

The CDRC carries out four core activities: data acquisition, data service, research and training and capacity building (CDRC-3).

Data acquisition involves both proactively and reactively procuring data from a range of commercial organisations (ibid). The CDRC works with over 30 data contributors from both commercial and public sectors, which regularly grant them access to their datasets (CDRC-1; Appendix F). These datasets can be found in the CDRC's purpose-built data portal, which provides an easily searchable catalogue of the data they hold based on format, topic and availability on national, regional or local levels (CDRC-5). It currently

hosts 69 data products with 32 open access, 25 safeguarded, and 12 controlled datasets (ibid, June 20, 2021 statistics). Notably, the datasets are continuously updated and expanded as new data becomes available (Appendix F).

The CDRC **data service** comprises two chief activities: value-adding data processing and data storage and protection systems (CDRC-3). Some of the main tools for enhancing data value utilised by the CDRC include data cleaning, addition of spatial references and other attributes, clarification of attribute descriptions and modification of data into more readily usable formats (CDRC-5). The extreme granularity at the level of geographical location obtained through these processes make the datasets particularly valuable for research at local levels and municipal authorities (Appendix G).

In terms of data security, the **CDRC classifies data they hold into three categories of access:** open, safeguarded and controlled (CDRC-3). Open data is freely available to all users for any purpose and is composed of non-sensitive data and derivative products produced by the CDRC (CDRC-4). Safeguarded data has restricted access due to licence conditions but is not considered personally identifiable or otherwise sensitive (ibid). Access to this level of data requires submitting a project proposal requiring approval by the CDRC before access to the data is granted (CDRC-6). Approved users are then able to access the data through a remote service (ibid). Lastly, controlled data is personally identifiable and is therefore subject to data protection legislation or is considered commercially sensitive (CDRC-4). Similar to safeguarded data, its access also requires pre-approval of the research proposal. However, as an extra security measure, controlled data can only be accessed on-sight at one of the CRDC's three secure facilities (ibid).

Project proposal requesting access to more secure levels of data must demonstrate the project potential to provide insights beneficial to society or scientific knowledge and proof of consideration for privacy and ethics issues (ibid). Moreover, users wishing to access controlled data must undergo safe research training by a licenced provider and present proof of accreditation (ibid). According to one interviewee, two-thirds of applications requesting such access to secure levels of data come from the academic sector, with the remaining applications coming in from the public sector and third-sector organisations (Appendix F).

The **research dimension** of the CDRC's work also consists of several directions. First, the CDRC conducts multiple internal research projects around its driving and priority themes (CDRC-3). Retailing and consumer consumption was the CDRC's original research focus, but areas of internal research have since then been expanded to also include health and well-being, urban mobility, population movement and housing infrastructure (Appendix G, CDRC-8). Secondly, in addition to internal research, the

CDRC also engages in collaborative research projects with commercial and other research partners (CDRC-3).

The CDRC's **training and capacity building** scheme encompasses several different projects. On a smaller scale, they provide annual training programs and online tutorials for both academics and practitioners aimed at giving participants relevant skills for data analysis (CDRC-7). These materials range from introductory to more advanced and specialised courses (ibid). The CDRC also operates a **Masters Dissertation Scheme**, which links Master's students with commercial partners to execute research projects important to the retail industry (CDRC-7). The scheme has been running as long as the CDRC has existed (Appendix G). It is presented to businesses as an opportunity to maximise the value of their privately-held customer data and, through the help of a Masters student, make progress on major current issues or long-desired projects (CDRC-6). Similarly, the CDRC also offers a Doctoral training program (CDRC-7).

All four areas of activity incorporate elements of linking measures that the CDRC employs in the overarching process of facilitating a flow of private data to public needs. One such is the principle of the "**Ladder of Engagement**", which was emphasised as a vital linking measure by both of the interviewees from the CDRC (Appendix F, Appendix G). The ladder of engagement is "*an approach to develop relationships with organisations, to start them and then increase them over time*" (Appendix F). It stemmed from the realisation that commercial companies often had data analysis needs that they did not possess enough resources to meet (Appendix G). By providing them with these resources through the Master's dissertation scheme, the CDRC shifted the conversation from a request for data to an offer of support in a way that demonstrated to private organisations the value in developing longer-term relationships with them (ibid). What this means in practice is that the CDRC initiates their relationships with private organisations through their Master's Dissertation Scheme (Appendix F). If that process goes well and the private company feels that they have benefited from the delivered insights, the CDRC will offer them a new partnership with a PhD student, which presents a more long-term commitment. This gradual process enables the CDRC to more organically approach the possibility of the organisation letting them host some of its data (ibid). According to the second respondent, the ladder of engagement is also a valuable tool for trust-building with members of the commercial sector (Appendix G). Both respondents emphasised that the initial connections formed through this approach made the future conversations around possible data-sharing agreements much more likely to be fruitful (Appendix F, Appendix G).

User forums also emerge as a linking measure for the acquiring of private data (CDRC-3). Data partner forums are an annual event that brings together CDRC's existing and potential data partners and academics to discuss research opportunities (Appendix G). These events present attendees with successful case studies from the CDRC's recent practice to bring attention to the value of consumer data analysis (ibid). Similar events are held for representatives of academia, encouraging them to engage with the datasets available through the CDRC (ibid). Notably, the interviewer emphasised the utility of such case study examples as a way of communicating the benefits of their work (ibid).

The CDRC also annually holds a conference around their Masters Dissertation Scheme, demonstrating the projects that have been completed by the students in partnership with commercial partners within that year (Appendix F). Similar to the user forums, this event aims to showcase the value of partnership with the CDRC to potential new partners as well as develop an alumni network and talent in the area of consumer data analysis (Appendix G). One interviewee noted that as organisations often hire the students that partnered with them through the Master Dissertation Scheme, this further facilitates data sharing by virtue of access to a contact inside the organisation (ibid).

Overall, the CDRC is quite active in terms of networking, holding three or four major national events annually and participating as guest speakers in many more events (ibid). Both interviewees **identified networking events, professional connections and word-of-mouth exchange among data contributors as the leading methods through which data sharing relationships are established** (Appendix F, Appendix G).

Notably, one of the interviewees emphasised that such networking events also present a considerable value proposition to the data contributors (Appendix G). As the networking events hosted by the CDRC bring together individuals interested in data-driven development in concrete focus sectors, these events also double as opportunities for private sector representatives to network with industry colleagues and pursue their agendas (ibid). If an organisation ceases to cooperate with the CDRC, it will also be deprived of access to these opportunities, which could be a significant loss (ibid). Additionally, through membership in the CDRC's network and participation in the annual events, **members get early access to a wide body of research** produced directly by the CDRC or completed through the various agreements signed by them (ibid).

In relation to data sharing relationships with the public sector, one of the interviewees observed that **connecting with highly regarded data specialists** within public departments or governmental agencies often simplifies communication and aids partnership development (ibid). In the CDRC's experience, this often presents a more effective linking measure than directly approaching high-level public management.

Financial resources also play into CDRC's data acquisition process. As an underlying ethos and part of their financial agreements, the CDRC does not pay commercial rates for data provided to them by industry contributors (Appendix F). However, they do offer a certain degree of financial compensation for some datasets, mainly to the extent of covering the data providers processing costs (ibid). Thus, the CDRC does possess a data budget, but it is generally spent quite conservatively (ibid). However, in some instances, the CDRC might obtain additional funding to support a private company in setting up a new research project, from which the CDRC can also benefit (ibid).

Furthermore, the CDRC's financial resources allow it to, on occasion, finance stakeholder attendance at its conferences, who otherwise would not be able to participate in the event (ibid). Most often, this issue arises concerning governmental employees, who operate on stringent budgets (ibid). More recently, with the onset of the Covid-19 pandemic, financial resources became most relevant for developing pre-recorded educational material for delivering CDRC's training programs in an online format (ibid).

According to both interviewees, the CDRC's creation and its focus on the retail sector are closely related to the government's strong concern over issues of data underuse and the future of the retail sector at the time (Appendix F; Appendix G). Both interviewers also stated that strong political and social resonance of research issues has, in general, been facilitating to the process of collaboration forming by making the value of the data more demonstrable and allowing data contributors to associate themselves philanthropic and socially beneficial activities (ibid). In light of this enabling effect, one of the interviewees identified **keeping up-to-date on the developments in political discourse and anticipating key topics at the top of political and governmental agendas** in the near future as an operational practice for negotiation of data sharing agreements and acquisition of financial recourses in the form of grants (Appendix G).

In terms of research, one of the interviewees stated that the offer of a **greater understanding of the data** they hold, through knowledge of the methods and outcomes of the CDRC's conducted research on it, is a strong motivator for sharing of datasets for both private and public stakeholders (ibid).

The CDRC's research activities also aim to support the public sector data users challenged by the lack of resources and expertise limiting their ability to take on the data hosted by the CDRC (ibid). To address this, especially in light of the new joint project around issues of Covid-19, the CDRC has been developing **analysis-ready data tools** which local governments can use to more easily work with the data the CDRC has on offer (ibid). These issues are also addressed in the training courses offered by the CDRC,

all of which incorporate a CDRC dataset and are aimed at giving users practical knowledge on how to utilise the information provided by the CDRC in practice (ibid).

Table 7: Linking Practices of the CDRC

The Consumer Data Research Centre (CDRC)	
Economic Capital	• Funding from the Economic and Social Research Council; • Operational funds, inducing funding for the networking events; • Funding to cover data contributors' expenses; • Funding for initiating jointly beneficial research projects with commercial partners; • Funds available to provide key stakeholders with financial assistance to attend organisation's events; • Financial resources for the production of online training material.
Social Capital	• "Ladder of Engagement" approach to relationship building; • Leveraging connections inside companies created as a result of their Master's Dissertation Scheme; • Annual networking events for the presentation of CDRC's work/as a value proposition for data contributors and users (User forums); • Early access to research through the virtue of membership in CDRC's network; • Engagement in collaborative research projects with industry and academia partners; • Developing connections with data specialists within the public sector.
Cultural Capital	• Affiliation with high-ranking UK universities; • Human recourses composed of distinguished university academic personnel; • The operation of a Masters Dissertation Scheme and Doctoral Training Program; • Hosting of a wide range of training and capacity building activities for practitioners and academics; • A portfolio of successful case studies; • A high standard template for license agreements, covering different models of data sharing with different levels of security; • Pre-screening and approval of projects requesting access to secure data; • A requirement for providing proof of accreditation on safe research practices as a pre-requisite for secure data access.
Symbolic Capital	• Awareness and anticipation of prevailing themes in political and governmental agendas. • Strong online and media presence through – blogs, social media and participation in outside events.
Technological Capital	• An easily searchable and accessible datastore; • Secure data facilities at three locations; • Advance granularity of datasets at the geographical level; • Development of analysis-ready data tools on offer to potential data users.

5.2 Dalberg Data Insights

Dalberg Data Insights (DDI) is a part of the global Dalberg consulting and advisory firm, supporting better decision-making in the development sector through data analytics (DDI-3). DDI possesses a robust international presence in 17 countries across Europe, Africa, Central and South America and South-Eastern Asia (DDI-1). It was founded in 2017 by the team behind the Belgian start-up Data for Good at Real Impact Analytics, joining the Dalberg firm (Appendix I, DDI-2).

DDI works with a host of developmental topics and issues, chief among them being: **mobility, global health, financial inclusion and gender inclusion** (DDI-6). To help effectively address pressing challenges in these areas, DDI builds innovative data ecosystems based on local community needs (DDI-1). For these purposes, they leverage private and public data sources through partnerships with various stakeholders and analyse it via advanced artificial intelligence and big data technology (ibid).

DDI conducts two core activities: obtaining access to a range of data from private and public organisations and developing new insight-delivering data analysis tools and ecosystems based on this data (ibid).

In terms of **data access**, from the private data contributors' side the DDI mainly partners with telecom operators (Appendix I). More specifically, they work with anonymised Call Detail Records and Mobile Money transaction records (Appendix I; DDI-4). In addition to that, they also work with other privately-held data sources, including both open access and closed private company data (Appendix I).

In the process of securing access to privately-held data, DDI utilises "**data as a service" agreements** (Appendix H). These agreements establish frameworks for secure, ethical and regulatory compliant data sharing between a private organisation and DDI (Appendix I). They entail the signing of one chief agreement at the high representative level that outlines the rules for data sharing, its potential permitted purposes and the data security measures needed for compliance with General Data Protection Regulation (GDPR) and local regulations (Appendix H). Once such an agreement has been put in place, following requests for dataset access for specific projects are processed through a more straightforward application form, which requires less time as it does not call for legal signatures (ibid). These agreements allow for easier long-term collaboration between DDI and private data contributors on various projects (ibid). Notably, one interviewer highlighted that all data-sharing agreements have to additionally comply with **DDI's internal standards of data security and privacy,** which in certain aspects of a higher level than internationally recognised standards (Appendix I).

Once the data access has been legally agreed upon, DDI will arrange access to the data by depositing servers at the data providers facilities and accessing them through a remote connection (DDI-4). In terms of data security against malicious breaches, DDI stays reliant on the data operators' facilities (ibid).

In the process of acquired **data analysis**, the primary tool utilised by DDI's is AIDA (DDI-5). **AIDA** is a **machine learning data platform** developed by DDI that brings together big and small data from private and public sectors and algorithms that generate

real-time insights and allow the building of new tools (DDI-1, DDI-5). The technical tools of AIDA, which include machine learning and artificial intelligence, allow it to assess, monitor and predict key dimensions of different developmental issues (DDI-5). It has been utilised in practice to support data source integration for many of DDI's projects, such as the Food Security Manager, an interactive tool that helps predict areas at risk of food shortages (DDI-7).

In developing such analytical tools, DDI **closely cooperates with end beneficiary organisations** and, through continuous feedback, strives to better account for the needs of all relevant stakeholders (DDI-1). On occasion, DDI will also cooperate with other partners, such as third-sector organisations present on the ground, with previous histories of working with the issues in question (DDI-7).

The data tools produced by DDI as a result of this process are notable for several factors: they provide **comprehensive and real-time data**, as opposed to more traditional field studies, which usually only produce snapshot views; the data is graphic, processed, and stabilised, making it more informative for practitioners; and the data tools are often interactive in nature (DDI-8). However, it should be noted that as a standard practice, DDI **never shares with their public sector partners the underlying private data**, but only the insights derived through its analysis (Appendix I).

The interviewees emphasised several noteworthy linking mechanisms utilised by DDI in the process of delivering private data insights to public entities.

Firstly, in finding and forming projects with public sector actors, they **cooperate closely with established international and third-sector organisations on the ground** with histories of partnerships with public bodies in a given country (Appendix H). These organisations help DDI identify the areas of public actors' needs (ibid). According to one interviewee, particular public bodies such as ministries will usually have a close partner/donor organisation to whom they will communicate their particular challenges in a given area (ibid). The partner organisation will publish a call for project proposals, detailing the particularities of the issues and desired solutions (ibid). DDI will then utilise this information to provide an initial data-based solution proposal (ibid).

In most cases, the same partner organisations will also put forth the **funding** necessary for DDI to execute the project (ibid). DDI is a for-profit organisation, however, in certain instances where it perceives an opportunity for making a significant impact or setting the groundwork for future work, it also provides services in a **pro-bono** manner (Appendix I).

DDI's relationships with major telecom providers and international donor organisations are supported by the **connections developed with these organisations by the Data for Good team through the years before merging with the Dalberg group** (DDI-2). Furthermore, one interviewee noted that the leveraging of previous connections with partners and building long-term relationships is also positively aided by the **conferences and similar networking events** that DDI regularly participated in previous to the Covid-19 pandemic for connecting with potential clients and for sharing key takeaways and learnings from their work with the larger community (Appendix I). In a similar vein, in the process of collaboration forming with clients, DDI also benefits from the **Dalberg group's recognition**, especially in the geographic regions where the group holds a particularly long track record of activity (ibid).

Another factor referenced as a significant enabler in establishing relationships with beneficiaries by both interviewees was DDI's **strong presence on the ground** in its focus counties (Appendix I Appendix H). According to one interviewee, most staff in a country's office are made up of locals who bring to DDI a deep understanding of the local contexts and ecosystems, which elicits respect in their clients and simplifies the process of communication and partnership (Appendix I). This mechanism relates to another important measure of **connecting with the right champions** within public sector organisations, who believe and understand the potential of data-driven transformation (Appendix H). In this regard, referencing the DDI's **portfolio of similar use-cases** from the same or different geographies presents a valuable tool for communicating the viability of the process to potential champions in the public sector (ibid).

Once a partnership has been established, depending on the geography, the sector of activity and the needs of the public sector clients, the datasets required for the delivery of actionable insights vary greatly, and in some instances, relevant data could be highly scarce (Appendix I). According to one interviewee, to mitigate these challenges, the DDI adopts two main approaches: they **continuously exercise to stay up to date** with new technologies and trends on the data market to have as wide an array of possible solutions available as possible; and they try to expand on the available data by allocating part of their project budget to **gathering primary data** through the support of external partners, that they later analyse in combination with other available large datasets (ibid).

Appealing to a company's sense of corporate social responsibility, especially in relation to its mission alignment with that of DDI's given project goals, and to **regulation mandating private company disclosure of data** that may exist in a given country are some other methods of accessing privately held datasets utilised by DDI (ibid).

A further important linking mechanism emphasised by both interviewees from DDI is their **focus on the sustainability** of the data ecosystem produced through their services (Appendix I, Appendix H). According to one interviewee, the most common apprehensions expressed by potential beneficiaries in the process of collaboration forming are doubts around issues of long-term financial and capacity sustainability of the data-driven solutions produced as the result of the partnership (Appendix I). In response to this, at the end of their projects, DDI helps their clients develop strategies for sustaining the solutions long-term, including potential sources for obtaining additional necessary funding and support in drafting corresponding project proposals (ibid).

Table 8: Linking Practices of DDI

	Dalberg Data Insights (DDI)
Economic Capital	• Project Funding provided by donor organisations; • Funds available for the acquisition of telecom data; • Recourses available for the commissioning of primary data gathering; • Recourse for completing pro-bono work.
Social Capital	• The partnerships with large donor organisations previously developed by Data for Good; • Co-designing of solutions with end beneficiaries and other partners; • On the ground presence, staff with a strong understanding of local context; • Regular participation in conferences and other experience and learnings sharing activities; • Finding the correct champions within public-sector organisations for collaboration.
Cultural Capital	• Membership of the Dalberg Group and access to its wider talent pool; • "Data as a Service" data-sharing agreement format with private sector data partners; • Principle of sharing insights but not underlying private data; • High level of internal standards on data privacy, security and ethics; • Company practice of continuously staying up-to-date with new technologies, trends on the data market and the requirements of the clients; • A portfolio of case studies of similar use cases Collecting supplemental primary data for analysis; • Focus and assistance on the long-term sustainability of data projects.
Symbolic Capital	• The Dalberg group reputation and recognition; • Appealing to the alignment of DDI's work with corporate social responsibility missions of telecom companies; • Appealing to local regulatory mandates on data disclosure.
Technological Capital	• The AIDA machine learning data platform; • Real-time and interactive data analysis tools; • Secure data access through on-site servers

5.3 DjustConnect

DjustConnect is a Belgian data-sharing platform focusing **on the agricultural sector**, based at the Flanders Research Institute for Agriculture, Fisheries and Food (ILVO) (DJU-1, DJU-2). It is a neutral platform open to all potentially interested users from the agri-food sector (DJU-4). They aim to provide farmers, horticulturists, private and public

sector actors with a high performing, secure ICT infrastructure and transparent rules of the game to facilitate innovation in the agricultural sector and, in particular, the development of more targeted and smart applications (DJU-1, DJU- 3).

DjustConnect grew out of a **European Fund for Regional Development financed project** on the potential of data in the agricultural sector and was further developed through the support and additional funding of five private companies in the agricultural sector (DJU-1, DJU-4). It was officially launched in 2019 as a permanent public-private body, with ILVO in charge of platform maintenance and supervision (DJU-1).

DjustConnect **gathers data directly from farmers, horticulturalists and agricultural companies, such as machinery producers, service companies and retailers** (DJU-2, Appendix J). Its data sharing model is based on the EU Code of Conduct on Agricultural Data Sharing, in which the farmer is recognised as the owner of the raw data (DJU-1). In light of this, DjustConnect grants farmers the ability to make final decisions about data access, including confirming data disclosures by private companies holding their data and allows them to withdraw their permission to share data at any stage (DJU-2). This process is simplified in practice by the online dashboard available to farmers that simplify the process of overviewing data access requests (DJU-1). One interviewee referenced this approach as an essential measure in creating trust among data providers and facilitating exchange (Appendix J).

The chief user base of DjustConnect consists of private companies interested in improving various services, however, there are also users from the public and civil sector (Appendix J). Potential data users can access available datasets via the DjustConnect data shop, through which interested parties are able to request consent for access from data owners (DJU-2). In filing such data access requests, users must clearly define their intended use of the data (DJU-1).

The data shop allows data owners and contributors to connect, but it does not store data itself (Appendix J). According to an interviewee, this was a strategic choice to avoid regulatory limitations and keep DjustConnect free of responsibility for ensuring data quality (ibid).

Data users gain access to the data shop by payment of a **yearly subscription fee** (ibid). The rationale behind the subscription fee is solely the coverage of DjustConnect's operational costs, as they operate on a non -profit model, with the goal of **cost neutrality** (ibid). One interviewee emphasised that this cost neutrality approach is a factor that differentiates DjustConnect from many similar initiatives globally and an indicator of their overall commitment to neutrality in all aspects (ibid). The impartiality of

DjustConnect is meant to be additionally guaranteed through the involvement of ILVO and the deliberate refrain from developing data fusion-based applications by DjustConnect itself (DJU-1). According to an interviewee, this **neutrality is a crucial linking measure** as it enables DjustConnect to bring together different actors and competitors and initiate discussion around data sharing (Appendix J).

The interviewees emphasised several additional key linking measures. One of these was the process of communication and sensibilisation of data owners about data sharing (ibid). This process consists of two elements. First is the **adoption of a modified "train the trainer approach",** whereby DjuctConnect communicates the advantages of their data sharing model to several farmers and recruit them to use word-of-mouth to further disseminate this information among other workers in the agri-food sector (ibid). According to the interviewee, this model is successful because potential data contributors are more trusting and open to information from persons from the same field (ibid). In relation to working with companies, DjustConnect prefers to approach them directly (ibid). Second is **the improvement of the digital and data literacy** of stakeholders (ibid). This area of activity focuses on several directions: building trust towards the particular digital application and systems used by DjustConnect, creation of an understanding of the potential value of data held by data owners and the possibilities of the ICT landscape and knowledge development around issues of data quality, formatting and integrity (ibid). According to the interviewee, filling such digital knowledge gaps is essential to enable data exchange and cooperation across stakeholders (ibid).

Additionally, DjustConenct facilitates the creation of extra **incentives and benefits of data sharing for data owners,** both financial and non-financial (ibid). In the financial form, these additional incentives might take the form of premium payments to suppliers for the disclosure of the origins and growth details of the products (ibid). Other, not directly financial forms might be expressed through guarantees of acceptance of product for sale by retailers, arrangements that result in the saving of time (ibid).

A notable technological measure is the ability to **build other applications on top of DjustConnect** that offer data contributors added value (ibid). An illustrating example of this used by DjustConnect is IKM, an application aimed at easing the administrative burden on farmers by automating the process of reporting documentation collection by letting all the necessary documentation flow directly to auditors on the basis of data owner's consent provided through DjustConnect (DJU-2).

Table 9: Linking Practices of DjustConnect

	DjustConnect
Economic Capital	• Initial funding from the European Fund for Regional Development; • Financial support from five private-sector partners; • Income from yearly subscription fees enabling the goal of organisational cost neutrality.
Social Capital	• Adoption of a modified "train the trainer" approach; • Emphasis on communication and education of potential data contributors;
Cultural Capital	• Affiliation with the Flanders Research Institute for Agriculture, Fisheries and Food; • Adherence the EU Code of Conduct on Agricultural Data Sharing and a high degree of discretion to farmers as data owners; • Model based on not storing data themselves; • Facilitation of financial and other benefits as incentives for data sharing.
Symbolic Capital	• Assurance of complete impartiality; • Expressed commitment to stimulate the agri-food sector.
Technological Capital	• User-friendly Dashboards for data contributors • Data Shop Platform • Capacity for build-on applications, such is IKM-NET

5.4 Positium

Positium is an Estonian-based company that provides data solutions and analytics based on anonymised mobile data for decision-making in mobility, tourism, and population management (POS-1). They aim to provide innovative solutions to common challenges based on their flagship methodological and technological tools (POS-3).

According to the interviewees, the inception story of Positium relates to Estonia being selected as the next host of the Eurovision Song Contest in 2001 (Appendix L, Appendix K). This created a need for statistical information on the amount of event generated inflow of tourists and their movement (ibid). As there were no available relevant data sources for producing such necessary insights, several researchers at the University of Tartu in Estonia with backgrounds in human geography, engineering, and architecture proposed the use of mobile phone data as a possible solution (ibid). However, after this initial idea, it took several years to negotiate access to privately held mobile datasets, the first of which were analysed at the University of Tartu in 2003 (Appendix K).

In 2006 Positium became a **spin-off company of the University of Tartu** (Appendix L, Appendix K). Though the company is since independently governed, they continue to closely collaborate with the University of Tartu, adopt similar research methodologies and employ many members of staff linked with the University (Appendix L).

Positium's clients come from both the public and the private sector (POS-1). These include regional development organisations, cities, municipalities and national banks (POS-3). However, they **chiefly partner with public entities**, namely national statistical offices and different governmental ministries (Appendix L). Moreover, their client base is international, with projects completed in 12 countries across four continents (POS-3).

Positium operates on a **commercial model,** however, they also complete some projects or portions of work on a **pro bono** basis (Appendix L). Initially focused on data brokerage, they recently began transitioning to a more software and methodological company model, supplying technical tools and expertise for mobile data analysis (Appendix K). However, pertinent to their long-term relationships with mobile operating companies, their work continues to include elements of facilitating data exchange connections (Appendix L, Appendix K). In such cases, Positium acts as a middleman connecting two actors and does not sell the data itself (Appendix L).

With respect to private data, Positium mainly works with **mobile positioning data** sourced from different mobile operators (Appendix K, POS-1). These include domestic data from subscribers within the country, inbound roaming data from foreign subscribers and outbound roaming data (POS-3). In addition, they also utilise public datasets, particularly those held by their public sector partners for supplementary analysis (Appendix L).

The primary tool employed by Positum in the process of mobile data analysis is their flagship product, the **Positium Data Mediator,** a scalable big data processing and analytics engine (POS-4). It enables data to be anonymised, aggregated, quality assured, methodologically analysed and presented in a comprehensible way via dashboards (POS-3). In other words, it is a custom software consisting of over 94 methodological steps that processes billions of points of raw mobile operator data to produce specific insights (Appendix K). Notably, it can analyse both historical data and near-real-time data feeds, making it capable of giving long-term views on areas of activity (POS-4).

As a result, Positium produces statistical indicators used by policymakers or similar analytical teams (POS-2). For example, in the field of tourism, such indicators might include detailed information on tourist destinations, origins and movement patterns (POS-3). Such indicators can be used for various aims, as demonstrated by the dashboard developed by Positium for disaster management for Estonia, capable of providing three-layered daily and hourly views on population location (POS-5).

In terms of linking measures, there were several practices outlined by the two interviewees as significant to their practice.

Both interviewees referenced presence and **active participation in international conferences** as the chief way of connecting to new public sector clients (Appendix L, Appendix K). They are mainly present in global forums centred around new data sources for official statistics (Appendix K). Positium's attendance at such events is further enabled by its **membership and partnerships in international organisations** such as the United Nations Global Working Group on Data for Official Statistics, Eurostat and International Telecommunications Union (Appendix K). Through such membership, they generate awareness of their activities, allowing public sector actors to easily reach out to them once they develop data needs (Appendix L, Appendix K). In connecting with potential clients, Positium is further aided by its reputation as one of the very few companies doing this type of mobile data analysis and its long experience in the field (Appendix L).

Furthermore, Positium also actively engages in **public sector partner outreach** (ibid). In particular, they aim to facilitate connections with well-placed individuals in government structures, usually at a high level, capable of championing partnerships (Appendix K). Whenever possible, they also aim to work with statistical officers since this simplifies communication due to shared background knowledge of data standards and requirements (Appendix L).

In dealing with international public sector partners, Positium relies on **local partner agents on the ground** for the handling of logistics and communication issues (Appendix L). This approach helps Positium mitigate the lack of resources for establishing subsidiaries for their international partnerships (Appendix K).

In terms of **obtaining access to private data sets**, Positium relies on four chief linking measures: **i.** reimbursing mobile operators for data access (covered by their clients) **ii.** appealing to national regulations that mandate the release of mobile data to local statistical offices **iii.** relying on public sector clients to contact mobile operators and negotiate the access to data and **iv.** appealing to mobile operators' sense of corporate social responsibility (Appendix L, Appendix K). According to one interviewee, in the latter case, it is instrumental to clearly demonstrate how the data links to output and societal benefit, in which process Positium references the United Nations Sustainable Development Goals and the objective to which their activities contribute (Appendix L).

Assurance of **data security** is another critical measure undertaken by Positium in forming collaborations with private data providers (Appendix L, Appendix K). Positium approaches this issue from both the angle of personal privacy and the protection of the business interest of telecom operators included in the data (Appendix L). According to

one interviewee, Positium holds an excellent track record in this regard, not having encountered any privacy and security issues over their years of operations (Appendix K).

Apart from its reputational effects, Positium's **long-term experience in the field** also influences the sophistication of its data analysis methodologies and capabilities (Appendix L). Foremost, the Positium Data Mediator itself represents a conglomeration of 15 years of work experience that ensures the consistent quality of produced insights and allows Positium to keep up a regular production of statistics and differentiate itself from competitors who mainly focus on one-off projects (ibid).

These experiences allow Positium to reassure clients of the viability of insight that can be delivered from mobile data and its capability of resolving most issues that can occur in the mobile data analysis processes (Appendix K). One interviewee emphasised that this high level of expertise and use-case experience across countries grants them **trustworthiness** in the eyes of potential partners (ibid). In addition, in the communication of these factors, the portfolio of past use cases has been found to be a valuable tool and an added value offer to clients in the form of illustrating examples (ibid).

The quality of Positium's insights output is additionally ensured by their strong focus on **confirming the high quality of input data** they process with the Positium Data Moderator, which one of the interviewees referenced as a critical area of focus in their communication with potential clients (Appendix L).

Table 10: Linking Practices of Positium

	Positium
Economic Capital	• Financial resources for completing pro bono work.
Social Capital	• Membership in prominent international organisations working on data and statistics; • Active participation in conferences around topics of data for statistics; • Partnerships with local agents on the ground; • Relying on the influence of public sector partners for negotiating private data access; • Good personal connections with existing partners; • Connections with high-level individuals within public structures and statistical officers.
Cultural Capital	• Affiliation with the University of Tartu; • Staff with academic background; • Diverse knowledge and expertise background of founding members; • Utilising data owned by public sector clients as supplementary data for analysis; • Vast portfolio of past use-cases.

Symbolic Capital	• Harnessing the societal resonance of the need for statistical information around the time of their founding. • Over 15 years of experience in the field of private mobile data analysis for public needs, uncommon in the field; • Referencing the contribution of Positium's work to United Nations Sustainable Development Goals; • Perception of trustworthiness.
Technological Capital	• Positium Data Mediator; • Insurance of high quality of input data; • Cloud services for the hosting and processing of data.

5.5 Urban Big Data Centre (UBDC)

The Urban Big Data Centre (UBDC) is a research centre and national data service based at the University of Glasgow (UBDC-1). Similarly to the CDRC, the UBDC was established in 2014 with funding from the United Kingdom Economic and Social Research Council as part of the Big Data Network Initiative, which aimed to make data routinely collected by businesses and local government organisations more accessible for research purposes (CDRC-2).

The focus of the UBDC lies in developing novel solutions for the use of urban big data in innovative research aimed at improving social, economic and environmental well-being in cities (UBDC-1, UBDC-9). It hopes to facilitate such innovation through the lowering of the barriers of entry to any social scientists interested in exploring urban issues through data-based approaches (Appendix M). For these purposes, they bring together a multi-disciplinary team of experts and foster collaborations with a diverse group of government, industry and third sector partners (UBDC-1).

According to the interviewee from UBDC, their work comprises two principal functions: conducting research on urban issues and providing a national data service, with the latter being the original function of the centre on the remit of which they received funding (Appendix M).

In terms of their **research function**, the key research directions of the UBDC run the whole gambit of urban issues (UBDC-9). These include housing, neighbourhoods, transportation, infrastructure, educational disadvantage and most recently, the urban impacts of the Covid-19 pandemic (UBDC-5). **The UBDC conducts research both independently and in collaborations with different partners, among them from the public sector** (Appendix M). An example of a joint UBDC-public sector research partnership was a project with the Glasgow City Council that aimed to develop a method

of evaluating the impacts of a long-term infrastructure initiative within Glasgow city (UBDC-10).

Related to their research function, the UBDC also conducts various educational and training activities (UBDC-9). Namely, the UBDC runs two Master's Degree Programs in urban analytics and urban transport (UBDC-6). They also provide a range of free online tutorials and seminars around data analytics techniques, especially those relevant to urban data (UBDC-7). Notably, within this framework, they have produced **several knowledge-sharing presentations specifically focused on identifying opportunities for urban data analysis-based collaborations between academic researchers and local public authorities** and the benefits of such collaborations (ibid).

The **data service dimension** of the UBDC's work consists of several elements. Firstly, the UBDC cooperates with both commercial and public sector partners to gain access to datasets with potential value for delivering insights into urban issues (UBDC-11). They reach out to potential data contributors based on their research or intuition of what datasets exist and their potential value in relation to urban issues (Appendix M). However, notably, they also provide **a data sourcing service**, whereby researchers or other potential data users may approach the UBDC with research issues they have identified and the type of dataset they required for its further analysis, and the UBDC will use its specially designated resources to procure the relevant datasets for them (UBDC-9, Appendix M).

The UBDC offers all potential data contributors **a standard data-sharing agreement with the possibility of the addition of bespoken terms** (UBDC-2). Such bespoken terms might limit the period and permitted types of data use to mitigate any perceived risks from the data contributors' side (UBDC-2, Appendix M). This makes some datasets available only for academic research purposes, whilst others might be accessible for public sector bodies or commercial research (Appendix M). The UBDC also negotiates sub-licensing terms based on which data can potentially be published as part of academic research (ibid). The UBDC always aims to obtain wide-reaching sub-licensing agreements in order to be able to supply as large a community of researchers as possible (ibid).

Based on the sensitivity of the data and the needs of the data provider, the **UBDC offers several levels of data access**: open, safeguarded and controlled (Appendix M, UBDC-11). These different types of data sharing models entail different types of restrictions and freedoms in terms of data access environments, types of usage and governance approvals (UBDC-11). Open data collections are the most easily accessible, requiring no pre-screening before use (UBDC-2). Safeguarded data involves the UBDC assessing an applicant's eligibility to access the data per the licensing agreement before data is handed

over (ibid). The application processes can vary in complexity from case to case, but all of them include the presentation of the research aims and process (Appendix M). Finally, at the highest level of security, in addition to eligibility screening, data has to be accessed remotely in a controlled environment (UBDC-2). In such instances, the **controlled access is provided virtually by the centre's safe haven' partner**, with the UBDC not requiring physical presence at premises from any of its data users (Appendix M, UBDC-2). Moreover, to access higher security levels of data, all users are required to agree to the UBDC's terms of use, which include provisions on penalties for breaches, up to and including legal action (UBDC-2).

Once the data has been acquired, the UBDC maximises its value by assessing, documenting and enhancing its quality (UBDC-11). In particular, the UBDC **enhances data value** by extracting data from multiple organisations, linking it, and conducting real-time data analytics and generating synthetic datasets (UBDC-9).

The UBDC operates a **data catalogue** where potential users can find all datasets in UBDC's collection categorised by thematic areas (UBDC-3). As of June 2020, the UBDC's data service was supporting around 800 users of data (UBDC-11). The majority of data users were representatives of academia, however, there were also a significant number of users from the public sector and third sector organisations (ibid). Additionally, though the dominant part of data users were UK based, there was also a widespread international representation of the user's countries of origin (ibid).

The linking practices that the UBDC employs stretch over both of their principal functional activities. According to the interviewee, one of the most influential linking measures utilised by the UBDC in the process of data acquisition is their considerable **financial resources** (Appendix M). The UBDC often cooperates with organisations whose commercial models are, to a certain extent, based on the sale of data as a service (ibid). This means that the UBDC's budget for data license purchasing is often an effective and thus common method in securing datasets (ibid). However, the interviewee noted that this approach is not ideal and that recently the UBDC has been taking steps to establish more long-term and cooperation-based relationships with data partners (ibid).

One method for doing so in UBDC's experience has been developing interpersonal connections and social familiarity with partners, which facilitate the starting of new partnerships as well as contribute to their positive outcomes (UBDC-10). In addition, the interviewee emphasised the social aspect of professional relationships as also being important for the ensuring sustainability of collaborations over longer periods of time, especially in smaller company settings (Appendix M).

On a similar note of interpersonal connections, according to the interviewee, one of the significant challenges in data acquisition has been the identification of the right persons for communication in terms of data literacy and competence to make decisions on data sharing (Appendix M). Thus, an essential linking mechanism that the UBDC employs is the **identification of the relevant contact persons** in organisations, through staying aware of emerging research completed in partnership with commercial companies of interest; following up on personal recommendations and scouting through online platforms such as LinkedIn (ibid).

For these purposes, the UBDC regularly conducts different events focusing on recent examples of big data application in urban issues and, more specifically, on the centre's work and past successful projects (UBDC-8). The UBDC also aims to ensure widespread awareness of its service across the research community and beyond (UBDC-11). In this aim, they participate in external conferences and keep an active presence on different social media platforms (ibid). Moreover, they publish **calls for invitations of expressions of interest** in their newly acquired datasets to create awareness among the research community about the possibilities on offer (UBDC-4).

An additional big aspect for establishing more trust-based, long-term partnerships for the UBDC is ensuring that all partnerships produce mutual value for all sides (Appendix M, UBCD-10). In this regard, one value proposition from UBDC towards private data contributors is expressed through the provision of **high academic level validation of data credibility,** through assessment of its representatives and value (Appendix M). If the UBDC deems the dataset as a credible basis of robust evidence, that can carry weight within the academic community and thus enhance the value of the data on the market (ibid). The UBDC can also find potential audiences for the data and increase awareness around the availability of these datasets in subjects interested in its potential insights but not authorised to access it based on the licensing agreement with the UBDC (ibid). This, in turn, can result in new business for the data contributor organisation (ibid). Moreover, as an additional by-product of the research conducted by the UBDC, **they can offer data contributors feedback on how to more effectively organise their datasets** and handle issued of data privacy (ibid).

Notably, according to the centre's director, within the UBDC's research functions, a particular characteristic of the UBDC is that it is led more by social rather than data scientists, as is more typical of similar types of research centres (UBDC-2). This, in turn, makes them closely connected to the urban societal changes they work to address. Thus, when it comes to data acquisition, they are more **problem-centric** and start by focusing on the issues they aim to address and seek out datasets accordingly (UBDC-2). However,

although the UBDC sometimes grounds its data outreach activities in concrete issues or requests of specific researchers, as a rule, they broaden the scope of the data they request in order to generalise the obtained dataset and enhance its wider usefulness for the entire research community (Appendix M).

Lastly, concerning specifically governmental audiences in their capacity as data users, the UBDC has recently started developing new initiatives for making data more easily accessible to them through the **use of tools such as dashboards and visual interfaces (**ibid). According to the interviewee, such accessibility is an essential prosecutor to any partnership as it is the only way to initiate a comprehensive conversation (ibid).

Table 11: Linking Practices of the UBDC

	The Urban Big Data Centre (UBDC)
Economic Capital	• Funding from the Economic and Social Research Council; • Considerable budgetary recourses to allocate for the purchasing of data; • Budgetary recourses allocated for the development of data accessibility tools.
Social Capital	• Forging connections with suitably placed individuals within partner and potential partner organisations; • Developing long-term partnerships with local government administrations through networking events and joint research projects; • Scouting potentially valuable contacts through professional platforms and network events.
Cultural Capital	• Affiliation with the University of Glasgow; • Offering three models of data sharing with different levels of security; • Pre-screening and approval of projects requesting access to secure data; • User agreements including penalties for breaches of terms of data use; • Provision of data sourcing services to researchers; • Facilitating data analysis potentially beneficial for both sides of the exchange; • A multi-disciplinary team with a wide range of expertise; • A problem-based approach to data collection; • Providing feedback to data contributors on their data sharing business models and data organisation. • Educational programs, including Master's schemes and knowledge sharing presentation specifically focused on urban data analysis-based collaboration with public authorities.
Symbolic Capital	• Ensuring widespread awareness of the service across the UK research community and beyond; • Offering private data donors, the establishment of credibility of their data as a robust evidence base, thus increasing its value on the market; • Publishing announcements for calls of interest around newly available datasets.
Technological Capital	• Providing a thematic data catalogue; • Remote virtual access provided by a trusted third partner for secure use of sensitive datasets; • Maximise the value of the data collected through real-time analytics, data capture and linkage and synthetic data generation; • Broadening and enhancing the generalisability of datasets to make them valuable to a wide community of researchers; • Producing tools such as dashboards and visual interfaces to enhance the accessibility of data.

6 Discussion

The following chapter will provide a comparative analysis of the case findings and identify the commonalities and differences that emerged in linking practices across data intermediaries. This discussion will be structured in the following manner: Section 6.1. will identify the measures that were present in a closely similar manner across all or several cases; Section 6.2. will overview the measures that were also common across several cases but implemented with noteworthy variations in approach and lastly; Section 6.3. will outline the measures that were unique only to select individual cases.

6.1 Measures Present Across Cases

Access to public funding emerged as a vital facilitator to several data intermediaries' operations. Intermediaries benefiting from public financing included DjustConnect, which received initial funding from the European Fund for Regional Development and which continues to currently be run by a publicly funded body (Appendix J), the CDRC, and the UBDC, both of which are funded by the Economic and Social Research Council (CDRC-2). Public funding was also named an enabling factor by DDI, who rely on grants and financial support from dominantly international public donor organisations to deliver private data insights to beneficiaries (Appendix K). These observations are in line with the suggestions of the stimulating effects of government action on the functioning of data intermediaries expressed in literature (Janssen & Zuiderwijk, 2014).

Additionally in the area of economic capital, in both cases operating on a for-profit basis, namely DDI and Positium, availability of financial resources that permitted the conduct of **pro-bono work** was found to be an important linking measure, with interviewees from both intermediaries referencing their gratuitous work as a means of relationship building and entry into new markets of operation (Appendix I, Appendix L).

Regarding measures in the area of social capital, the importance of **personal connections** emerged as the most repeatedly referenced linking measure across all intermediaries. Particular emphasis was made on the maintenance of good personal relationships with existing partners by the CDRC, Positium and UBDC (Appendix F, Appendix M, Appendix L).

As a particular branch of personal connection development, the **research, identification and outreach to relevantly and suitably placed individuals within public institutions** appeared as an additional linking measure**.** In particular, the CDRC, DDI, Positium and UBDC referenced their directed efforts of connecting with individuals in public institutions with a degree of data expertise as a means of smoother

communication (Appendix G, Appendix M, Appendix H, Appendix K). Moreover, DDI and UBDC also emphasised the importance of connecting with persons with the competence to make decisions on data sharing and approve potential partnerships (Appendix M, Appendix H).

In a similar vein, the CDRC, DDI, Positium, UBDC emphasised their regular participation in different **networking events** and conferences as one of the principal ways of forging connections with both data donors and public sector users. Furthermore, DDI and Positium stressed their **memberships in various international organisations** focused on data for global development as a highly effective means of creating awareness around their work and enabling public sector users to easily connect with them once they develop a need for data insights (Appendix I, Appendix L).

The broad coverage and continued emphasis on the importance of professional networks and outreach events are consistent with the literature on data intermediaries, which also suggested social connections and histories of previous collaborations as being principal tools for the facilitation of data exchange (Klievink, Van der Voort & Veeneman, 2018; Robin, Klein & Jütting, 2016; Susha, Janssen, & Verhulst, 2017b; Van Schalkwyk et al., 2016).

Networking and related outreach events often link to an intermediary's **activities in the area of education,** which were also commonly observed across cases. For example, the CDRC and the UBDC both run Master's Dissertation Schemes in data analytics and provide other educational resources for practitioners on the subject (CDRC-7, UBDC-6). Similarly, DjustConnect organises data literacy workshops for its stakeholders (Appendix J). While the educational activities of DDI and Positium are not as institutionalised, they nevertheless emphasised their practices of sharing valuable lessons learned through their experience via various channels (Appendix I, Appendix L).

Relatedly, a common expression of cultural capital observed in all but one case was the **affiliation and close partnership with established higher educational institutions**. In Positium's case, such affiliation was mostly initial, with the organisation going on to become an independent spin-off company. However, Positium continues to closely cooperate with its founding university, particularly in terms of hiring (Appendix L). In other instances, such as the CDRC, DjustConnect and UBDC, university affiliations are more prominent, with all entities operating as their respective university projects managed by academic staff (CDRC-1, Appendix J, Appendix M). Notably, in the case of the CDRC, it possesses a close association with not one but several educational institutions.

Moreover, **high academic credentials, experience and knowledge of intermediaries' staff** were also highlighted as value propositions in the majority of the cases. The CDRC and the UBDC referenced this factor in connection with their conducting of high-quality research (Appendix G, Appendix M). DDI presented its access to the Dalberg Groupe data pool, which includes experts with a wide array of backgrounds, as a significant asset (Appendix I). Positium named the academic background of its founding members as an influential factor around its conception and creation, as well as identified their team's vast experience in data analysis and methodological approaches as the chief reason behind their trustworthiness and reputation among the public sector (Appendix L).

The proliferation of these two factors across studied cases is a valuable observation. Academic qualification or association with academic practices are the core essence of cultural capital (Bourdieu, 1986). Thus, although these factors have not been as broadly discussed in the literature on data intermediary linking measures, the findings suggest that cultural capital, in fact, presents a crucial resource for data exchange facilitation.

The development and access to a **portfolio of successful case studies** was also commonly identified as an operative linking measure, especially in the communication of the potential of private data insights to data contributors and public partners. The interviewees from CDRC, DDI and Positium directly acknowledged this fact, with all of them stating that illustrating examples of similar use cases from their experience serve as an effective means of acquiring new partnerships (Appendix F, Appendix I, Appendix H).

Adherence to high standards of data protection as a means of building trust and mitigating the concerns of data donors was observed across the board. All intermediaries reported adherence to the GDPR, with DDI and Positum additionally reporting compliance with all other relevant local regulations in their international projects (Appendix I, Appendix L). Select intermediaries, such as DDI, reported further compliance with higher-level internal rules of data security, privacy and ethics (Appendix I). DjustConnect presented a partial exception to this observation, as they do not store or manipulate data themselves, however, they still mandate adherence of all their platform users to European data security and privacy guidelines (Appendix J). These reflect the observations in literature of privacy measures being identified as an established practice among data collaboratives (Winowatan et al., 2020; Winowatan, Young & Verhulst, 2020a).

Lastly, in the realm of cultural capital, two practices referenced in the literature on governance measures were found to be present in the studied intermediaries. These were

the practices of the **negotiation of bespoken terms** and **the development of standard template agreements for data sharing**, as discussed by Martens & Duch-Brown (2020) and Groves & Neufeldi (2017). References to the regular practice of negotiating bespoken terms for data sharing were made by the UBDC, DDI and DjustConnet (in the last case in a slightly modified manner of allowing their platform users to freely negotiate their own terms of data sharing) (Appendix M, Appendix I, Appendix J). As for the second practice of reliance on standardised templates for data sharing agreements, wide use of such templates for regulating relationships with both data providers and data users were observed in the CDRC, DDI and DjustConnect (Appendix F, Appendix H, Appendix J).

The expressions of symbolic capital were found to be the area most divergent among intermediaries. However, all intermediaries confirmed that a high degree of relevance and symbolic value of targeted issues at the basis of a data-sharing considerably aided partnership establishment. A specific measure commonly observed in this sphere was to **appeal to data donors' sense of corporate social responsibility** through allusions to the societal value of the intermediary's work. For these purposes, DDI showcased the alignment of their work with a particular organisation's social mission statement (Appendix I), while Positium communicated the value of their outputs by connecting them with specific Sustainable Development Goals (Appendix L). For DjustConnect, this process took the form of expressed commitment to the stimulation of the agri-food sector (DJU-1).

In the realm of technological capital, the most common pattern that emerged across the board was the use of **data catalogues and shops**. CDRC, DjustConnect and UBDC were all observed as having easily searchable data catalogues, categorising the datasets on offer according to thematic areas and geographical location (CDRC-5; DJU-2; UBDC-3.)

Specialised and flagship software for data analysis was a common mechanism employed by the intermediaries DDI and Positium, through their AIDA and Positum Data Mediator platforms. Both data analysis tools were characterised by their ability to deliver insights in a comprehensible manner through artificial intelligence and the integration of a wide set of private and public data sources (DDI-5, POS-4). Relevantly, the use of analysis software was only found in the cases operating on a for-profit model.

Lastly, in the technological domain, **tools for data accessibility such as dashboards and visual interfaces** were found to be employed in four intermediaries. UBDC has recently started developing such tools to make data more accessible to the public sector and initiate educated partnership conversations (Appendix M). DDI and Positum already have an established history of utilising such tools within their data analysis platforms (DDI-5, POS-4). DjustConnect also referenced their use of dashboards, with the

difference that they use them to make data sharing more accessible for data owners rather than the end-users (Appendix J).

6.2 Common Measures Implemented Through Varying Approaches

A key financial linking measure reported by all intermediaries was the **availability of funds for the compensation of contributors for their data**. However, there was substantial variation in the implementation of this measure in terms of its degree of reliance, the amounts of compensation and the source of the payments. For example, the CDRC reported that it, in general, avoids payments for data, except in cases where it is meant to compensate for the data processing costs incurred by the provider, and even in such instances, it never pays commercial rates. (Appendix F). On the other hand, UBDC expressed strong reliance on their data purchasing budget, naming it the most valued resource for acquiring private datasets (Appendix M).

Furthermore, both DDI and Positium mentioned payments for private data as being a facilitating factor in accessing private datasets, though in their case, the payments were sourced from their clients, either directly from beneficiaries or partner organisations supporting the project (Appendix I, Appendix L). Lastly, DjustConnect also relied on the use of financial incentives for encouraging data sharing, however, in their case, this discussion is held directly between the owners and users of data, with DjustConnect playing a facilitating role (Appendix J). Notably, the remuneration of contributors for costs incurred in the opening of their datasets was identified as one of the potential means of overcoming the economic imbalance of private-public data exchange (Eckartz, Hofman & Veenstra, 2014).

Along the same lines of negotiation of terms of use, **assurances of data use only for specifically agreed purposes** was observed as a critical measure in gaining private dataset access in the area of cultural capital. This observation closely reflected the literature in the area of governance measures of data intermediaries. Said literature underline the importance of institutionalised mechanisms of adherence to agreed-upon purposes of data use as an essential element of building trust among data contributors (Klievink, Van der Voort & Veeneman, 2018; Susha & Gil-Garcia, 2019; Winowatan, Young & Verhulst, 2020a). In the observed cases, such institutionalised methods varied. CDRC and UBDC employed similar methods of classifying available datasets into three categories of access dependent on the sensitivity of the data and requiring submissions of project proposals detailing the purposes and methods of data use before granting of access (CDRC-6; UBDC-2). DDI managed the process through their “data as a service” agreements, with the intended purposes of data use being signed off on by the data owners through a simplified process on a case-by-case basis (Appendix K). Lastly, DjustConnect

approached the matter by presenting data owners with requests for data access and description of intended data use pending their approval in each separate instance of data requests (Appendix J).

One of the most overarching measures that emerged as possessing crucial importance in every intermediary's operation was the **provision of added value through some form of data manipulation**. How the intermediaries provide added value and to which stakeholders were the most variable factors among all observed linking measures.

Some value propositions were more technical and focused and data processing, as seen in the case of the CRDC's and its practice of clearing and adding very granular spatial referencing to acquired datasets (CRDC-5). The UBDC followed a similar method of clearing and categorisation of data, although their datasets were not as geographically diverse in scope (Appendix M). Data anonymisation and processing were also presented as essential elements of the services offered by the DDI and Positium through their flagship software (DDI-5, POS-4). These observations are unsurprising, as they closely reflect the central importance of technical data treating capabilities in the process of data intermediation expressed in both literature on linking practices (Eckartz, Hofman & Veenstra, 2014; Flipsen, 2019; Van Schalkwyk et al., 2016 and conceptualisation) and conceptualisations of data intermediaries (Sein & Furuholt, 2012; Womack, 2002).

However, apart from such technical services, several other measures for value generation were utilised across cases. For example, both the CDRC and the UBDC stressed the contributions of their data analysis to their private data donors. The CDRC saw these benefits in the broad reach of the insights derived through their research on the basis of contributed data, which they perceive as mutually valuable for both policymakers and industry actors (Appendix G). On the other hand, the UBDC saw their contribution to their data donors in the validation of the credibility of their datasets as a robust evidence base that increased their value on the market (Appendix M). Furthermore, DjustConnect perceived their proposed value as being the easing of the administrative burden on their data contributors through facilitating documentation flow among actors (Appendix J).

Certainly, this is not the full extent of the measures practised by data intermediaries in the area of added value, as all activities carried out by them serve to generate benefits in some form. The above mentioned were just a few of the measures that were referenced explicitly as a means of deriving added value for stakeholders. Notably, these observations seem to reflect the idea of value generation as an essential element of data intermediation (Klievink, Bharosa & Tan, 2016; George, Yan & Leidner, 2020; Verhulst et al., 2019).

In terms of social capital, a linking measure witnessed across intermediaries and referenced as a critical enabler of data sharing relationships was that of **on the ground situational awareness**. The concrete measures employed to achieve such close familiarity with the needs of stakeholders encompassed several different approach groups. The first was the pursuit of **joint research projects** with entities from the public sector, allowing for the building of cross-sectoral working relationships. Such joint research projects were reported by the CDRC and the UBDC (Appendix F, UBDC-10). Joint research projects with non-governmental actors with a good understanding of local contexts on relevant issues were also observed in DDI (DDI-7). The second common measure was **reliance on local human resources**. In DDI, this took the form of on the ground presence through regional offices staffed with local experts with strong awareness of the particularities of the given environment (Appendix I). Limited by lack of resources for the setting up of subsidiaries, Positium's approach relied on forging connections with a few local agents, sometimes within partner structures, that supported them in logistical and liaison matters (Appendix K). These findings seem to reflect that of the research by Van Schalkwyk et al. (2016), that high levels of social capital expressed in well-developed networks serve to help data intermediaries be more attune to the needs of their targeted segments.

In addition to the regulative aspect of safeguarding data discussed in the previous section, the **technical aspect of data security insurance** also emerged as a prominent theme. A particular variety of approaches characterised measures in this area: The CDRC approached the security of its most sensitive data through the provision of three physical data facilities, with users required to travel to them in person to access data (CDRC-4). UBDC addressed security through virtual controlled access to sensitive datasets provided through their safe haven partner (Appendix M). DDI arranged access to the data by depositing servers at the data providers facilities and accessing it through a remote connection while leaving the security of malicious breaches of data the responsibility of data owners (DDI-4). Similarly, DjustConnect avoids the responsibility for data security through the strategic decision never to store data themselves (Appendix J).

6.3 Unique Measures

There was only a single intermediary in which a notable distinctive expression of economic capital was observed: the **cost-neutrality and subscription fee model** of DjustConnect. Due to their public sector affiliation, the business model of DjustConnect operates on a non-profit basis with the goal of full cost-neutrality (Appendix J). It aims to achieve this through an annual subscription fee for access to its data platform, with a stable rate regardless of the number of times a user utilises it to facilitate a data exchange

within a given period (ibid). DjustConnect was also the only intermediary that explicitly reported receiving **funding from private donor companies** (ibid).

There were three unique measures observed in the area of social capital. First was the principle of the **"ladder of engagement"** employed by the CDRC. This refers to its approach of using its Master's Dissertation Scheme as a means of establishing associations with private companies, and through an emphasis on mutual benefit and gradual deepening of relations, build trust up to the point of sharing of datasets (Appendix F). Notably, UBDC also runs a similar Master's Dissertation Scheme, however, they did not report utilising it as a linking measure.

The second measure is also employed by CDRC and comprises the **annual user forums**, which bring together the CDRC's existing and potential data partners to facilitate exchange of experiences (Appendix F). Notably, an interviewee from the CDRC was the only one to present the social activities engaged in by the intermediary as not only beneficial for the intermediary's outreach objectives but as an added value proposition to its partners, as a platform that enables them to network in their own right among people in the same industry (Appendix G). This approach is in accord with that expressed in literature by Susha & Gil-Garcia (2019) that the broadening of social networks is one of the principal incentives for data sharing that intermediaries can offer to their partners.

The third measure was the **modified "train the trainer" approach** adopted by DjustConnect. This measure entails the intermediary making initial connections with select individual data contributors and training them in the method of communicating the benefits and process of agricultural data sharing (Appendix J). These trained individuals would then conduct outreach on a peer-to-peer basis to other potential data contributors, benefiting from a level of trust derived from their shared backgrounds (ibid).

In the realm of cultural capital, DDI was the only data intermediary to make explicit reference to a focus on **the sustainability of the data solutions** and the private data derived insights (Appendix I). One of the interviewee's detailed several measures taken by DDI as a regular part of their partnership process for ensuring their public sector client's ability to sustain the beneficial results of the private data analysis in the long term (ibid). This was achieved through support in the setting up governance mechanism and guidance in the means of accessing additional funding (ibid).

An additional distinctive expression of cultural capital was the **data souring service** provided by UBDC. This linking measure entails the UBDC actively seeking out and procuring privately held datasets at the substantiated request of researchers of public sector entities (Appendix M). This measure is remarkable in that it facilitates access to

datasets other than those already in possession of the data intermediary, as in the cases of the CDRC and DjustConnect. Moreover, the UBDC approach also differs from that of DDI and Positium. Although the latter two may also negotiate access to datasets based on specific user needs, as outputs they deliver only resulting insights and not the underlining data itself. In opposition to this, UBDC facilitates direct data access.

A few special measures were observed in the area of symbolic capital. The first one was highlighted by one of the respondents from the CDRC, who identified the **intermediary's regular practice of staying up-to-data on and in constant anticipation of prevailing themes in governmental agendas** as a valuable means of continuously building new public sector partnerships (Appendix G). The second were the **calls for expressions of interest** disseminated by the UBDC upon acquisition of new datasets, a practice aimed at publicising new research opportunities to academic and public sector actors and raising awareness around the UBDC (UBDC-4).

Lastly, there were two notable technological measures reported by interviewees. Firstly, DjustConnect underlined the **possibility of developing build-on applications** capable of delivering different types of added value on top of their data sharing platform (Appendix J). Such applications are potentially capable of serving a multitude of different stakeholder needs (ibid). Secondly, though several intermediaries were observed as possessing different data analytics tools, only the CRDC reference to being in the process of developing **analysis-ready data tools specifically for the use of potential data users from the public sector**, that might otherwise be lacking the resources or expertise for taking advantage of the available private sector data (Appendix G).

7 Limitations and Future Research Recommendations

Although the given research provided comprehensive and valuable insights into the little-studied field of trusted data intermediary linking practices, it also suffered from several limitations.

Firstly, in terms of methodological limitations, it is important to note that the conclusion resulting from a cross-case analysis and its generalisability depends mainly on the heterogeneity of the selected cases (Greene & David, 1984). This is especially relevant in light of the choice to adopt a cross-case analysis approach. Indeed, pre-determined case selection criteria were closely followed to ensure heterogeneity and consistent logic in case selection. However, **it is possible that the selection of cases based on the more general grouping of trusted data intermediaries was not sufficiently specific to ensure a high degree of heterogeneity**. Although there is limited literature on alternative classifications which could have resulted in more narrow categories for case selection, two such alternatives, namely the typology of trusted data intermediary business models by Flipsen (2019) or the two subtypes of trusted data intermediaries proposed by Verhulst et al. (2019) were available. However, due to feasibility considerations, specifically the significant limiting of potential cases that would result from such narrowing of selection criteria and the complications that would carry on obtaining necessary case evidence in the limited time frame allotted for conducting this research, the criterion was kept to include all potential cases that meet the definition of trusted data intermediary as well as other criteria presented in Table 4.

Furthermore, in connection with the case selection criteria, a limitation occurred concerning the criterion of restricting the selected data intermediaries to those operating in Europe. Although all selected intermediaries were registered or based in Europe, through the research process it emerged that most of them also had partnerships or additionally conducted operations outside of Europe. This factor might limit the comparability of observed practices, particularly in connection with the concept of field within the theoretical framework presented in Chapter 3.

The choice of interviews as the primary source of data was also limited in several ways. Although its choice was justified by the nature of the research questions and lack of existing research in the area, the effect of this choice was that it potentially **limited the pool of respondents participating in the study** compared to what potentially could have been achieved had another data collection method been adopted. This limitation was exacerbated by a lack of resources and time constraints. The result was that for one case, it was only possible to interview one respondent, and for all other cases, the number of interviews was two.

Moreover, despite the interviews being conducted anonymously, it is still possible that the interviewees' responses were to a degree influenced by tacit company expectations and pressure regarding the external communication of the intermediary's activity. Furthermore, due to feasibility issues, **the intermediary's beneficiaries nor its data donors were able to be interviewed.** This presents a notable limitation, as all the observations of the attitudes and needs of data intermediary partners come from the possibly biased or limited perspective of the data intermediary and not the partners themselves.

In supplementary data gathering methods of online documentation analysis, documentation collection was conducted manually, making it possible that relevant information was excluded from observation due to oversight. Moreover, it is also possible that the selection of studied online documentation was influenced by a potential bias included in the search engine algorithms and the resulting queries since online documentation was retrieved through their use.

Lastly, and most notably, it should be emphasised that **the observed and presented linking practices of studied data intermediaries were not exhaustive**. They present only those key measures that were accented by the interviewees and online documentation on the intermediaries' activities. It is almost certain that data intermediaries employ further important linking measures in the process of facilitating a cross-sectoral flow of information that did not emerge within the bounds of this research.

For this reason, further inquiry into the field, on a systematised and more specific basis, possibly focusing only on one of the expressions of forms of capital as presented in Chapter 3, is recommended. Moreover, future research should also explore the applicability of other theoretical frameworks to the study of data intermediary linking practices that could offer an alternative classification of measures and a more in-depth analysis of data intermediary operations.

8 Conclusion

This research has given insights into the linking practices employed by trusted data intermediaries in the fulfilment of their goal of facilitating a flow of private data to actors capable of utilising them for greater societal benefit. The importance of such cross-sectoral data sharing has recently come to prominence in parallel with the rise in collaborative governance and data-driven innovation. Realising that access to the vast amount of real-time data routinely gathered by private businesses could significantly aid public administrations in tackling critical societal challenges, various initiatives have emerged aimed at harnessing the potential of private data for the public good.

Trusted data intermediaries, which this work defined as agents enabling cross-sectoral (and particularly private-public) data sharing, trust building and collaboration, through some form of data collection and processing aimed at the generation of added value for partners and redressal of public needs, are one of such initiatives. They represent a particularly promising form of cross-sectoral collaboration due to their ability to mitigate many barriers to data-sharing and provide added value to all involved stakeholders. However, as they are a comparatively recent phenomenon, there is a lack of literature and understanding around their operations. In particular, there has been little research into the measures that they rely on in the process of connecting actors and facilitating the flow of data between them.

To fill this existing gap in knowledge, this research set out to identify liking practices employed by trusted data intermediaries in the process of forging connections and exchange of data between private and public actors. To gain a more systematised understanding of these practices, it also resolved to classify the observed measures into different types of capital in accordance with a modified framework of Bourdieu's theory of field, habitus and capital. Lastly, in order to derive generalisable takeaways on trusted data intermediaries as a form of cross-sectoral data sharing, it analysed the observed practices across cases to identify how they concurred and differed.

The research objectives were addressed through the conducting of a multiple-case study of five different data intermediaries. The intermediaries differed in their policy area of activity, operation models and history of practice. Information on their linking practices was gathered through a combination of interviews with their representatives and secondary online document analysis.

The results of the individual examination of studied cases revealed that each intermediary employed a variety of different linking measures across all types of capital,

with practices in the areas of social, cultural and technological capital being the most numerous.

The cross-case analysis of observed linking practices presented the following interesting findings. First, access to public funding was observed as a crucially important factor in the operations of almost all intermediaries. Moreover, cultural capital, particularly in the forms of university and academic affiliation, emerged as a broadly influential field of measures in contrast to the lack of attention to these factors in studied literature. On the other hand, various expressions of social capital such as personal connections and networking associations were reaffirmed as being valuable linking mechanisms in accordance with findings of previous research.

The results also demonstrated a significant and compelling variation in approaches to the implementation of certain common measures across cases. Most prominently, such variations were observed in the processes of generation of added value through data manipulation or analysis and measures for the development of on the ground awareness of stakeholder needs.

Lastly, the research identified some notably unique linking measures employed by particular data intermediaries. Select measures discussed in this section included a subscription fee-based model of cost neutrality in the area of economic capital, three different methods of stakeholder relationship building and outreach in the field of social capital, and data solution sustainability and data souring services in the realm of cultural capital.

Through these findings, this study contributed to the little-researched field of trusted data intermediaries. However, due to shortage of previous research, this work was limited in several ways, described in detail within the work. Therefore, further inquiry into data intermediary linking practices, in particular on the basis of alternative frameworks of classification for more in-depth and specialised analysis of measures, was recommended.

References

Adams, C. W. (2015). Conducting Semi-Structured Interviews. In Wholey, J., Hatry, H., Newcomer, K. (ed.) *Handbook of Practical Program Evaluation* (Fourth Edition). 365-378. Jossey-Bass.

Albert, M. & Kleinman, L. D. (2011). Bringing Pierre Bourdieu to Science and Technology Studies. *Minerva* 49(3). 263-273.

Alemanno, A. (2020). *Towards a European Strategy on Business-To-Government Data Sharing for the Public Interest*. Final Report Prepared by the High-Level Expert Group on Business-to-Government Data Sharing European Commission. HEC Paris Research Paper No. LAW-2020-1394.

Atkinson, W. (2019). *Bourdieu and After: A Guide to Relational Phenomenology*. Routledge Publishing.

Bailey, J. P. & Bakos, J. Y. (1997). An Exploratory Study of the Emerging Role of Electronic Intermediaries. *International Journal of Electronic Commerce* 1. 7–20.

Barrett, D. B. & Martina, A. C. (2012). Towards a Non-deterministic Reading of Pierre Bourdieu: Habitus and Educational Change in Urban Schools. *Policy Futures in Education* 10(3). 278- 291.

Bhattacherjee, A. (2012). Social Science Research: Principles, Methods, and Practices. *USF Tampa Bay Open Access Textbooks Collection*. Book 3.

Bourdieu, P. (1986). The Forms of Capital. In J. Richardson (ed.) *Handbook of Theory and Research for the Sociology of Education*. 241-258. New York: Greenwood.

Bourdieu, P. & Wacquant, L.J.D. (1992). *An Invitation to Reflexive Sociology*. University of Chicago Press.

Burt, R. (1999). Entrepreneurs, Distrust, and Third Parties: A Strategic Look at the Dark Side of Dense Networks in. In Levine, M., Messick, D., Thompson, L. (ed.) *Shared Cognition in Organizations: Management of Knowledge*. 213-245. Psychology Press.

Burt, R. (2001). Structure Holes Versus Network Closure as Social Capital. In Lin, N., Cook, K, and Burt, R. (ed.) *Social Capital: Theory and Research*. 31-57. Routledge.

Crain, M. (2018). The Limits of Transparency: Data Brokers and Commodification. *New Media and Society* 20(1). 88-104.

Data Collaboratives Explorer. The GovLab. Retrieved from: https://datacollaboratives.org/explorer.html

Davidsson, P. & Honig, B. (2003). The Role of Social and Human Capital Among Nascent Entrepreneurs. *Journal of Business Venturing* 18. 301-31.

De Clercq, D. & Voronov, M. (2009). The Role of Cultural and Symbolic Capital in Entrepreneurs' Ability to Meet Expectations about Conformity and Innovation. *Journal of Small Business Management* 47(3). 398-420.

Denzin, N. K. (1978). *The Research Act: A Theoretical Introduction to Sociological Methods*. New York, NY: McGraw-Hill.

Dobbin, F. (2008). The Poverty of Organizational Theory: Comment on 'Bourdieu and Organizational Analysis'. *Theory and Society* 37(1). 53-63

Dube, L. & Pare, G. (2003). Rigor in Information Systems Positivist Case Research: Current Practices, Trends, and Recommendations. *MIS Quarterly* 27(4). 597-636.

Eckartz, S., Hofman, W. & Veenstra, A. (2014). *A Decision Model for Data Sharing*. 13th International Conference on Electronic Government (EGOV). September 2014, Dublin, Ireland. 253-264.

Edgerton, D. J. & Roberts, W. L. (2014). Cultural Capital or Habitus? Bourdieu and Beyond in the Explanation of Enduring Educational Inequality. *Theory and Research in Education* 12(2). 193-22.

Eisenhardt, K. (1989). Building Theories from Case Study Research. *The Academy of Management Review* 14(4). 532-550.

Emirbayer, M. & Johnson, V. (2008). Bourdieu and Organizational Analysis. *Theory and Society* 37(1). 1-44.

Emmison, M. & Frow, J. (1998). Information Technology as Cultural Capital. *Australian Universities' Review* 41(1). 41-45.

European Commission. (2018a*). Towards a Common European Data Space*. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. No. COM (2018) 232 final. Brussels: European Commission, April 15.

European Commission. (2018b). *Guidance on Sharing Private Sector Data in the European Data Economy*. Staff Working Document. No. SWD (2018) 125 final. Brussels: European Commission, April 15.

Ferri, J. P., Deakins, D. & Whittam, G. (2009). The Measurement of Social Capital in the Entrepreneurial Context. *Journal of Enterprising Communities* 3(2). 138-151.

Flipsen, M. (2019). *Data Collaboratives: Trusted Data Intermediary Business Models* (Master's Thesis, Delft University of Technology).

George, J., Yan, J. & Leidner, E. D. (2020). Data Philanthropy: Corporate Responsibility with Strategic Value? *Information Systems Management* 37(3). 186-197.

Gil-Garcia, J. R., Chengalur-Smith, I. S. & Duchess, P. (2007). Collaborative E-Government: Impediments and Benefits of Information-Sharing Projects in the Public Sector. *European Journal of Information Systems* 16(2). 121-133.

Greene, D. & David, J. (1984). A Research Design for Generalizing from Multiple Case Studies. *Evaluation and Program Planning* 7(1). 73-85.

Greenspan, I. (2014). How Can Bourdieu's Theory of Capital Benefit the Understanding of Advocacy NGOs: Theoretical Framework and Empirical Illustration. *Nonprofit and Voluntary Sector Quarterly* 43(1). 99-120.

Groves, M. R. & Neufeldi, A. (2017). *Accelerating the Sharing of Data Across Sectors to Advance the Common Good*. Beeck Center for Social Impact + Innovation at Georgetown University. Retrieved from: https://bit.ly/3pNmx4N

Hagel III, J. & Rayport, J. F. (1997). The Coming Battle for Customer Information. *Harvard Business Review* 71. 53–55.

Hallett, T. & Gougherty, M. (2018). Bourdieu and Organizations: Hidden Traces, Macro Influence, and Micro Potential. In Medvetz, T. & Sallaz, J. (ed.) *Oxford Handbook of Pierre Bourdieu.* 273-298.

Harker, R., Mahar, C. & Wilkes, C. (1990). *An Introduction to the Work of Pierre Bourdieu the Practice of Theory*. Palgrave Macmillan.

Harper, R. (2002). *The Measurement of Social Capital in the United Kingdom*. ONS Report, Office for National Statistics, London, September.

Head, W. B. & Alford, J. (2015). Wicked Problems. *Administration & Society* 47(6). 711-739.

Henke, N., Bughin, J., Chui, M., Manyika, J., Saleh, T., Wiseman, B. & Sethupathy, G. (2016). *The Age of Analytics: Competing in a Data-Driven World*. McKinsey Global Institute. Retrieved from: https://mck.co/3cJTlWP

Hilgers, M. & Mangez, E. (2015). *Bourdieu's Theory of Social Fields Concepts and Applications*. Routledge Publishing.

Huang, X. (2019). Understanding Bourdieu - Cultural Capital and Habitus. *Review of European Studies* 11(3). 45-49.

Ignatow, G. & Robinson, L. (2017). Pierre Bourdieu: Theorizing the Digital. *Information, Communication & Society* 20(7). 950-966.

Jacob, S. A. & Furgerson, S. P. (2012). Writing Interview Protocols and Conducting Interviews: Tips for Students New to the Field of Qualitative Research. *The Qualitative Report 17*(42). 1-10.

Janssen, M. & Zuiderwijk, A. (2014). Infomediary Business Models for Connecting Open Data Providers and Users. *Social Science Computer Review* 32(5). 694-711.

Kern, G. F. (2018). The Trials and Tribulations of Applied Triangulation: Weighing Different Data Sources. *Journal of Mixed Methods Research* 12(2). 166-181.

Khanchel, H. & Kahla, B. K. (2013). Mobilizing Bourdieu's Theory in Organizational Research. *Review of General* Management 17(1). 86-94.

Klievink, B., Janssen, M. & Tan, Y.H. (2012). A Stakeholder Analysis of Business-to-Government Information Sharing: The Governance of a Public-Private Platform. *International Journal of Electronic Government Research* 8(4). 54-69.

Klievink, B., Bharosa, N. & Tan, Y. H. (2016). The Collaborative Realization of Public Values and Business Goals: Governance and Infrastructure of Public-Private Information Platforms. *Government Information Quarterly* 33(1). 67–79.

Klievink, A.J., Van der Voort, H.G. & Veeneman, W.W. (2018). Creating Value Through Data Collaboratives: Balancing Innovation and Control. *Information Polity* 23(4). 379-397.

Leuffen, D., Shikano, S. & Walter, S. (2013). Measurement and Data Aggregation in Small-n Social Scientific Research. *European Political Science 12*(1), 40-51.

Magalhaes, G., Roseira, C. & Strover, S. (2013). Open Government Data Intermediaries. *Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance*. October 2013, Seoul, Republic of Korea. 330-333.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. & Byers, A. H. (2011). *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute. Retrieved from: https://mck.co/2TvC4tN

Martens, B. & Duch-Brown, N. (2020). *The Economics of Business-to-Government Data Sharing*. JRC Technical Report, JRC Digital Economy Working Paper 2020-4.

Neveu, E. (2018). Bourdieu's Capital(s): Sociologizing an Economic Concept. In Medvetz, T. & Sallaz, J. (ed.) *Oxford Handbook of Pierre Bourdieu*. 347-374.

Noordegraaf, M. & Schinkel, W. (2011). Professionalism as Symbolic Capital: Materials for a Bourdieusian Theory of Professionalism. *Comparative Sociology* 10(1). 67-96.

Organisation for Economic Co-operation and Development (2013). *Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value*. OECD Digital Economy Papers No. 220. OECD Publishing, Paris. Retrieved from: https://bit.ly/3zoY0XZ

Prell, C. (2006). Social Capital as Network Capital: Looking at the Role of Social Networks Among Not-For-Profits. *Sociological Research Online* 11(4). 1-14.

Pret, T., Shaw, E. & Drakopoulou Dodd, S. (2016). Painting the Full Picture: The Conversion of Economic, Cultural, Social and Symbolic Capital. *International Small Business Journal* 34(8). 1004-1027.

Prieur, A. & Savage, M. (2011). Updating Cultural Capital Theory: A Discussion Based on Studies in Denmark and in Britain. *Poetics* 39(6). 566-580.

Ragnedda, M. (2018). Conceptualizing Digital Capital. *Telematics and Informatics* 35(8). 2366-2375.

Rieke, A., Yu, H., Robinson, D., & Van Hoboken, J. (2016). *Data Brokers in an Open Society*. Open Society Foundation. London. Retrieved from: https://osf.to/2RPHvTU

Robin, N., Klein, T. & Jütting, J. (2016). *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*. OECD Development Co-operation Working Papers No. 27. OECD Publishing, Paris.

Romele, A. (2020). Technological Capital: Bourdieu, Postphenomenology, and the Philosophy of Technology Beyond the Empirical Turn. *Philosophy & Technology*. 1-23.

Saldaña, J. (2013). *The Coding Manual for Qualitative Researchers* (Second Edition.). SAGE Publications.

Schell, C. (1992). *The Value of the Case Study as a Research Strategy*. Manchester Business School. Retrieved from: https://bit.ly/3h5mOfp

Sein, M. & Furuholt, B. (2012). Intermediaries: Bridges Across the Digital Divide. *Information Technology for Development* 18(4). 332-344.

Selsky, J. & Parker, B. (2005). Cross-Sector Partnerships to Address Social Issues: Challenges to Theory and Practice. *Journal of Management* 31(6). 849–873.

Sorrentino, M. & Niehaves, B. (2010). Intermediaries in E-inclusion: A Literature Review. In *Proceedings of the 43rd Hawaii International Conference on System Sciences*. January 2010, Honolulu, USA. 1-10.

Susha, I., Janssen, M. & Verhulst, S. (2017a). 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*. January 2017, Waikoloa Village, USA. 2691-2700.

Susha, I., Janssen, M. & Verhulst, S. (2017b). Data Collaboratives as "Bazaars"?. *Transforming Government* 11(1). 157-172.

Susha, I., Pardo, T., Janssen, M., Adler, N., Verhulst, S. et al. (2018). A Research Roadmap to Advance Data Collaboratives Practice as a Novel Research Direction. *International Journal of Electronic Government Research* 14(3). 1-11.

Susha, I. & Gil-Garcia, J. R. (2019). A Collaborative Governance Approach to Partnerships Addressing Public Problems with Private Data. *Proceedings of the 52nd Hawaii International Conference on System Sciences*. January 2019, Grand Wailea, USA. 2892-2901.

Susha, I., Grönlund, Å., & Van Tulder, R. (2019). Data Driven Social Partnerships: Exploring an Emergent Trend in Search of Research Challenges and Questions. *Government Information Quarterly* 36(1). 112-128.

Susha, I., Rukanova, D. B., Gil-Garcia, J.R., Hua Tan, Y., & Gasco, M. (2019). Identifying Mechanisms for Achieving Voluntary Data Sharing in Cross-Sector Partnerships for Public Good. *Proceedings of the 20th Annual International Conference on Digital Government Research*. June 2019, Dubai, United Arab Emirates. 227-236.

Susha, I. (2020). Establishing and Implementing Data Collaborations for Public Good: A Critical Factor Analysis to Scale Up the Practice. *Information Polity* 25(1). 3-24.

Tsai, W. (2000). Social Capital, Strategic Relatedness, and the Formation of Intraorganisational Linkages. *Strategic Management Journal* 21. 925–939.

Van Schalkwyk, F., Cañares, M., Chattapadhyay, S. & Andrason, A. (2016). Open Data Intermediaries in Developing Countries. *Journal of Community Informatics* 12(2), Special Issue on Open Data for Social Change and Sustainable Development. 9-25.

Van Schalkwyk, F., Willmers, M. & McNaughton, M. (2016). Viscous Open Data: The Roles of Intermediaries in an Open Data Ecosystem. *Information Technology for Development* 22. 68–83.

Vaughan, D. (2008). Bourdieu and Organizations: The Empirical Challenge. *Theory and Society* 37(1). 65–81.

Verhulst, S. & Sangokoya, D. (2014). *Mapping the Next Frontier of Open Data: Corporate Data Sharing*. Report for Internet Monitor 2014: Reflections on the Digital World. Retrieved from: https://bit.ly/2RPHxeu

Verhulst, S. & Sangokoya, D. (2015). *Data Collaboratives: Exchanging Data to Improve People's Lives*. GovLab. Retrieved from: https://bit.ly/3pTCyWT

Verhulst, G. S., Susha. I, & Kostura, A. (2016). *Data Collaboratives: Matching Demand with Supply of (Corporate) Data to solve Public Problems*. Retrieved from: https://bit.ly/3gs7Ila

Verhulst, S., Young, A., Winowatan, M. & Zahuranec, J. A. (2019). *Leveraging Private Data for Public Good: A Descriptive Analysis and Typology of Existing Practices*. The GovLab. Retrieved from: https://bit.ly/2TsuEay

Waldstrøm, C. (2003). *Social Capital in Organizations – Beyond Structure and Metaphor*. Working Papers 7. University of Aarhus, Aarhus School of Business, Department of Management.

Webb, J., Schirato, T. & Danaher, G. (2002). *Understanding Bourdieu*. Allen & Unwin.

Winowatan, M., Young, A. & Verhulst, G. S. (2019). *A Data Collaborative Case Study – Accelerating Medicines Partnerships: Improving Drug Research Efficiency through Biomarker Data Sharing*. TheGovLab. Retrieved from: https://bit.ly/3cDvrwq

Winowatan, M., Young, A. & Verhulst, G. S. (2020b). *A Data Collaborative Case Study – Global Fishing Watch: Pooling Data and Expertise to Combat Illegal Fishing*. The GovLab. Retrieved from: https://bit.ly/3xkAwS8

Winowatan, M., Zahuranec, J. A., Young, A. & Verhulst, G. S. (2020). *A Data Collaborative Case Study - Leveraging Telecom Data to Aid Humanitarian Efforts: Lessons Learned from the 2015 Earthquake in Nepal*. TheGovLab. Retrieved from: https://bit.ly/3pPTJZk

Winowatan,M., Young, A. & Verhulst, G. S. (2020a). *A Data Collaborative Case Study - the Atlas of Inequality and Cuebiq's Data for Good Initiative: Using Location Data to Quantify Segregation of Urban Social Spaces*. TheGovLab. Retrieved from: https://bit.ly/3whGhzM

Womack, R. (2002). Information Intermediaries and Optimal Information Distribution. *Library and Information Science Research* 24(2). 29–155.

Yazan, B. (2015). Three Approaches to Case Study Methods in Education: Yin, Merriam and Stake. *The Qualitative Report* 20(2). 134–152.

Yin, R. (2003a). *Case Study Research: Design and Methods* (Third Edition, Applied Social Research Methods Series 5). Thousand Oaks: Sage.

Yin, R. (2003b). Applications of Case Study Research (Second Edition, Applied Social Research Methods Series 34). Thousand Oaks: Sage.

Yoon, A., Copeland, A. & McNally, J. P. (2018). 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.

Zaidah, Z. (2007). Case Study as a Research Method. *Jurnal Kemanusiaan* 9. 1-6.

Zang, W. (2010). Technical Capital and Participatory Inequality in E-deliberation: An Actor–Network Analysis. *Information, Communication & Society* 13(7). 1019- 1039.

Appendix

A Overview of papers analysed in the literature review

Keyword	Number of Articles	Results
Data Collaborative	15	Susha, Janssen & Verhulst, 2017a; Susha, Janssen, & Verhulst, 2017b; Susha et al, 2018; Susha, 2020; Klievink, Van der Voort & Veeneman, 2018; Velhurst el al., 2019; Eckartz, Hofman & Veenstra, 2014; Flipsen, 2019; TheGovLab, 2017; Susha, & Gil-Garcia, 2019; Susha, Grönlund & Van Tulder; 2019; Winowatan, Young, & Verhulst, 2019, 2020a, 2020b; Winowatan et al. 2020.
Trusted Data Intermediary	1	Flipsen, 2019.
Data Intermediary	7	Yoon, Copeland, & McNally, 2018; Van Schalkwyk et al, 2016, Van Schalkwyk, Willmers & McNaughton, 2016; Magalhaes, Roseira & Strover, 2013; Sein & Furuholt; 2012; Sorrentino & Niehaves, 2010;
Infomediary	4	Janssen & Zuiderwijk, 2014; Magalhaes, Roseira & Strover, 2013; Bailey & Bakos, 1997; Womack; 2002;
Business-to-Government Data Sharing	4	Klievink, Janssen & Tan, 2012; Robin, Klein & Jütting, 2016; Susha et al, 2019; Groves & Neufeldi; 2017.
Data Philanthropy	3	Verhulst & Sangokoya, 2014; Susha, Grönlund & Van Tulder, 2019; George, Yan & Leidner, 2020;
Data Broker	3	Rieke et al. 2016; Crain, 2018; Susha, Grönlund & Van Tulder, 2019.

B Exemplars of Different Types of Capital

Economic Capital	• Possession of income, savings, intellectual property and tangible business assets (Pret, Shaw & Drakopoulou Dodd, 2016); • Possession of assets capable of being immediately or directly converted into money or of being institutionalised in the form of property rights (Bourdieu, 1986); • Possession of the general elements of production and could be things such as land, factories, jobs and inheritance (Huang, 2019); • Possession of assets usable for the hiring of human resources and expertise (Albert & Kleinman, 2011); • Availability of means to pay organisations for access to the desired data (Van Schalkwyk et al., 2016).
Cultural Capital	• Holding valued dispositions, cultural goods, skills and education (Pret, Shaw & Drakopoulou Dodd, 2016); • Having extensive knowledge and mastery of information, concepts, trends, theories and history in the relevant field (Atkinson, 2019); • Holding high academic qualifications or affiliations (Bourdieu, 1986); • Holding leading positions in competitions (Bourdieu, 1986); • Possessing habits or skills which facilitate and simplify the continued acquisition of education and knowledge (Prieur & Savage, 2011); • Holding of certifications (Neveu, 2018).
Social Capital	• Having memberships in societies, relations, networks and alliances (Pret, Shaw & Drakopoulou Dodd, 2016). • Having an extensive network of connections which one is capable of mobilising for their assistance and benefit (Bourdieu, 1986). • Having a network of connections made up of agents who in turn themselves possess high levels of capital (Bourdieu, 1986). • Having interconnections with other companies (Atkinson, 2019). • Having the right connections that allow for the facilitation of collaboration with reputable experts (Albert & Kleinman, 2011). • Having the image of trustworthiness, resulting in other agents being more open to cooperation, less formalised cooperation and reciprocity in resource sharing (Prell, 2006). • Possessing privileged access to knowledge resources by virtue of network connections (Ferri, Deakins & Whittam, 2009). • Having network mediated access to individuals capable of playing an influential role in the agent's success, for example, business advisors, venture capitalists, customers or suppliers (Ferri, Deakins & Whittam, 2009). • Having access to third-party organisations that provide support and funding through the virtue of the network (Ferri, Deakins & Whittam, 2009) • Being well informed about relevant local, organisational, civil, club activities and possessing an ability to influence them through the virtue of being part of a network (Harper, 2002).

	• Having the capability to speed up the time frame for certain processes/activities by calling on personal connections (Huang, 2019). • Membership in professional unions or guilds (Pret, Shaw & Drakopoulou Dodd, 2016).
Symbolic Capital	• Holding of awards, trophies, publicity, reputation, prestige (Pret, Shaw & Drakopoulou Dodd, 2016); • Being the subject of advertisements and publicity in reputable sources (Pret, Shaw & Drakopoulou Dodd, 2016); • Having membership of an esteemed group or organisation (Atkinson, 2019); • Having a reputation for being a dominant actor in a particular field (Atkinson, 2019); • Possessing a status of professionalism (professional knowledge, skill, experience and ethics) (Noordegraaf & Schinkel, 2011); • Having a reputation for conducting high-quality research work (Van Schalkwyk et al., 2016); • Having the ability to impose definitions of phenomena on other field participants, establish new field rules and dictate best new practices (De Clercq & Voronov, 2009); • Having a stand out position in comparison to other actors within a field (De Clercq & Voronov, 2009); • Having a reputation of being actively involved in charity and philanthropy (Pret, Shaw & Drakopoulou Dodd, 2016); • Having a long history of industry experience (Pret, Shaw & Drakopoulou Dodd, 2016); • Possessing customer or client loyalty despite the availability of alternatives on the market (Pret, Shaw & Drakopoulou Dodd, 2016).
Technical Capital	• Possession of technologies, particularly those designed for or those that can be used for an entity's particular processes and objectives (Romele, 2020). • Having authorisation for the use of particular technological artefacts in ways that is not available to other actors (Romele, 2020). • Possessing knowledge and skill in regards to digital information, communication, safety and problem-solving (Ragnedda, 2018). • Possessing the abilities of data extraction, reorganisation, anonymisation, validation, treatment, formatting and processing (Van Schalkwyk et al., 2016). • Having the knowledge and skill to facilitate the data publication in useful formats (Van Schalkwyk et al., 2016). • Having the knowledge and skill to develop applications for interpreting large datasets (Van Schalkwyk et al., 2016).

C Interview Questions

Dimension	Interview Questions
Introductory Questions	1. Could you briefly describe your role and responsibilities in your organisation? 2. Could you provide some background information on your organisation? (e.g.how did stakeholders first come together to develop the common interest of its establishment? Is it affiliated with a parent organisation?)
Circle of Partners	3. Is the list of your organisation's private sector data contributors static or continuously expanding? If the latter is the case, what activities does your organisation conduct to outreach to new data contributors? 4. Is the list of your organisation's public sector partners static or continuously expanding? If the latter is the case, what activities does your organisation conduct to outreach to new public sector users? 5. What are the main methods employed by your organisation for finding potential data contributors and/or public sector partners that could provide valuable data or benefit from the data you already have access to?
Collaboration Forming	6. Could you give a walk-through of a generalized process of how a collaboration comes to be? How do these processes differ among private sector and public sector partners? 7. What factors do you believe constitute a major draw for collaboration with your organisation for your partners? 8. What mechanisms does your organisation employ to communicate the potential value of data sharing and the services you provide to potential partners? Is there any aspect of this process that proves to be particularly challenging in practice? 9. Do you believe that a personal-professional connection contributes to the establishment of new partnerships? / Do you believe that word-of-mouth contributes in any significant way to the expansion of your network? If so, how? 10. Do you find that financial resources meaningfully enable or restrict partnership establishment efforts? If so, how? 11. Do you find that strong symbolic or societal importance of targeted issues meaningfully enable or restrict partnership establishment efforts?

	12. In your experience, what are the most common apprehensions (if any) expressed by stakeholders in the course of forming a collaboration? What approaches do you take to mitigate any such concerns?
Concluding Questions	13. Are there any other aspects that you believe to be relevant to your stakeholder linking process that I failed to ask about or we have not yet addressed that you could elaborate on? 14. Could you put me in touch with one of your colleagues working on similar issues that you believe would be open to also speaking with me?

D The Gathered Online Documentation

Case	Online Document	Reference Code
The Consumer Data Research Centre (CDRC)	1. About the CDRC, cdrc.ac.uk/about, Accessed on June 1st 2021	CDRC-1
	2. Big Data Network Phase 2, esrc.ukri.org/research/our-research/big-data-network/big-data-network-phase-2, Accessed June 20th 2021	CDRC-2
	3.Power Point Presentation, hummedia.manchester.ac.uk/ institutes/cmist/ researchdocs/fleming.pdf, Accessed June 20th 2021	CDRC-3
	4. CDRC Data Service Guide Version 6, data.cdrc.ac.uk/sites /default/files/CDRC-Data-Service-User-Guide-V6.pdf, Accessed June 1st 2021	CDRC-4
	5. Power Point Presentation, dpconline.org/docs/miscellaneous/ events/ 2016-events/1489-transdata-presentation-sheppard/file, Accessed June 20th 2021	CDRC-5
	6. Information for Businesses, cdrc.ac.uk/education-and-training/masters-dissertation-scheme/information-for-businesses, Accessed June 20th 2021	CDRC-6
	7. Education and Training, cdrc.ac.uk/education-and-training, Accessed June 1st 2021	CDRC-7
	8. Research, cdrc.ac.uk/research, Accessed June 1st 2021	CDRC-8
Dalberg Data Insights (DDI)	1. About, dalberg.com/what-we-do/dalberg-data-insights, Accessed June 26th 2021	DDI-1
	2. Riaktrs' Data-for-Good Joins Dalberg, riaktr.com/real-impact-analytics-data-good-joins-dalberg, Accessed June 26th 2021	DDI-2
	3. YouTube Video: Dalberg Data Insights: Conscious Data for a Better World, youtube.com/watch?v=B0yyMBI264s&t=77s, Accessed June 26th 2021	DDI-3
	4. YouTube Video: Q&A with Christophe Bocquet, youtube.com/watch?v= aYTW-ApF7iE, Accessed June 26th 2021	DDI-4
	5. AIDA, aida.dalbergdatainsights.com, Accessed June 26th 2021	DDI-5
	6. YouTube Video: Presentation Dalberg Data Insights by Lucio Melito, youtube.com/watch?v=qrHQXynKEx0, Accessed June 26th 2021	DDI-6

	7. Food Security Manager, dalberg.com/our-ideas/dalberg-data-insights-identifies-areas-risk-food-insecurity-using-mobile-phone-data, Accessed June 26th 2021	DDI-7
	8. Dalberg Data Insights Maps Urban Mobility Across Haiti and Around Kampala, dalberg.com/our-ideas/dalberg-data-insights-maps-urban-mobility-across-haiti-and-around-kampala-uganda, Accessed June 26th 2021	DDI-8
DjustConnect	1. Launch of 'DjustConnect', ilvo.vlaanderen.be/en/news/lancering-djustconnect-draaischijf-voor-het-delen-van-data-in-de-agrovoedingsketen, Accessed July 15th 2021	DJU-1
	2. DjustConnect homepage, djustconnect.be/en, Accessed July 15th 2021	DJU-2
	3. YouTube Video, DjustConnect English Introduction, youtube.com/ Watch?v=bB7ViJVgdiQ, Accessed July 15th 2021	DJU- 3
	4. DjsutConnect 'Who Are We?", djustconnect.be/en/who-are-we, Accessed July 15th 2021	DJU-4
Positium	1. About Positium, positium.com/en, Accessed July 12th 2021	POS-1
	2. YouTube Video, Erki Saluveer Presenting Positium at ICT2015 Lisbon, youtube.com/watch?v=RhLvJkUQzHU, Accessed July 12th 2021	POS-2
	3. PowerPoint Presentation. Mobile Big Data for Human Mobility and Official Statistics.	POS-3
	4. Positium Data Mediator, positium.com/research/positium-data-mediator, Accessed July 12th 2021	POS-4
	5. Factsheet of the Estonian Case Study of BuildERS Project, buildersproject. eu/assets/content/FactSheetsBuildERS(EN).pdf, Accessed July 12th 2021	POS-5

Urban Big Data Center (UBDC)	1. About UBDC, ubdc.ac.uk/about-ubdc/our-work, Accessed June 12th 2021	UBDC-1
	2. Information for Data Owners, ubdc.ac.uk/data-services/data-services/ information-for-data-owners, Accessed June 12th 2021	UBDC-2
	3. Data Catalogue, ubdc.ac.uk/data-services/data-catalogue, Accessed June 12th 2021	UBDC-3
	4. Call for Expressions of Interest, ubdc.ac.uk/data-services/data-services/ call-for-expressions-of-interest, Accessed June 12th 2021	UBDC-4
	5. Research Projects, ubdc.ac.uk/research/research-projects, Accessed June 12th 2021	UBDC-5
	6. MS Programs, ubdc.ac.uk/education-and-events/education/msc programmes, Accessed June 12th 2021	UBCD-6
	7. Online Tutorial, ubdc.ac.uk/education-and-events/online-learning/online-tutorials, Accessed June 12th 2021	UBDC-7
	8. Events and Courses, ubdc.ac.uk/education-and-events/events-and-courses/upcoming-events, Accessed June 12th 2021	UBDC-8
	9. YouTube Video: Welcome and Introduction to UBDC, youtube.com/ watch?v=GT6gIvTD99s, Accessed June 15th 2021	UBDC-9
	10. YouTube Video: Urban Big Data Centre and Glasgow City Council's CCTV Object Detection Project, youtube.com/watch?v=7uuQzyhFDdM, Accessed June 15th 2021	UBDC-10
	11. YouTube Video: Building Experiences with Government Data Owners, youtube.com/watch?v=kusGRI0x62s, Accessed June 15th 2021	UBDC-11

E Data Coding Definitions

Code/Subcode	ID	Definition
Economic Capital	EC	References to a data intermediary's financial assets and resources.
Operational Budget	EC-OB	References to funds that enable the day-to-day operation of the data intermediary.
Funds for Data Acquisition	EC-FDA	References to funds that enable intermediaries to reimburse potential data contributors for their data.
Social Capital	SC	References to a data intermediary's network relations and any of its resulting effects.
Professional Unions	SC-PU	References to membership/participation in stable professional unions with other entities in the field of data exchange.
Personal Connections	SC-PC	References to personal professional and other connections among representatives of the data intermediary and other organisation representatives.
Networking Events	SC-NE	References to organisation or participation in networking events in the field of data exchange.
Partnerships	SC-P	References to organisations with which the data intermediary enjoys professional partnerships and/or good working relationships.
Trust	SC-T	References to the existence of trust between a data intermediary's and other organisation's working relationship.
Privileged Access	SC-PA	References to a data intermediary's privileged access to information, funding or decision makers by the virtue of social connections.
Cultural Capital	CC	References to a data intermediary's knowledge and academics related activities and resources.
Human Recourses	CC-HR	Reference to a data intermediary's possession of skilled/ academically highly qualified human resources.
Educational Activities	CC-EA	Reference to the educational activities with a data intermediary organises or takes part in.

University Affiliation	CC-UA	References to a data intermediary's affiliation/partnership with a higher educational institution.
Experience-Based Knowledge	CC-EBK	Reference to a data intermediary's possession of knowledge derived from personal experience of operating the data exchange field not available to other actors.
Symbolic Capital	SYC	Reference to the public's awareness and perception of a data intermediary's activities.
Accolades	SYC-A	Reference to any awards, honours or acknowledgements held by a data intermediary.
Societal Resonance	SYC-SR	References to the societal/political resonance or actuality of a data intermediary's activities.
Reputation	SYC-R	Reference to a data intermediary's reputation among other entities.
Media Coverage	SYC-MC	References to both outside media and social media coverage of a data intermediary's activities.
Technological Capital	TC	References to a data intermediary's technological capabilities and resources.
Hardware	TC-H	References to a data intermediary's possession of hardware enabling data processing, storage or insight delivery.
Data Storage	TC-DS	References to a data intermediary's abilities of securely storing data.
Data Processing	TC-DP	References to a data intermediary's abilities of data extraction, reorganization, anonymization, validation, treatment and formatting.
Insight Delivery	TC-ID	Referenced to a data intermediary's ability to deliver insight from different methods of data processing.

F Interview with CDRC Project Delivery Manager

Interviewer: Could I ask you to briefly, for the record, describe your role and responsibilities for the CDRC?

Respondent: I work as Project Delivery Manager for the CDRC, the Consumer Data Research Centre. My role revolves around a number of different things, the main aspect is user applications. So, these people, researchers, or people who want to come and use the datasets - I manage the process to give them access to the confidential data. I also used to have responsibility for data partner management, these are the organizations that provide the data. My role has changed slightly, I am still a bit involved with that, but most of that is now being done by my colleague, Oliver. But that's only very recently, so I'm happy to share my experiences. I'm also kind of responsible for some of the project administrations in terms of the paperwork and the ethical requirements as well, either for external projects or internal projects that staff working for the CDRC put together.

Interviewer: Could I ask you to provide me with some background information to the extent of your knowledge about the origins of the CDRC? I am aware that you were established through the founding of the Economic and Social Research Council within this Big Data Networking Initiative, but do you have maybe any insights on how, for example, stakeholder came around the issue of consumer data, how the first data contributors came on board? How did the connections and stakeholders come together in the very first place?

Respondent: I wasn't actually involved back 2014 when it first came together. But my understanding is, there was a lot of discussion around big data at that point, in terms of, at that point it was more partly a technical challenge of how do we work with these massive datasets. But, from the CDRC's point of view it is very much, you know, there is this data out there, usually with private companies, and often they use their data for certain aspects of their business. But there's so much more we could do with that data. But there is no infrastructure to manage this process. And then the companies don't have the time or resources for individual researchers to come along and say, oh, I've got this really interesting project, it'd be great, can I access your data to do this amazing research. These companies don't have instructions to do that on a case-by-case basis. But if we act as a gatekeeper, we can offer your data for academic research, and this is how we do it, and we've got various procedures and so on. And the company is involved in the "can the research be done?" phase, so they can say yes or no to whether a piece of research is done. But it gives them a structure to work with, it gives them the resources to be able to have their data used for world changing research.

Interviewer: So, would it be correct to say that your Centre is kind of a practical embodiment or a continuation of relevant political discourse at the time and that there was quite a big political will around data driven policymaking and analysis that started back then?

Respondent: Yes, yes, I think that's a reasonable thing to say.

Interviewer: Let's delve into questions that deal more with your previous role managing partnerships with data contributors. According to a list that I recovered you work with around 30 private data contributes. Is that list more or less static and well established, or are you continuously trying to expand and outreach to possible new data contributors from the private sector?

Respondent: It's continually growing, you know, we're always interested in new collaborations. And we have to maintain the existing ones as well, because most of the data are snapshots that we have. So, it's really useful to get updates of data at points throughout the project and to get more up to date data as more data sources becoming available as well. We've recently been involved with the local data company looking at footfall and we've got a quite big project with them, to help them in installing floor monitors, so devices that will measure the quantity of people have arrived to different places, and we we've got a process where we collect that data and aggregate it and make it available as a research usable product. Yes.

Interviewer: Could you go into more detail about how this process of acquiring new data donors work? Do they contact you or is it more that you take active measures and outreach to them? Do factors such as your reputation from being established for quite some time now play a factor or this process in your opinion?

Respondent: Most of the links are through, you know, networking relationships from our PR who know a lot of people in the area that we work in. We do quite a lot of promotional work about what the CDRC does, we have the blog, we do Twitter and all the rest of it - just saying this is stuff people have done with the data. We also have what we call a ladder of engagement. This is an approach to develop relationships with organizations, or start them and then increase them over time. So, for example, we might start an organization off with a Master's dissertation in that - we have the master's dissertation scheme where we pair master students who want to do a dissertation with companies who've got some interesting data. And so, it's like a three or four-month project for the company, they get a master's students for that period to do some interesting data analysis that they may not have time or resources to do. The master's student gets a great dissertation and gets access to data that they probably wouldn't have otherwise. So, you know, we might start with that. And then if that goes well, we might say, okay would you would you like to have a PhD student work with you? That is a three-year project, much more in detail and they will put some financial resources into that, they might fund or co-fund the studentship and how it works. And then if that goes well, or they want to do something more, they

might then host some data with us or a small subset, or a larger dataset, depending on what it is. We kind of build that collaboration gradually over time.

Interviewer: Yeah, so it's a much more long-term process. Do you believe that networking and personal connections, and even maybe word-of-mouth, play a significant role in this process?

Respondent: Oh, yes, definitely. That's definitely been our experience. You know, it's really hard to go into a big organization with no contacts and say - can we share your data? But if we met someone at a conference, or we know someone already, we can say - hi, Joe, we've got this, this is who we are, what we do, and we think this particular data set might be really useful for research, are you interested? And it's a way of starting that conversation. Yeah, we still have to go through the legal agreements and all the rest of the formal arrangements, but it's getting that initial connection is where the that's where the networking comes in and the yeah, it's really, really key.

Interviewer: And is there also a networking aspect between your data donors for them to share the benefits that they possibly have gained from this?

Respondent: Oh, yes, definitely. Yeah, that's good that you mentioned that because we do, we have the master's dissertation scheme and we do an annual conference, in association with a group called DUK where the students come and present what they did. And we have presentations from a number of the data partner organizations and we're developing an alumni scheme for that event. And then we also have the data partners forum which we run, or we used to run every year, we've hit a bit of a roadblock with COVID, and all the rest of it, but we usually do that. And that gives an opportunity for the data partners to come together. And we do a kind of day conference. And we you know lunch, attend session we do networking, all that kind of stuff to foster data collaboration.

Interviewer: A majority of the work you do focuses on delivering public value, but does your research in certain cases also deliver insights to the private data contributors? The study you did on green labelled products, for example comes to mind.

Respondent: I think one of the main benefits is a greater understanding of the data they already hold, and the option to provide more advanced analysis. And they can have the skills for in house or even new types of analysis as well, particularly with a PhD project. That's not just saying we'll apply this existing method to your data set it's - we're developing a new method based on this data set to give us some useful answers and there are those benefits to society but there are probably also benefits to the sector possibly benefits to that specific company as well.

Interviewer: Do you find that financial resources meaningfully either enable or restrict partnership establishment development efforts? Beyond the basic things like the funds necessary to hold conferences or networking events.

Respondent: Well, it depends. We don't pay commercial data cost rates. So that is part of our kind of underlying ethos, and part of our funding agreement is that we don't pay commercial rates. So, if someone wants us to pay commercial rates for data, then we can't do that. We do pay for some datasets, and it depends what the data is, you know, how much and what kind of range of aspects, how useful the data is, how many updates we get, and that kind of thing. One example, of the payments we do make is to help with the organization's costs, rather than to make money per se. So, for example, it's not typical but it's a good example, it's with the local data company. We secured some funding, some additional funding, which will go to them ultimately to help them install this network of sensors. So, you know, they will get the benefit of getting the data, being able to use it in their commercial products, but we also get the data and we do research with them to develop new insights and new ways of measuring. In the end it varies, we have a budget for data, but we always were relatively conservative about how we go about spending it, because we want to get the most out of that budget that we possibly can. And a lot of the organizations we work with appreciate that. So, they see the need - we have to get the best value for money and that the money is there to help with their costs rather than to give them an income. And I suppose you were asking about enabling partnerships, I suppose that that presence of that money does allow us to do more partnerships, bit it's very common for there to be other benefits as well, not just financial. So, it might be that we do a joint PhD with them, or we do we have some other sort of partnership with them, or they get some money from another kind of business-university collaboration fund, or something like that. So, it's usually kind of multifaceted.

Interviewer: Do you find that collaborations are possibly easier or quicker to negotiate if the data set has potential value to target a problem or an issue that's very societal relevant at the given moment? As an example, we've been observing a lot of data collaborative data sharing initiate quite quickly emerged around the issues of COVID-19, presumably because of its high resonance.

Respondent: Yes, definitely. Specifically, with COVID we've been involved in the local data spaces initiative and that was all very COVID driven. Having a clear aim was really, really useful with that it helped a lot of processes go faster and there's lots of examples of COVID related stuff. I know the US and the UK Data Service has done various different things in relation to Coronavirus. And equally, having a clear problem is really useful in general for negotiations as well, kind of outside COVID. If, in general, we go to a data partner and say, you know, this data is really interesting we think there might be some good research in it, we tend not to get as far as if we go to data partners say - we've got a specific problem that we want to use your data to answer. This is what we think the research can do, are you interested? So, it's having that specific question is usually a good mechanism for starting that conversation.

Interviewer: Maybe now let's move on to your current capacity of managing the access to the data sets and more onto the side of the users who gain benefit from this data sets.

Do you find are the applications mostly coming in from scholars, master's students? Or do public sector organizations also actively want to gain access to them?

Respondent: It is probably about two-thirds academic one-third other and the majority of that is public sector. Okay. During Coronavirus, we've had a lot of interest in ethnicity data a related to the fact that you know, the impact of COVID on it for black men and minority groups was higher. A lot of data that organizations held, including the NHS, lacks ethnicity data, or is patchy so - can they use some of this data to try and supplement what they know about ethnicity to better understand its relationship? So, we had quite a lot of interest from that. We occasionally get what we call third-sector organisations. These are private companies, but who are contracted to do work for some bit of government, usually local government, it's usually people doing analysis for local authorities when they outsource that analysis. So, we classify those as third sector in our analysis.

Interviewer: Just a few clarificatory questions. You also have different type of partnerships, when the public sector not only gains access to your data sets, but sometimes you also provide research and analysis to help them make sense of it and then give them actionable insights. So, you find that this is also a big draw for partnerships with you as these entities might often not have the capacity to analyse the data and gain insights from it themselves?

Respondent: Yeah. And that's the big underpinning of the local data spaces initiative, in many local authorities they're tasked with doing analysis and work related to COVID impact COVID. But they don't have the resources to do this. They don't have the ability to take the data we can give them and do some meaningful insights of that. And we've got to two of our local staff seconded to the local databases initiative, to develop kind of analysis ready tools to allow them to do this work.

Interviewer: Do you find it difficult? Or do you have any particular measures that you do to communicate the potential value of the datasets you have to the public sector?

Respondent: A good way is by giving people examples of what they can do with the data sets. So, most of the training courses that we run, use some of the CDRC datasets. It is tricky because it's difficult to safeguarded datasets in an open training course. So, we generally use the open data. But that does act as a way in. I'm not sure off the top of my head, but I suspect they probably are examples of people who've been on training course and then we've gone on to form a kind of longer-term partnership with them. And that's the kind of area we're trying to grow and develop as well.

Interviewer: Have there been situations where a public organization reaches out with a general issue or problem and asks if you have any data that can help us tackle with this? Or relates to this problem?

Respondent: We do we do get that occasionally, yeah. It's sometimes difficult to respond effectively to that, because, by definition, their query is a bit general, because they don't know what they want. We have very limited time and resources to be able to spend with that specific individual to work out what it is they want to do. And so that it does get tricky to kind of respond that effectively and efficiently.

Interviewer: And lastly, to clarify, I would assume that if they come with any issue, you can only offer them the data sets you already have available, you do not work as an intermediary between possibly getting a new data set from one of your partners and might respond to their issues.

Respondent: We are not opposed to doing that, but we generally don't. More recently, we've had a reduction in resources. So, probably in the first phase of a grant, we'll been more likely to do that. There was a case study where we actually did do that. There was a group looking at cancer rates and they needed some loyalty card data from us. And it's an interesting analysis, and they were looking at - can you link rates of cancer with the purchases people make? It is really interesting, but there's a lot of complexity in that. The bottom line was that there is some link and you can do some modelling of that. But we didn't have the names of the individuals with those accounts we had, what they bought, and where they bought, and we could link one individual to all of their purchases, but we didn't know who that person was. So, the researchers wanted to link that to their medical record to see if there is a link, but we couldn't do that, because we didn't have the data. But we put them in touch with the data partner and they went on to do a much bigger project where they could access the names directly with a data partner. And that was a really, really beneficial relationship. And we kind of introduced them, but ultimately, they did it as a direct link rather than through us. Because the data partners saw the benefit of what they could do. And, you know, the academics involved may have good pictures, too, why they need more details. If the data provider didn’t want to make that available through us, they will have to do that on an individual basis.

Interviewer: Just to clarify, I think you mentioned right now, did you experience some cutbacks in human resources that you're finding challenging?

Respondent: Yeah, so it’s funding across the board. So that, you know, from a practical point of view, it translates as people really. So, we have fewer staff now than we used to. I think we have less money for data as well. But in the setup that we're in people are quite expensive, so that's where we've had much of the cuts.

Interviewer: Lastly, a bit more general question. In your experience, when dealing with forming those connections, and negotiating those partnerships, whether it be through you or connecting them independently, what are the most common apprehensions expressed by stakeholders through this process of forming a collaboration? If any?

Respondent: I think it's the data security aspect. Particularly with GDPR, over the last couple years essentially. And so, we've had to update a lot of our agreements to reflect

that. Part of that is kind of top level, do they have confidence in our processes, and we have a discussion with them and explain the application process and how it works. And the kind of second element is the legal agreements – to cover everything that they want them to cover. And usually they do, but it takes a bit of coordination, to get that through the process and to get both them and us as being UCL in this case, to sign the agreement to share the data. So yeah, that's probably the most common one.

Interviewer: Any apprehensions from probably the user side?

Respondent: In terms of how they apply to use the data, do you mean? Yeah, well, some of them find the process a bit onerous. You know, we have to balance the needs of the data buyer and the needs of the user. 10 years ago, the user would real out to us and say just give us your data, and we'll do some interesting research, and they might have gotten away with that 10 years ago. Now, that's just won't happen. And there is a process to go through it can be time consuming. So, for our safeguarded data, we usually say, up to four weeks, and for the secure data, I think it's six to eight weeks, but I'd have to check on the website. The users who are aware of that and engage with the process do get the data. There are some users who don't realize that, and then the timelines are just not feasible for what they want to do with the data. And it's really hard to quantify how many users see that on the website, and then decide not to apply for the data, because we've got no way of capturing that information. Because we don't know who doesn't apply, because they don't make contact with us. So, we've got no way of measuring that. We are trying to make open data more accessible. So, we're developing an API at the moment for open data. And we're looking at the options of extending that to some safeguarded data as well. But there will still have to be some sort of application process for that. And we haven't worked out the details of that yet.

Interviewer: Thank you. That was all the questions I had. Do you feel like there's something I didn't ask what you feel is relevant that you would like to share in addition?

Respondent: We do have a template that we use for license agreements. And most partners are happy with that. I think it took quite a long time to develop in the first place. And when we, when GDPR came in, we had to change quite a bit to actually reflect the terms GDPR and various details. But in general, many of the data partners are happy with it, but sometimes some are not and they want us to use a different template or different example. It's more common that most people are happy with the templates.

G Interview with CDRC High Level Management

Interviewer: Firstly, could I ask you for the record to briefly describe your role and responsibilities at the Consumer Data Research Centre?

Respondent: Okay, so I'm _____, one of my roles that I am a co-director of the Consumer Data Research Centre, which is a research project, in effect, funded by the UK Economic and Social Research Council, a government research body between three universities London, Liverpool and Oxford. And the objective is to create a bridge between the data resources of commercial and public sector organisations, with a focus on consumer data and academics and others doing work those areas to try and promote capacity building and kind of engagement between the two groups.

Interviewer: I was wanting to ask you about some background information. I'm already aware about how you were founded through the Economic and Social Research Council as part of the Big Data Network Initiative. But could you possibly provide me with some more background information, how the stakeholders first came together to develop this common interest in the Centre's establishment? And a shared interest ain the consumer data in particular?

Respondent: So, the ESRC develop certain broad themes for its research interests. And around about six, seven years ago retailing was very much on the on the agenda. And it was a particularly concern that there were a lot of inside resources locked up in commercial organizations that weren't available to the academic sector that could have provided mutual benefits, if they had known each other's interests.

As you said this is a part of the Data Network Initiative, so there were several Centres established, of which I think only a couple currently remain of which CDRC is one. But they were aimed at some particular aspects of the economy. So those are the ones that administration and government wanted. And from the point of view of the academics involved, I had from my position known the colleagues at UCL in London in the geography department for a number of years. And they approached me and said - would I be interested in being partner in this, the original collaboration was between geography departments, and so they wanted a business school involved, because of the connections that we have with business. So, they were looking for a kind of leads for the kind of data they might actually acquire. And they also wanted the Oxford connection to be responsible for the training capacity building, to scale up some of the academics that would then go on to use this data, giving them training all the way through to skills in analysing – through cluster analysis, and so on. So that was in effect, a proposal put together by four of us, two from UCL, one from Liverpool, myself (from Oxford), to the ESRC in a bid to run the Centre. And we were along with Leeds at the time we were

successful. Since the last couple of years, the ESRC has funded the Leeds CDRC separately from the Oxford, UCL Liverpool Consortium, because they have slightly different objectives. But all four institutions are still kind of continuing with the original mission. Although, of course, the last year it's being refocused slightly around some of the COVID themes emerged with the pandemic funding that's coming through. So, it's been particularly interested in things like deprivation, disease, accessibility, green spaces, all of which are relevant to the COVID agenda.

Interviewer: To build up on what you said, would it be correct to say that back in 2014, the fact that the political discourse and the government agenda was quite focused on data analysis and more so on the potential private and retail data, did that that kind of relevance and actuality play into getting the stakeholders on board?

Respondent: Yeah, I think that's right. It's a conversation so it's one side deciding, the other side responding. So clearly, the academics involved were keen to influence CDRC to choose retail as one of the areas for attention partly because at that time, it's still today, the retail sector is in some significant difficulty in this country as it is elsewhere in its traditional form. Store-based retail is being replaced by online activity. And there was concern around the policy implications of this at the time in government. And so, there was actually a policy environment that was concerned about what we call the future of high streets at the time, which involves a few consultants or reports being written for government about this and possible solutions. One of the errors that was identified was actually that there was a shortage of data and insight into this problem. There were a couple of small companies that had set up operation to collect data in this field. But they had to do that because the government itself had pulled out of a lot of data collection. During the Conservative government, they cut back on the costs of data. And in particular, they had gotten rid of a census of distribution. So, there was no national census of retail or commercial activity available, other than very kind of partial samples. And government felt that they were missing a lot of insight, because lot of the data was being created by commercial organizations and sold at a price rather than being government data that could be more freely shared. So, they were encouraged by the ESRC to, I think, create these networks in order to reduce the barriers between the stakeholders, but also save government some money – as the data was already being collected by commercial organizations.

Interviewer: Speaking of those networks, and collaborations could you possibly generalize the process of how a collaboration with the CDRC comes to be? Of course, I would assume this process differs based on whether it's a private or public sector partner.

Respondent: Yeah, it is to some extent. Although actually, we have what we call a ladder of engagement as a model for this. So, what we say is it's about building trust between CDRC and the commercial partner, or the public sector partner. Public sector partners are usually easier because usually there are some open-source kinds of data exchange contracts available, which we can piggyback on. Often in the commercial environment, the challenge is that no one has time to think about these things, often the data is lying

around and it's not quite clear who to contact. So, it took quite a while, this took 18 months to two years to actually established properly, from 2014, longer than ESRC had expected. But we had predicted this in our proposal. We had to build trust between the organizations we were targeting and ourselves, given what we wanted to do with that data. Much of the data was regarded as commercial and in confidence. So, one of the data sets that we acquired, for example, this tool called High Street Retailer Data, because we can't name the company that actually originates the data. But it's a database that contains sales information by category, contains loyalty card data. And so, it's quite a commercially significant data set. They were willing to release some historical data in order to allow researchers to add some value to the analysis that they could conduct. So, there's a good deal of retailers that are generally not very trusting of outsiders. It's a very secretive industry and in some respects quite defensive. So, you got to ask for something and the default answer is - no. So, one of the reasons for having us on board as business school was to try and open doors basically, to start the conversations off. And what we found in building this ladder of engagement was that by better understanding their needs, because often you found that commercial companies had analysis needs as well, but they didn't have enough resources to analyse the data they couldn't get value out of it. When we discovered that we developing what we call a master dissertation scheme, where we could put graduate students as interns into these organizations to do work for them. And in return, we will get access to the data set that was seen as a more even-handed discussion. So, it wasn't just about "please trust us and give us your data", it was about saying "what can we do for you?" - How can we understand your problems and your analysis issues as a as a group of academics who are skilled in this case, in spatial analysis or commercial data. So, it was a way for us in some senses to act as consultants for these organizations, as a value for them, so they could see that actually it was worth developing a longer-term relationship with us. And so, the dissertation scheme has been running as long as the Centre and we're just going into 2021 exercise and we have about 25 to 30 students placed in organizations that support our work, and often we will go on to employ their students after they finish their dissertations. So, there's a kind of virtuous circle there. Once you get someone actually in the organization, it's much easier to have a conversation. and again, one of the big challenges around the legality of the setting up and securing data storage and protection. So even though we've exited the EU, we're still bound by GDPR regulations. And so, the control over data is still quite a sensitive issue. One of the things that we had to put in place to encourage trusted engagement was we had secure facilities to look after the data if necessary. So, as you may have seen, we have kind of three tiers of data, public or open, because anyone can access and download is on demand; secure, where you actually have to apply to use the data, justify why you want to use it; and then the safeguarded level, which is where you have to physically go to a secure lab, and access the data from a laptop, which is not connected to networks and so on. And in fact, those facilities are probably more secure than the data facilities that the companies themselves were operating. So, we had to add additional security precautions to reassure the lawyers in those companies that we would look after their data quite carefully. So that that aspect was again part of the trust building process.

Interviewer: Could you also briefly talk to me about finding public sector partners especially in light of public sector often lacking human capital. Of course, it's great to salvage this private data, but it needs to go to where the need is.

Respondent: That's right. And I think government data holdings in the UK are both hot and cold. At the moment we're in a reasonably positive place, so the Office of National Statistics, particularly given the COVID pandemic, is actually being very agile in developing new data sources and techniques for data collection. They also have a data campus now in South Wales, where they are taking in their own interns and building up expertise to try and build resource. They've always had data in government, it has always been mostly poorly resourced. But actually, again, the pandemic has led to some really quite innovative, creative solutions to get to collect and disseminate data. And that has meant the government was more open to API's and to open-source activities. We went through a phase where things like the census, they were going to charge users phenomenal numbers of money to access this information. But increasingly, we're looking at some mutual benefit arrangements. In some areas, like for example, employment data and unemployment data, where there's some sensitivities, there's still a need to go through a kind of a contractual arrangement - service level arrangement with the public sector. But often, curiously, the relationships with individuals in the public sector are better than with the commercial sector, because those people stay around longer. So, you find that ONS contacts are often sometimes they're previous academics or sometimes they are would be academics and thus they have more sympathy, if you like, and more empathy with academic objectives. And they do see the benefit in working with universities institutions to transfer value. But the communication is not so much where the government departments but with the assisting agencies and the data holding people within those departments. The conversations we have with departments are less helpful because they don't really understand the data environment as well. So, you talk to a Minister or to a senior civil servant, who isn't data connected, and they're afraid that they're unsure about the environment and what the implications are. So, connecting with specialists, especially if their highly regarded by government, together is the way to do that. And our National Stats Officers have been very cooperative and very encouraging of the work we've been doing.

Interviewer: What about in instances where the in public sector is not as data contributor, but a potentially data users? Your colleague previously mentioned that around only one-third of the applications to access your secure databases, are from public sector actors.

Respondent: Yeah, this comes down to resources, I think. I think he's also probably combining central and local government in that natural response, because what we've found over the last few years is that people at the local level of government where there is an interest in local differences, for example, our datasets are very, very granular generally, they go down to quite a lot of detail at the local level. Local governments don’t tend to have resources to even onboard this data and analyse it. So, think about Oxford

city council, for example, it's my local council. There's a city council and the county council. So, two levels of administration. Within both departments they're probably two people responsible for economic development analysis. They simply haven't got the time, money, skills, resource to onboard this data and analyse it effectively. So, they often outsource it to another actor, to a consultancy or to third party at central government level.

Yes, you're right. That doesn't happen either as much as it could. Partly again, I think, because the awareness may not be high enough in certain departments. Partly again, I think, because of time pressures and resources, the body simply aren't there to do the analysis is my is my sense. And they don't necessarily frame the questions in a way that they see – we're the answer. They'd rather commission McKinsey or Deloitte to do this work, if you like, rather than using an existing academic consulting. And that may well be our problem in not having raised awareness well enough within central government, of our capabilities and resources. But bear in mind that we're an intermediary, although we do a bit of research ourselves. We're an honest broker in between the data holders and the data users. And so, the government, we can make our presence felt as an intermediary. And we do on all kinds of conferences, platforms and so on. But it's often harder in central government to find the key individuals who are really dealing with those questions. And so often, we tend to go through the statistics people because they know who the internal client is, within the department for police services.

Interviewer: Would you say for example, because I think though your Centre also does education initiatives, specially targeted at public sector workers, do those kinds of events courses produce any kind of connections that then develop potentially into public sector partnerships?

Respondent: A bit, but not very much? Because, again, for example in Birmingham there are probably two or three people responsible for current development initiatives within the local authority. There are lots of other quangos, lots of other sort of quasi-public administrative bodies, responsible for economic developed and so on. But they, again, their habit is tending towards to go to known consultants to do additional work for them, rather than to university bodies, universities to be lower down the list, partly because they're seen often as being slower than a commercial consultancy to produce results, partly because the awareness isn't there, because the universities are not as good as marketing themselves as consultants are, potentially.

Interviewer: Would you say that professional and personal professional connections, as well as networking, really do contribute to the efforts of establishing collaborations.

Respondent: It's word of mouth, absolutely, it's literally word of mouth. So pre-pandemic, we were holding probably three or four kinds of national events a year, we were doing a lot of outreach, speaking at any organization that wanted to ask us to speak. I would represent Oxford at two events a year, one was for existing data providers and we would invite prospective data providers along to those meetings to hear the good practice and the case studies. So, they would learn how effective it could be. In the other

half of the year, I would run sessions for academics, where we would talk about, again, the good practice cases and encourage individuals to download the data as end users. The bit we found this difficulty with, I think, was actually getting a good supplier public sector representation in the either data supplier event or the customer event. Again, targeted to identify the individuals that could be invited, but some of them are on such tight budgets they couldn't even physically afford to come to Oxford. Because their budgets couldn't let them leave Birmingham or wherever. So, you don't have to subsidize some of these things sometimes. So, we'll come along and if there's no charge to attending, and we'll pay welfare, because actually, local authorities in the UK are extremely financially constrained. And that's led to a lot of cutbacks in what are seen as not essential services. And further reporting about the pandemic, now probably about a dozen local authorities are in danger of going bankrupt. So, there's very little kind of financial resource to pay for this kind of stuff. If you look five, six years ago, you'd have found local authorities at the local level with specialist data officers, for example, Southampton, Hampshire County Council in the south of England had specialized data officers who were responsible for assembling data hubs within their counties, most of those have gone because they could no longer be afforded by the authority. And if you wanted to, I can introduce you to one of them, who used to work for Hampshire who now works as a consultant, about his experience of hand-to-mouth existence with a local authority in trying to get the resources to assemble data resources.

Interviewer: It's so interesting that you brought up resources, because that was actually going to be my next question of whether you find financial resources meaningfully enable or restrict the Centre's partnership establishment efforts?

Respondent: Yeah, so like all the Centres, we had to cut back in our budget. When we applied for a last renewal, we were asked for 20%, cut in resource, so they'd like us do the same with less recourses, basically. So that's meant that we've lost lose a couple more heads, administrative heads. We lost our chief administrator. The reason for this is because funding has being redirected towards more explicit COVID relevant research. And also, the government's research budgets are being cut anyway, in the current economic situation. So, there is a sense in which it's getting harder to do the job we were set up to do, because of the incremental cutbacks. And that's also affecting national government, as well as local government, as its efforts are redirected towards the pandemic and then pandemic recovery in terms. The National Stats Services for example, are spending a lot of its time developing new datasets to support COVID awareness, whole new dashboards and just as Netherlands and Belgium are doing in terms of massive investments in new dashboard technology and surveys of infectivity and deprivation and so on. Which are really important to understand the incidence of the pandemic. So, a lot of the resources that were there are being redirected and reoriented, which has meant that some of the business-as-usual stuff is being neglected or is being cut back.

Interviewer: So apart from the kind of payments to data holders and administrators is what other outreach measures is the lack of resources affecting the most?

Respondent: Apart from data acquisition, I would say it's probably well, it's hard, it's hard to be definitive, because actually, during the pandemic, most of our activities have been online, which has been a lot cheaper to run than physical events. So, we haven't had to incur kind of room hire and catering costs and so on. And curiously doing it that way has allowed us to attract more people than we would have been able to do with face-to-face events. Because people can just switch on zoom or teams and join the webinar. So curiously, it's had a somewhat liberating effect to some extent, which is offset some of the constrained resources. So, I probably would be able to do more in a year, with about with half a resource that I would have been able to do pre-pandemic, because most of the training and so on has been done online. Where the costs do start to ramp up, at the moment, is in terms of creating what we call asynchronous content. So, this is recorded material, because that requires some production technology, editing software, and a fair degree of overhead in creating that content, compared to the face-to-face training that we used to be doing. Because you do it differently when you do this online. So, we're planning for example, at the end of April, a Master's dissertation scheme alumni event online, which will bring together the last kind of six or seven years of alumni and we've put a database of where they ended up, as part of the exercises to see which companies have taken them on, their projects, what they have done. It’s a kind of a reunion for most of them as individuals, and to demonstrate the effectiveness of the scheme, improving the talent base in this area, but also to allow us to share some of the latest findings from some of these companies. The work that these previous interns are doing now they're working for the organizations they used to research for.

Interviewer: And quickly to circle back to the overarching issue that keeps coming up with COVID-19 and the pandemic. Even before COVID, did you find that possibly partnerships were easier to negotiate if the targeted issue had a strong symbolic or societal importance?

Respondent: Absolutely. That's why retail was the opening gambit, because at the time government was desperately concerned about the future of high streets. Since then, clearly the health agenda and deprivation has become first and foremost social justice, and so on. And so, part of the skill in running these projects is anticipating, you know, the sort of the fashionable themes, and the current topics that are going to get traction, both in public sector and in private sector. And, you know, let's say their food service companies worried about diet, accessibility to their operations, closeness to public gardens, and open spaces, and so on. All those kinds of geographical factors start to become more relevant. We did pre the pandemic, for instance, we had a Health Index called AHA, which is sort of Accessibility and Health Index for the for the UK, which has actually proved quite interesting. And we've done a lot of work with the existing deprivation indices that the government already collects, to try and repurpose those for COVID. circumstances. So, it's important in these kinds of Centres to be very alive to the evolving with themes and hot topics and to anticipate those. So, when a Minister says - I've been reading about the 15 minutes city, for example, we can already say - well, we've been doing work on this and here's the data to prove that kind of thing. So, it's anticipating some of the sort of

political and governmental sort of topics that are going to be followed on top of the agenda.

Interviewer: Do you find this also relevant with public sector data donners?

Respondent: Things that make it actionable, easy to understand. But also, in English we call these brownie points where the company does this, they're seen as being good citizens. So, if they engage with a say a health agenda, and get involved in this and that becomes known, it's a good tick mark against them in government. They're seen as being good citizens and good partners. So, there's also a kind of a bit of a citizenship agenda for private companies as well in this, although obviously, that primary motives are commercial.

Interviewer: My last kind of really big question, apart from the extensive data analytics that your service performs, they already talked about, and also the secure storage that you have offer that kind of elicit stress, what other factors do you believe constitute a major draw for collaboration for you partners?

Respondent: There are a few, one of them are the network benefits. So, you're not just getting data, you're also through the events, through communications, and so on, you're joining a network of people that do the same kind of thing. And so, it sets up the opportunities for communication and contact across sectors. Our client base is not just retailers any more, it goes through food, service, through anyone that has some consumer data has a spatial dimension to it. And so that includes people like house sales, data providers, insurance companies, banks, and so on. And so, by onboarding them as data providers and creating networking opportunities, they get to meet each other and get to collectively address and solve problems, perhaps lobby for changes in policy, this kind of thing. So, it's a, as I said the network benefit is one that I think is probably undervalued, but actually is quite useful, as well challenge at the moment, but it's, I think, even online valuable, because this is a small world, ultimately people that do these kinds of things, wherever they happen to be. I chair a small professional association for location analysts, which sounds like a very specialized superset. But there are 400 members of that in the UK. And that's it. That's almost the entire spatial analysis sort of skill set in the UK, so it's a small world. And so often in an organization, if you're the insight manager or the analysis director or whatever, you're very lonely, you're on your own in that organization. And so, there's some real benefits to meeting others who are in the same kinds of roles, whether that's in private or public sector organizations, because you often share common problems and concerns, and could perhaps collectively think about some solutions. It also helps you when negotiations with commercial software and mapping providers, in a sense in which “what do they charge you for your”. So, whether it's commercial contract negotiations can be very useful to have those kinds of networks of users.

Interviewer: I think that's a very interesting point, a lot of people bring up networking as a method of acquiring partners, but not so much present it as a value proposition to them as well**.**

Respondent: I think it really a way of maintaining them. Because if there is no value in it, they won't continue. Otherwise, you'll just get a one-off data set with logical view or a one-off project, you have to demonstrate that continuing value. And part of that is actually this "if you don't do one this year, you'll lose contact with all these people." So, it's to make it a bit stickier, I suppose.

Interviewer: A lot of your academic research and analysis also offers two sided benefits, the public and the private sector.

Respondent: So, in those events, what you end up doing if you're inside the network, then you get early access to the research. So, you'll see some of the things that happen, not just in our research and the consortium, but the research that's been permitted through the various agreements that we sign. So those annual events, we wouldn't just have Oxford, Liverpool, UCL presenting, we'd also have some other academics who did work presenting back to the network about what they've done. So that something else that will be lost, if you kind of dropped out as it were the relationship.

Interviewer: That's all the questions I had, would you like to comment anything if maybe I feel something's relevant, and I failed to ask a question about it?

Respondent: I think, one of the things is the kind of the national and cultural context of this stuff. I've always found that Belgium has quite an enlightened approach to some data. There is a culture in some countries for looking to universities as a point of access. In other countries this collaboration is seen as being too remote or too academic, too distant, or take too long to do analysis, and so on. So that dimension, I think, is quite an interesting one. It does vary culturally, internationally, around the world.

Interviewer: That's very interesting. And would you say that in the UK, that culture is stronger with government and university partnerships?

Respondent: It's getting better, one of the reasons for the establishment of these networks was to try and improve it, because actually it was pretty lousy historically. It was about who knew what and who knew who and it was very strength mentored, if you like. I think the research councils have tried to improve these initiatives. For example, funded partnerships between small and medium sized firms and university departments. So, funding scheme which allows digital academics to work with a medium small to medium sized company, to help them with a particular problem. And so, their growth in those kinds of collaboration schemes, that level as well. And I can't remember the name, but a few more. So, I think UK is getting better, but if I think about other examples, I think about Australia, I think about Japan, that were much closer connections between universities and governments.

H Interview with High Level Management at DDI

Interviewer: Firstly, could I ask you to briefly for the record to describe your role at the Dalberg Data Insights.

Respondent: Yeah, sure, good to virtually connect. So, my name _____ and I co-lead the Dalberg Data Insights Team. I basically lead the kind of business development in the firm, but also work very closely with the project managers and the delivery of the project. So very closely between the partners, external and internal, as well as the project management team.

Interviewer: Okay, lovely. And to start off with might be the most general question, but could you walk me through the generalized process of how collaboration comes to be, but most specifically involving public sector actors? Of course, I understand most commonly, you get approached by clients themselves, and you provide the service they're interested in. But _____ mentioned that in certain cases, there are examples, when you actively reach out to beneficiaries if you think they can benefit from the insights or the private sector data access, you have to provide.

Respondent: Yeah, you know, there are so many different and divergent examples. To be honest, the field where we operate is still quiet, very nascent and not much advanced in terms of understanding, especially in the in the public sector. So, it takes education, right, so sometimes they will reach out, but in most cases, it's through conversations that we realize where the opportunities are, where the challenges are, where the needs are. And then it's just kind of a very complex ecosystem of funders, the public sector, that usually will figure out what a solution could look like and then start working on it, while we're still you know, very much refining throughout the whole process. Very often, and you probably know that, different foundations and partnership, funders, their work it already the public sector whether it be a Ministry of Health, a Ministry of Agriculture, they already have their preferred partners and funders that they're working with. So, they kind of engineer together requests for proposals that very much reflect the needs of the public sector. And often, organizations like ours will go ahead and respond to those. So, it comes really demand driven. So that's one way, to be a little bit more structured, and the second way to go about it is to be more of an agenda setter and try to proactively look for those opportunities, by already knowing in depth what the challenges are and proposing to funders to be able to find such solutions.

To be very blunt with you, the public sector, they don't even utilize their own data very well today, and then talking to them about bringing in external data sources to fill their gaps when they're not even using current data, it's very complex. So, you know, we've learned a lot on the way and usually you need to take our organization on a journey, if you are to talk about data or introducing any kind of new and innovative solutions around data. And that journey usually will start with making sense of their own data and

information and finding the right champions within the public sector that see the value in data and see the value of kind of the more advanced analytics, but again, you need to start from the basics, and then bring in the examples of how for example, telecom data can fill data gaps in poverty mapping or, you know, population will want to work methane, the potential for the solar panel industry in Uganda etc. So, again, I think it's just like anything else, if you need to introduce a new innovative solution it is to relate to people and it will be hard to bring in the advanced analytics to a partner that has not even used survey data on their own. And it's really important to start from the basics, take them on a journey, and then introduce them to new topics.

Interviewer: It's also something that I've been finding that a lot of the times intermediate organizations don't only grant access to data, but also provide the expertise and the help and trying to help assess the private sector, their own data needs and capabilities. And of course, education is a big part of it. That also comes as a hindrance, especially with public sectors, limited resources, and human capital. We're going to touch upon this again in a minute when we talk about value proposition, but just to briefly specify how you partner with foundations or third sector organizations on the ground to forge this first links with public sector and to understand their needs. If you could specify a tiny bit on that. If that I understood correctly, that's what you do.

Respondent: So yeah. So, I kind of tried to explain that a bit, but maybe I need to be a little bit more concrete. I work in the global south, right? And most of my experience comes from East Africa, but also in Asia, and in Latin America. But so, we are talking about the global side. And normally, let's take the Ministry of Health in Uganda, they have a certain need. Ministry of Health, Uganda works already 10 years plus with certain development partners, and foundations, and they are all very much interlinked. I think the WHO's of the world, think UNICEFs of the world, think an umbrella organization like the Global Fund, or the Gates Foundation, or the Clinton Health Initiative. So, they only work with these partners and there are competing kind of interests, obviously, among those partners, but that's a separate topic. Then we have a Ministry of Health that will come up with like "Hey, we need now, during the epidemic, we need devices for data entry, because we don't have laptops and people cannot enter data for, all we need, and we struggle with data management or data analytics and specific topics", then these partners will kind of come in together and brainstorm who can bring in what and how this can come together. Often, if there is a need UNICEF, and I've seen it happening multiple times, UNICEF or WHOs, you know, partner organizations will actually publish an RFP that will formulate and that will be together with the ministry, there will be a request for proposals, right, that will formulate what is actually needed, what the challenge is. And then organizations like ours will bid to propose a solution, and the winning organization will go ahead and provide the solution. That's how it usually works. Sometimes ministries have their own money to fund projects, but often in the global south they do that through partners, which work very closely with them and have been working with them for a while. And again, you know, we talk about WHOs, UNICEFs, if we talk about the public health space, and these are again, there are individuals behind them that are very close,

related with the public sector, and understand very well the politics and the dynamics. And again, they will publish an RFP, and then things will go sometimes.

So that's one way to go about it, which is a common way. There is also other way, which is what kind of I'm very much excited about. For example, again, example formation three, there is this global fund for Kovacs develop initiatives. So, they're saying “hey, how can data help support ministries of health with equitable vaccine distribution?”, that's the main objective. So, the fund comes with that very big objective. What I've done myself is even yesterday, when at the Ministry of Health, talk to the technical partner and told them “Hey, guys, I want to understand everything that's happening within the facts and distribution, kind of value chain, what are the main challenges you're facing when it comes to data”, and then I've come up with ideas, use cases, solutions that can potentially solve those problems. And I heard actually a lot of interesting things. And now based on everything I'm hearing, I'm going to go ahead and write a proposal that might include non-traditional data sources and data sources from the private sector, but also might not it might just include data that is that is out there, right? So, there are just two ways of how the public sector is basically working with external organizations in solving their own problems. And maybe there are others but these are the ways that I have seen work.

Interviewer: Thank you, that's very specific and very clear. Apart from those you also have direct relationships you firm by public sector actors directly approaching you, I would assume so, is that right?

Respondent: Yeah, if they've already worked with us, you know, it might happen. But they don't usually approach us and say “hey, look, this is kind of the challenge we are facing” and there will be probably no funding at that point of time for it. So, they will help us also look for funding, which is kind of more the second way, which I'm telling, like, I found that there is the foundation or the fund that is funding certain types of challenges. And then, you know, I'll reverse engineer and go to the ministry and approach them and figure it out. But it can also be that the ministry approach does, indeed.

Interviewer: So, it is possible that when a ministry or the public sector body approaches, you can also help them find the funding for the support you provide through linking them with other organizations.

Respondent: It's possible, but again, it's more likely that the ministry finds the funding, than a private organization like Dalberg, because the ministry already works with some of the biggest partners, and those partners would like to see the buy in from the Ministry, not the private sector, because the private sector have their own interest, right. So most likely, it will be the ministry that will partner up with the funders versus the other way around.

Interviewer: Okay, great. And now to quickly move to maybe the other side, which is your agreements with private data donors, and specifically _____ mentioned that you have this data as a service agreement that if I understood correctly includes the possibility

of accessing and leveraging telecom data for a wider set of purposes, not just very specific purpose, if I understood that correctly, or maybe it's a mix. And if you could tell me more about that, and in such cases, how you meet the donor's privacy and commercial needs, and concerns in those regards?

Respondent: It's a good question. I'm not sure how _____ articulated it, but the reality is that the data that we would have access to is for very concrete purposes, they would never just share data widely and do whatever you want with it know. What happens, and I secured this data collaboration agreement with the biggest telecom operator here in Uganda, the data as a service model that was referred to you, you will have their willingness to collaborate on data, considering certain regulatory, you know, policies, so data privacy, both from our side, our side we are GDPR compliant, because we're also a Belgian company, and then Ugandan regulations, because we're in Uganda, and we have this kind of a broad data sharing agreement that specifically outlines the process and the way for us to be able to get access to the data. And that will include something we call a project approval form, which does not need to be, and this is very important, because, you know, that legal agreement that big data collaboration agreement signed by the legal team of the telecom operator in this case, but then afterwards, we set a concrete meet for a project, and that's the project approval form, on a project-by-project basis. And that's actually the really important part, because I cannot just use the data for anything, and it's one of and then, you know, like, very concretely, I will cover asked, let's say, by a UN agency on this kind of a data set for research purposes, I need to write it up, specifically, who has access to what kind of data how are we preserving the data privacy, what data is required. And then that's signed, but not legally, and that's the tricky part, because that doesn't require legal signatures, so it takes much less time, and that is signed by the commercial teammates, and then we move ahead. And again, this is done on a project by project based, very well defined the use case, they use the data privacy, the end users, everything in those projects'' Google Forms.

Interviewer: Okay, so you've created a sort of larger framework within which you can apply on a case by case, project per project basis for approval, but it's governed by a larger overarching…Okay, great that clarifies it. That's great. And what do you believe constitutes a major draw for collaboration with you for partners? Because of course, I think there's maybe much more straightforward on the side of the beneficiaries, but for example, what's for the telecom partners? What are the benefits of sharing their data? Because you both commercial actors, so there could be a potential clash?

Respondent: Yeah, definitely a few things, you know, first of all, for any kind of a partnership, you need to find the incentives of each partner, nobody will agree on anything if there is no win-win. And as you mentioned correctly, these companies are private entities that are very profitable, and know, that doesn't necessarily is mutually exclusive of any kind of a social element, but the commercial element is there, for sure. So here, there is definitely a cost that we pay for the data. So, there is a commercial component to those data sharing agreements. And that, I believe, is the way to go because the telecom

operators or the entities, actual data also incur costs of being able to share that data. And then the other one is, you know, especially when you work with the public sector, and I think that's your interest, there is a thing of the greater good - like okay, if we share that data, for example, in Belgium, and you are aware of this project with the Ministry of Health and the three telecom operators, it there was no commercial component, but they knew that understanding almost real time the population movements in COVID pandemic, it's really important for understanding how the lockdown measures are being respected or not. And that came as a request from the Ministry. So, the telecom sector was willing to collaborate without the commercial component. So, there is this thing of drives the social agenda of whatever government, whatever ministry, and we are going to do it. So, it's a bit of both.

Interviewer: And just to a small clarification when we're talking about commercial reimbursement for the data, are you are you paying the full commercial price or reduced price? Because intermediaries, some pay full commercial price, and some because it's being used for public sector needs, or l the greater good pay reduced prices. Or is it also a project-by-project basis?

Respondent: It's case to case, and to be honest, for us it really depends on again, willingness to pay, but we are a commercial entity, and regardless whether we work with the public or the private sector, we need to pay our bills as well. So, it will be very rare that we work pro bono, sometimes we do when we have the resources. And, again, you know, let's say we have the resources to do that. And we really, truly believe in the project and the impact that that project would have, then we're willing to collaborate. In most cases, it's got us as we work with the private or the public sector, we will charge for the services.

Interviewer: Maybe just to clarify my question, of course, I'm aware you're don't really work pro bono, what I meant, of course, you charge the beneficiaries in the end for but when you're accessing the private data sets, in that case, do you pay for commercial price?

Respondent: I mean, it's really case by case. And the model is also very hard to talk about the pricing model that is very standard, just because the field is also so early stage that everybody's still figuring things out. But definitely though, data providers are also charging based on who's willing to pay what and there are a lot more compromising in the public sector. But that's not always the case.

Interviewer: My last question is if there are any other factors that come to mind when you are from personal experience that help in the process of negotiating these partnerships with both private and public actors, such as personal professional connections, or maybe the Dalberg name, reputation or the by financial resources you have at your disposal for outreach activities, like anything along those lines that you've found to be a relevant factor?

Respondent: Yeah, maybe a couple of things that come to mind. One is showing the viability of those use cases in other places and geographies. You know, and bringing the examples of “hey, we did this in Ghana, and it works, we did this in Belgium, and it worked, and, you know, you might not see the value today, but it's, it's proven to work in other places”, a very straightforward. There is also the perspective of the leadership, which in all the partner organizations, so there is a telco or the champions in the ministry, like these people need to be on board, and this needs to be their priority. And they need to understand also what this is precisely. And then kind of the stars are aligned.

And then definitely, you know, the connections make a difference. And the more you know, people, the more they trust you and you trust them, and they want to work with you. So, I think, yeah, those three things are definitely important.

You know, second being, again, for a telecom operator, the task model is not even a priority, they're not going to make space for it, because probably doesn't bring them enough money. Maybe because they have other 10 competing priorities, and they can only, you know, spare as much time. So, all these factors play a role.

I Interview with Senior Project Manager at DDI

Interviewer: Could I ask you briefly for the records to describe your roles and responsibilities at the Dalberg Data Insights?

Respondent: I'm a Senior Project Manager at Dalberg Data Insights, and I'm based out of the Nairobi office. We are a management consulting firm part of the Dalberg Group, but then, I'm personally part of the data science business of the group. So, most of the projects that I'm working on are having a really strong data digital angle for different verticals, as we say. So, we work, for example, in agriculture and food security, employment, gender, financial inclusions, public health, and others. And I have also, I would say, a bit of a focus on agri-tech, as we say, so technologies and data for agriculture and food security. So as a Senior Project Manager, to answer to your questions of roles and responsibilities, I'm basically managing people and projects to conduct our activities, especially in East Africa, playing the intermediary role between the execution team, and then the client, to make sure that basically our operations meet their requirements and their expectations.

Interviewer: Great. Thank you. And before we jump into the more specific questions, could you give some background about how Dalberg Data Insights, this branch kind of came to be, how it first developed, and how the stakeholders came together to develop discover shared interest in this endeavour?

Respondent: So, Dalberg Data Insights, was actually before the kind of part of another company, that's called Real Impact Analytics. It has been rebranded as Reactor at the moment. So, at that time, it was a Belgium's start-ups/scale-up. And the team that joined Dalberg, to create Dalberg Data Insights was actually called Data for Good. The Data for Good department for Reactor, that company that at that time was still Real Impact Analytics. The whole story is, is a bit complex, or just basically also related to investors of that firm, but there was a willingness to spin out of this Reactor firm. And then at that time, there were different options. One was to just go into more standalone mode. And then the other one was maybe to join another firm that is looking to add a certain, let's say, set of data related capabilities. And that's how the Dalberg group was contacted, because their main business is what they call Dalberg Advisor, which is more strategy consulting. But then they also wanted to add this angle of data science and those capabilities around digital transformation and broader, let's say, technologies to solve the social challenges that Dalberg as a group is aiming at solving. So, the integration happened about four years ago, basically, and that's also when I joined. And so, I personally joined at the moment of the of the transition from reactor to Dalberg.

Interviewer: Alright, great. And then before I jump into the other questions, could I first asked you to give me an idea of what kind of private data sets you work with? I know

they're very diverse. But do you establish some sort of long-term partnerships with private companies to regularly supply you with data or how that works?

Respondent: Let's say, by the essence of the work that we used to do with reactor, the main private data sets that we've been working with are the data generated by the telecom operators. And so, the telecom operators have different sets of data. But let's say that the main one is data on the customers, so just standard CRM data. Then another set of data is what we call the events or the detailed records, the call detail record. So, every time that someone is giving a call, receiving a call or sending an SMS, receiving an SMS going on the 3G and so on.

Then you also have, especially in the geographies where we operate what we call mobile money data. So, it's the money and the transactions of money that can be carried on the phone with mobile wallets. And then you have like other types of data more related to the infrastructure of these mobile network operators. So, it's kind of to give you a sense of the data from the telecom operators. And that's where we've created a series of more long-term partnerships to access that data.

The main one that we are pushing is what we call the “data as a service agreement”, which is an agreement between Dalberg Data Insights, and then different telecom operators to access telecom data for different purpose. So, it's basically establishing a framework to securely, ethically, and also in alignment with the local regulations, access data that we are accessing, but also defining a commercial, of course, agreement. Because it's, we are a for-profit company and telecom operators are also for-profit, so of course, it also comes down to some commercial agreements, to define a bit the way we access to data.

Those are the private data, the main data set that we access. But then of course, we work with other privately held data sets that are sometimes open. For example, Twitter has an open API, it's what we call a privately held data set, because it's owned and handled by Twitter, which is a private company, but of course, it's publicly available. Same goes for Facebook data that we work with as well. And then depending on the project, we access data from our clients. So data, for example, from banks, those are privately held data as well, but that remain usually within the context of projects where we work. Yeah. And then the last source of data that we've worked, as well as the satellite images, some of them are public, like the ones generated by the NASA and ISA, which is the European equivalent European satellite agency, Space Agency. And then, yeah, but there also like some privately held satellite images. And then companies like Planet Digital Globe, and so on. So yeah, just quickly giving you a sense, yeah. That's very diverse, but to quickly to understand.

Interviewer: So, you are continuously seeking new partners and donors, private sector data donors, is that correct?

Respondent: Well, yeah, it's so by the essence of our business. So, we are a consulting firm. So, we work in different geographies, different verticals, different clients. And for

each project, we need to assess what kind of data we need, basically. And so according to the sector, the geography and the client, we need different data sets. So, we are always looking for new partnership for the data, for the finance of the project and for the operations of our project. So yeah, it's a, let's say, an ongoing exercise and where we always need to be up to date with the technologies, the trends on the market, and then the needs and requirements of the clients that we are aiming at serving.

Interviewer: So, if I understood this correctly, once again, when you're delivering insights to public sector, local government, it's more that they approach you with the issue, and then you seek out data sets based on those needs and the project needs and not so much that you just offer the data sets that are already available to you from a third party?

Respondent: I would say it's a mix. It's really hard to give you a standard way to approach a project. Sometimes we work in context where there is a huge data scarcity. So, there is basically very little available. And then in that case, we indeed have to look for data sets that can answer to questions that a client has - what is the gender gap in certain part of the country, what is the impact of the investment of a national bank, for example, in terms of poverty? So, we want to measure the poverty level, but we don't have data on the poverty so we need to find it. So sometimes there is nothing and then we have to come up with the solutions. So, we can look for a kind of big data or data that we can remotely access. We can also do primary data collection with some partners. We usually don't do the primary data collection ourselves, but then we can allocate part of the budget that was initially for us to do so primary data collection, to do a phone survey, for example, or going with numerators on the ground. And then we can just use that or sometimes use that in combination with big data sets.

And then in some contexts, the client itself also has data to bring to the table and expects us to actually leverage that data. Sometimes it's only the data, sometimes it's that data in combination with other datasets. So, there is no one size fits all. I would say we are completely agnostic, when it comes to the data we work with. What we are not agnostic about is, of course, the privacy, the ethics and the security. So regardless, the data we work with, we want to make sure that kind of the international standards and also our own standards, which are put at a higher level than international standards, I believe, are respected. But when it comes to the source of the data, and, yeah, and the way the clients would like us to, to work with it, then we just adapt.

Interviewer: So, to make it clear, the service you provide is more delivering insights based on the data sets you have, and less so about delivering or granting access to third parties to the private data sets, you have?

Respondent: Oh, no, no, we don't give access to the data that we have. This is in most of the agreement that we have. This is not basically possible and not allowed. Yeah, we've

just delivered the insights from the base of your analysis and data exactly in the context of the contract that we have with our clients, and so on under anonymization, for different ways that you can anonymize the data. But, all this is regulated by both the agreements that we have with the data holders and the agreement that we have with our clients.

Interviewer: Okay, great. And do you conduct any active outreach measures to gain public sector partnerships? Or do they generally come to you themselves?

Respondent: Okay, and it's probably a mix. We do we do both? Yeah, we do both. So sometimes, yeah, I would say part of the work that we do is, is looking at, for example, request for proposal, or the call for proposing IDs that are open, and that can come from different kinds of clients. And then also, when we have IDs that we want to propose to certain clients, we also try to reach out to them, so maybe more or so being proactive. In other cases, it's also being aware of an opportunity or the need of a client and then trying to answer that. So, it's a mix.

Interviewer: And do you find that personal and professional connections contribute any in a significant way to the establishment of new partnerships, word-of-mouth or previous networking connections?

Respondent: Yeah, I mean, of course, especially with certain clients, like, if you work with clients who are not, let's say, forced to go for open call, let's say, which is usually the case, for the private foundations or some private clients, and you can leverage, let's say, previous relationship and build trust with time, I think it's a big part of the work that we do – is basically, we try to build relationship on the longer-term, also, by the essence of our objectives as a firm. So, we want to bring sustainable and responsible change, and not only work on the short term, so we try on every project to establish a long-term relationship that can, of course, facilitate in some cases, let's say the business.

Interviewer: And is there any type of networking or outreach events that you also hold, for example, maybe training seminars or something like that, to really draw in new connections in a more social context?

Respondent: Yeah, we sometimes attend, I would say conferences, a bit less since COVID, of course, but conferences where we demonstrate what we do and where we give a sense or webinars. So, these kinds of events like we are doing them on a regular basis. But, it's again, a mix of ourselves willing to do it to potentially connect with new clients, but also ourselves genuinely willing to share some takeaways, some key learnings and doing some kind of new transfer.

Interviewer: And lastly, in the area of the personal or network connections, do you believe, or to the extent of your knowledge, because Dalberg Data Insights is part of the bigger Dalberg group, does that have a positive influence or impact when it first came to establishing this long-term stakeholder relationships, especially with data providers?

Respondent: You mean, the Dalberg brand? Yeah, definitely. So, Dalberg, as a firm has been out for, like, 20 years now. So, t's of course, a brand that we benefit from. We have, of course, to comply with some quality standards and other procedure, standards to be part of the group. But in exchange, of course, the group benefits from our own experience and our own expertise. But we also benefit from the brand itself, the recognition that it has in the presence that it has in certain geographies. If you take the example of Kenya, where I'm based out of, Dalberg, has a really strong presence and a long experience and track record in this country. So of course, it's easier to open doors or potentially connect with a new client under that brand, then, if you just start your business.

Interviewer: To move on to a different aspect, you mentioned that you're a for-profit group, of course, but when it comes to public sector, stakeholders or partnerships, are financially reimbursed for the insights you provide the guidance, is that correct?

Respondent: Well, just the insights is a bit more, it's a bit restricted. So, the way we are presenting what we do is that - we are building an inclusive data ecosystem. And in the data ecosystem, you have different components, we usually split them in three key pillars, you have the community, so the set of users, but also contributors, managers, administrators of the ecosystem, then you have the governance pillar. So, what are the processes that you actually put in place to have this ecosystem running and then the technology is just one aspect, the insights are the very, very final part of the data value chain, but you have a lot of steps, before having insights. You have the collection and the access to the data, then you have the cleaning the pre-processing, then you have the story, then you have the analysis. And then the provision of the insights is just the very last, let's say step of the data value chain, which is just one of the three pillars of what we are doing, which is the technology pillar, but we also do governance, we do community, and then the technology itself, we offer services throughout the value chain.

So, we cannot restrict what we do to just delivering insights for some clients, we are just developing an algorithm, the algorithm, it says is not any insights, right? It's just a kind of a logic that can load on existing data or on data itself. And then generating insights. But what we sell is the intelligence, the artificial intelligence behind the algorithm. But there are no insights on that front, in some cases, mobile application.

Interviewer: Of course, I understand that you do read much greater work, but for the purposes of my research, I'm more interested in the processes where you give, very specifically privately held data insights or conclusions to public sector actors. But yeah, of course, as you said, you provide services where that could not even be a part of the private data related aspect.

Respondent: It's difficult for me to extract just that that part, because there are very few projects where client just expect us to come up with insights. Usually, they expect us to come with, with something bigger and the creation of this ecosystem around the insides.

Interviewer: But for this overall greater service, you are financially reimbursed, is that correct? Do you find that financial resources play an enabling or restricting role in partnership establishment efforts?

Respondent: It's the core of course, we are a private consulting. So, we are not working pro bono and of course, in I mean, in some cases, we are working pro bono, for example, when we want to do let's say business development, like any kind of consulting firm is doing when we see a greater opportunity, when we really want to have the impact, and we have the capacity to offer services, in some cases, we are working pro bono. But of course, the overall activity of the firm is not to do pro bono work. But in some situations, we do work for free, that's for sure. So don't get me wrong on that side. But yeah, we need to pay our salaries, like every employee of Dalberg needs to get a salary from Dalberg. And for that reason, we basically need to also be paid in exchange of the new services that we deliver.

Interviewer: Moving on to a different area, because as you said, sometimes you get a project and you have to reach out to private sector donors to specifically request data for this issue. Do you find partnerships easier to negotiate if a targeted issue has a strong symbolic or societal importance at the moment? And we know that most of the issues that you deal with do but as an example, we saw a lot of data collaboratives, and data sharing initiatives pop up really quickly around issues of COVID-19, for example? Have you encountered that kind of impact in your experience as well?

Respondent: Totally, yeah. So, the financial incentive that we just discussed is one important part. But then, of course, the mission alignment is in its own an incentive for stakeholders to partner. So, some people will do it because they are required to do it. We've been seeing, for example, countries where there is a COVID-19 pandemic, and you are a telecom operator, you are forced to share your data, if you don't do it, the country will just remove your license to operate in the country. So that's more of a heavy-handed approach to get the data. In other cases, it will be let's say, the Ministry of Health reaching out to the telecom operators being like “look, there is this pandemic, you have a lot of data that can be used to answer burning questions, could you share it?” and of course, it's a bit of a win-win, there is alignment on the mission, because I mean, every large company wants to solve this problem for their own benefit, because no one is gaining from lockdown, but then also for the impact on the society, right. And you have more and more of CSRs, for the private companies. So, this Corporate Social Responsibility initiative, that those companies are running, sometimes just for their image, but also sometimes genuinely for the potential positive impact that they can that they can bring. So, it's never binary, of course. But it's important, definitely, the alignment on the mission is an important incentive to facilitate and catalyse the establishment of partnerships.

Interviewer: On a kind of related note, do you find a national or cultural context influence this forming of partnerships at all? Especially considering your operations in different countries?

Respondent: Well, what culture doesn't impact? So of course, yeah, culture impacts everything and impact the establishment of partnerships as well. Yeah, the culture around the change, the culture on being data driven, rather than assumption driven. All the all those are just like examples. But yeah, it has a big a big impact on the way partnerships are run. It's indeed, indirect, right. So, it's not like as such on the contract of the partnerships, but it has an important influence on the way those partnerships are designed, the way they are run and the way they are sustained on the longer term.

Interviewer: For example, I was told by in other interviews, that it depends of how the culture is in the country, in terms of where you go to ask for assistance and what kind of bodies you go to, being a consultancy, I think might help with that. Because they said, for example, people will go to a consultancy first before they go to a research university in some instances.

Respondent: Oh, yeah. Yeah, indeed. Yeah, that's that for sure. In some countries, for example, need to have a public a member of the public sector on board to have a successful partnership. In other countries. It's a bit the other way around. So yeah, the perception that people have from private consulting firm, for example, will vary a lot from one country to another. That's true.

Interviewer: What are the major most common apprehensions expressed by public sector organizations or bodies when coming into a partnership with you? If any?

Respondent: Yeah, no, no, of course there are apprehensions. I think there is a bit. One barrier, I would say is, I don't know if we could classify that as an apprehension but part of it is on the digital literacy and the broader data literacy. So, for our project to be successful and the partnership we established to be successful, I think it's important for the members of the partnership, including the beneficiary to fully trust what we do, and really believe that this is a proper, legit and good approach. If you don't get that trust, then it's really hard to go anywhere, because eventually what you are trying to bring is new approaches to solve problems. And that trust is highly linked to digital literacy. So, if you are proposed to, let’s say, bridging the top-notch machine learning algorithm to predict what is the food and security in a country, but the client is actually not trusting this kind of algorithm, then it's really hard for you.

Another one is probably a bit of apprehension on the potential to integrate and leverage what we deliver, and the kind of broader sustainability from a capacity perspective. Will my organization be able to sustain what you've been delivered and provided for us? And so that's kind of a segue to also the dependency potentially on solutions that are not sustainable on the longer-term, for financial reasons, maybe from capacity reasons, because then you leave, and you come with your resources for the financial resources in the capacity but then you leave resource to be able to Yeah, Yeah, the sustainability is definitely important and requires sometimes to establish a new offer.

Interviewer: And do you offer any solutions and support in terms of the sustainability?

Respondent: Yeah, so, at the end of the project, of course, it's a very common discussion that we have with our beneficiary client. I'm saying beneficiary client, because we also have, in some cases, situations where the beneficiary client is different than the paying clients. Like if you were, if you work in context, where the public sector, for example, might not have the money to pay for our services, but has the need, then you can also find a way to bring in some financial big foundation or some bilateral, or different ways to cover the cost of our services. And then for this, of course, you want to make sure that this can be sustained on the longer term. So, something that we do is to help our clients and beneficiary client, especially to think of the way to sustain it. So, to think of the changes that will need to happen in their organization to sustain it, thinking of the ways they can actually find more longer-term financing and funding to sustain the work that we've been doing. So, you can, for example, end the project with a concept note or kind of helping the or beneficiary client to pull together a proposal that they can then be presented to funders to sustain it on the longer term. These kinds of things.

Interviewer: To end off, one last kind of broad question - what do you believe personally, constitutes the most major draw for collaboration with you for partners? And I do believe that's probably differs for that we're talking about private sector or public sector, but just your general thoughts.

Respondent: Yeah, well, I guess there is it's a bit of a mix, there is definitely the Dalberg approach of things that is probably quite unique, bringing a lot of different competencies to the table. We have experts in Human Cantered Design, we have expert in strategy, we have experts in implementation, expert in data science, expert in in in primary data collection. So, we have those different businesses, then we also have a really strong presence on the ground, especially on the African continent and that's something that our clients respect. So, they know that they will work with people who really understands the local context. So even for example, if I'm not the best example of that, as I'm based out of Nairobi, and I'm not Kenyan, but in the office here, the vast majority of the employees are actually people who grew up here and who have a very good understanding of the ecosystem. So, I think that's also something that is extremely important.

And then probably just the excellent, I think that the firm itself has very high standards. We always try to speak and that comes for us from the recruitment, recruiting the right people, and then the passion and the motivation for the work that we do. So, I think there is also a certain level of trust in the market, on the quality of the services that we deliver, the reputation indeed.

Interviewer: Thank you, that was all the questions I had.

J Interview with Current and previous Project managers from DjustConnect

Respondent A: Hi, my colleague ___ will be 10 minutes late, but he's coming, so maybe we could wait and, in the meantime, we could introduce ourselves.

…….

Interviewer: You are physically located in Ghent if I'm not mistaken is that correct?

Respondent A: ILVO, the institute for Agriculture, Fisheries and Food Science, it's in Merelbeke, just below Ghent.

Interviewer: And what is the nature of your affiliation with ILVO?

Respondent A: DjustConenct started as a project, an ILVO project, I don't know the full name, but it started as a project a few years ago and now in September it will end and for now ILVO is the manager of the platform. But because, ILVO is a research institute linked to government it can't make profit, so DjustConenct can be a company, but it will never make profit, I don't know the English term. So, for now it's still a project and maybe it will be postponed or extended.

Interviewer: Does this mean that DjustConnect team is entirely staffed with personnel from ILVO?

Respondent A: Yes.

Interviewer: And the funding, you mentioned it received it externally from an EU project.

Respondent A: It's the European Fund for Regional Development. So, it's regional for Flanders. But we have five founding partners, also because of their investments the platform is made.

Interviewer: On your website it mentioned that you have data buyers, but since you're non-profit, does that mean those buyers do not pay monetary compensation for access to data?

Respondent A: They're not really buyers, but users, and they need to connect to DjustConenct, and they will pay for connection, but it's like a yearly subscription fee, but it's just to maintain the platform, services, management all the costs that we acquire and it's not made for profit.

……

Interviewer: Do you, we already briefly talked about this idea, but could you give me a short background on how DjustConnect came to be? How did the stakeholders first come around the interest in the idea of its creation, what warranted this focus specifically on agricultural data? And yeah, we already spoke a little bit about your affiliation with the Research Institute, but if you could give me a maybe a little bit more background information.

Respondent B: Yeah, well, DjustConnect came to be, purely, well, originally as a concept - that we saw farmers struggling because they have to fill in different forms, multiple data points, double data points they have to transfer data, there's manual work, there's a lot of administration burden, there are companies which are requiring data sets, which are looking for qualitative performance data sets to build additional business services on, and there was a big mismatch between these different actors and that's my, the idea of DjustConnect came to be. We as a platform, we try to achieve data flow, so as a data highway, but because we are affiliated with ILVO, well it's actually hosted by ILVO, run by ILVO at the moment and we are, in a very small part, part of the government's -we really have the ambition of being cost neutral, so we don't want to make profit, and we really want to make it as a tool which enables the transmission and data exchange between different actors in the value chain.

Interviewer: Interesting. And if you could give more detail about your data contributors as far as I understood it if from your website you have two types of data contributors - either the farmers who can directly give you access to their data or more organizations or agricultural companies who with the permission of the of the farmers, give you data. So, if you could give me background of what kind of actors are your data contributors?

Respondent B: You've got that a bit right, is basically anybody who has data can contribute, in the whole value chain. There could be retailers, there could be any cutting machine companies, there could be service companies, it could be veterinarians - it's really diverse. But what we try to achieve is that once we access data, we make sure that everybody involved gives their permission for that data transfer. So, we really want to respect the farmer, or anybody in that case, regarding data ownership. Try to make it as fair and transparent as possible, that's why we also have multiple actors involved for a single data transfer action.

Interviewer: Okay, I understand that on principle, data contributions are open to a wide variety of organizations, but in practice, and I don't need concrete statistics but on the general, what type of companies do you see more dominantly contributing data at this stage?

Respondent B: It's quite diverse, but it really depends on the size of the company, it could be farmers but it could be represented by a contractor for example who delivers certain services on the farm itself. It's mixed between the farmers and the people that are delivering are providing services directly to the farmer. So that's the machine building companies, that crop raw companies, it's advisories, it's really diverse. That's the main

bunch. They all have the wish and need to connect as well with the retail, and the retail itself would like to be involved as well. But Europe is still struggling to get a decent policy out, which causes some delays and postpone decisions on the retail side, which means we have a bit of leeway regarding or towards the farm and the first service providers, let's say, in the value chain.

Interviewer: In terms of data sets and I also understand that that's quite broad, because I think I've read that it's a very vast amount of documentation and data sets that you gather, but what are kind of the general type of data that you work with, that the data contributors provide?

Respondent B: There's no such thing as a single general dive, it's very diverse depends on the value chain if you going into everything regarding milk, then obviously it's letters and protein grams and bacteria number and etc., and how the cow raised and what feed was she eating etc etc. But if you start thinking of growing potatoes, then all these value points or data points are useless, so there is a totally different value chain. What I can say maybe is what the data is usually used for - what you see as a trend in the whole thing is - it's quite often used as a methodology to determine premium value of the product. Let's say if it's organic grown, if certain pesticides are not used or used, is there an additional chance of, let's say, toxic products that are involved, you name it, things like that. Then these parameters are taken into account to create a premium value to the product. And that's a trend that you see both on the livestock farming as on the crop farming itself. So that's, that's where the data is useful but depending on the value chain, you're talking for the data sets are completely different.

Interviewer: I was going to ask this a little bit later but since we kind of brought it up, I was also going to ask you about the profile of your data users or buyers. ____ explained that it's much more of a yearly subscription, just to cover the costs of data maintenance, not so much that they're paying commercial rates for the data, but that being said, what is the profile of your data users? I suppose there are some retail companies and there's also probably, maybe some third sector, private, public-sector presentation. What you have seen in practice.

Respondent B: Ah, quite often is a private sector, public sector of course it's also interested in the problem there that we see is trust, governments aren't really the most trusted bodies in terms of data capturing. They have the reputation of it being used to penalize the farmers instead of helping them, support them or rewarding them, so there we see a bit of reluctance data sharing, because you need to keep in mind that the permission for data transfers is one of the biggest aspects in DjustConnect. We see quite often just private companies, that there are of very different size and magnitudes, that are looking for data to build or expand their services. That's really the most important proof for data consumption.

Interviewer: But to specify, you do also have, not necessarily maybe the government, but also third sector, public sector data users?

Respondent B: There are and obviously they themselves and a lot of interest, because, just on the administrative burden it could be a lot of improvements that are possible. For example, we have a yearly obligation in Belgium and it's I think it's similar in a lot of countries, but you have to declare where you grow and what crops, so, which means that they have some policies, there's like, if you plant or sign seeds straw, you get additional flour burns there to increase biodiversity you get additional subsidies and things like that. But, you have to declare all that, if that could be automated, there's a lot of administrative gain that you could get in terms of workload that you could take away from the farmer, plus the quality of the data could be improved a lot. The problem is, these are positive things that could be possible but farmers usually start thinking about, fertilizer, and all the limitations after that, or so, it's often the discussion between yeah there's a lot of good to the public sector third sector could do, but quite often, it's negative that at least a perception that is more prominent there, so what we're doing at the moment and it's a lot of effort that's going towards status is spending a lot of time communication and sensibilisation and making sure that we get over that hurdle, and then try to showcase actually things that could….

Interviewer: And to briefly go back because I think that's very interesting point you mentioned, what is the framework or the nature of the agreements that you have with your private data contributors because I think you have some interesting elements, whereby it's not just one time agreement of sharing of data and they can be retracted or changed by the contributor throughout the process if they change their mind.

Respondent B: There are two types of contracts that we have at DjustConnect, one that we are off with and the ones that we are not at all. It's the ones that we say if you want to transfer data over DjustConnect you have to play by our rules, which means transparency, which means press etc. And there we make sure that you're in line with all the privacy regulations, with all the data regulations and you name it - that's legally bound. Then there is also the value of the data itself, and that's the business contract that needs to be set up between the two companies. And what the rules of engagement are there is that it’s basically up to the companies that exchange data, that's freely up to them. It could be one year on or could be just one off, it could be a subscription, that's really up to them. At DJustConnect we work with a yearly subscription fee, so even if you use it only once, it's goint to still cost you for a yearly fee, so that's probably not the best idea. But because our farmers and everyone involved actually has a possibility to retract their permissions, we coach them and support them in finding a business model that really rewards having your permission still there, but obviously if you're not feeling comfortable or you don't want to share your data for whatever reason, then you're still allowed to subtract your or retract your permissions.

Interviewer: So, if I understood this correctly, the data contributors provide kind of their terms of use for their data, which is in which has to meet the standards you provide and then your data users make individual connections with them and then individual contracts and agreements for data sharing are formed.

Respondent B: Yeah, in theory, yes, in practice, not really. It's more like on a collective level. It's not like they're making individual connections, we try to streamline the whole process so we're not connecting to every individual farmer - we connect to farm management systems, and then as a data user wants to collect some data, they come to DjustConnect and say look I want from these farms that kind of data and then we act as an intermediate. Also, the legal activities you connect usually on the farm management level, where you get some additional rewards if you share that data so there's an additional service could be some money reimbursement or something, but it could also be a non-monetary benefit that they get like certainty of purchase of their products or some loyalty fees, rewarding system or you name it, there's a lot of creativity in that area.

Interviewer: Okay, interesting, and once the data has been shared, can you tell me a bit more about the details of the digital infrastructure for secure transport that you provide, because that was mentioned on your website.

Respondent B: Actually, we had an external auditor and we are fully up to standards and things like that. But if you really want to have some details on all the technical security bits that should get you in touch with someone else. We don't store data, and we never do. If you're talking about IKM-NET, that's an application that we have, where we actually showcase how we can ease the administrative burden for farmers. So, to make it really specific, every three years, I believe, in the dairy companies, you have to get it out, to make sure that everything is still up to standards and everything is confirmed regulations, there's a lot of administration things that you have to provide - and we automate the whole process so if you DjuseConnect, and just let your data flow, then the IKM-NET tool collects all the administrative documents that you need and puts it all together. But DjustConnect is a different thing, so that's one of the example applications that you could build on top of DjustConnect, but DjustConnect itself is purely data highway, we don't store nor manipulate data that's purely the responsibility of the data collector/data user. We don't manipulate that data; we just ease or make it possible that data flows between A and B. And that's a really strategic choice that we made there, to not store data because of except apart from all the regulations and things like that, once we start storing data and manipulating the data, then we also become responsible for the data, the quality of the data. The format of the data, because there's also a lot of creativity there, etc. So, and also regarding the trusts, because we're still part of the government. We really just said – "Look, we're not going to store the data which is going to connect for the AVP and allow the transfer of data".

Interviewer: Yeah, that's very interesting. So just to make it clear. So how exactly does IKM-NET connect to DjustConnect? If you share your data through DjustConnect there is a possibility for it to automatically be imported into IKM-NET?

Respondent B: Yeah, I believe so, it's just so IKM request data from all these different sources, the farmers get their request to share the data. Obviously, it's slow because the farmer initially logs in, it makes a IKM request the data then the farmer has to give the permission to connect to all different sources, could be collectively but it could also do it

one off, depending on what he is comfortable with or not. And then afterwards, the documents are connected directly from the suppliers, so it could be invoices could be laboratory reports, it could be cleaning verification certificates, things like that.

Interviewer: Interesting. So, yeah, because that's the next question I want to ask are kind of in the area of creativity of trying to get stakeholders on board, especially with a new project and what kind of offers you provide them so I guess this kind of simplicity simplification of the process of audits, through this application could be one of them.

Respondent B: That was actually a showcase, it's not a strategy to develop all these things. It's quite often if you talk to companies that they have ton of ideas themselves. The main service that we provide and that's where the biggest need currently is, is actually act as a neutral party that allows all these different actors and competitors to talk to each other and to really start working towards that trust those different actors in the value chain. But, it's really also competitive information that could be exchanged, that they are comfortable in sharing some part of their data sets, without having to expose themselves to all the free-market regulation things etc. Because as a neutral body we are acting between, and it could be that you as a data user want to collect data from different parties in the value chain, for example farm management systems there are several out there, you as a company either need to make every single connection with every one of them and then do all the negotiations and then it goes well for one and it goes horrible for the other. We as a neutral body try to be the middleman, that gets all the connections going and make and facilitate that process, and that's where really that need is. And then the second part is that we try to explain what the business values could be for that data, because we see, and that sounds negative but as the best description we see quite some data holders, which have like a huge amount of data available but don't basically do anything with it. And they just stored it for someday, which is ethically questionable, but anyway. And we try to explain them look if you unlock that value of the data, it could be insights that could be generated, it could be feedback loops, it could be triggered to throw out a value chain, and try to guide them through that process. So, and then the last one, and it's also very interesting one, it's actually trying to balance the digital skill levels throughout the value chain because we're talking here about data sharing, but in quite some cases, it's still paper that's put in a cardboard box and then driven to a certain area, so it's yeah, there's some gaps that still need to be covered.

Interviewer: And if I could ask, if you to briefly go into more detail about the creative measures you of getting stakeholders on board I think you mentioned, or certainty of purchases or loyalty points or some kind of like the area of those kinds of incentives?

Respondent B: Again, it really depends on the sector that we're talking about, but for example if you talk about potatoes - If a farmer comes to this trailer full of potatoes to the to the factory, when he wants to drop them off, it could be that the just the size of the potato is not suited for whatever the factory of that day has planned, so it could be the day they're looking toward mashed potatoes, or they want to make French fries. And then there are different specifications of the potatoes need to have so it could be done into is

to drive back and forward to deal with this, which is obviously a waste. And what we try to do is first of all try to match whatever the farmer has fits the needs of the factory and tell them when to deliver, and how much etc. And even better - what to grow. And then secondly, we guaranteed a farmer, even if we screw up in planning or whatever that whatever he drops up at the factory that we'll take it, so that he doesn't waste a trip and his time and he has a comfort of it I dropped. Now, if I now go to the factory, I know for certainty that I get my money and that increases ease of mind, things like that are a possibility. It's just that certainty that could be enough and it's not that you pay a premium for the datasets or anything like that. It's just that that guarantee that he doesn't waste time driving back and forth that's enough. And that's for example, a use case that you have. But obviously, in some cases with dairy, there are premiums that are being paid, if you if you share your data, so they can use that data to improve their product quality, then you get a premium that is paid on top of it. There is also the other way around. Some buyers really work from the system, if you don't share your data, you don't get to sell your product for us. So, it's really different. I usually tend to towards the positive reimbursements. And it doesn't have to be money, as explained, but yeah, not everyone is convinced about that principle.

Interviewer: Interesting. And what are other areas of activities that you're conducting to facilitate this kind of flow of data? You mentioned that you're very actively involved communication around the issues of trust - if you could say a little bit more about what your communication strategy.

Respondent B: It's different levels. First of all, we have everything regarding digital skills, that's something that we tried to cover. If the farmer is still using pen and paper to use all the cliches, then whatever brilliant system you have in place like robots in the value chain, it's not going to work. So, we tried to work on digitalization and automation, from the beginning. And then we have different workshops where we have, where we provide information and experience on how to build connections, because usually they say – “look what if I just send you my Excel file, then you can start using that”. That doesn't make too many people happy. So, we tried to explain to them about what the possibilities are in the ICT landscape, what connection you could build, but also regarding data integrity, formative data, and things like that. And then, yeah, overall, we try to work on different levels. Regarding DjustConnect, how you want to carry your message outwards - there we take the approach of not really train the trainer but the same principle, where we try to use word-of-mouth, so we convince a few people, shown in benefits, and the best approach with farmers that you can have is if they convince sort of each other. They usually tend to believe each other a lot more than anybody standing on a stage, claiming whatever. We work directly with the suppliers, machine suppliers, sensor suppliers, you name it, obviously the big buying companies are the follow up process and factories that we work as well and that's something that we approached directly, but that's usually more a company approach and is less, the feeling part of it - emotions are less involved, it's pure rational business. But for the farmers, it's really, showcasing, demonstrating, trying to find those one or two brave souls which are trying to take a step

forward and then allow them to convince their peers, telling them – “Look, I didn't go bankrupt, it really was a benefit for me etc”.

Interviewer: So, you're saying personal network connections network effects are very facilitating factor.

Respondent B: Literally going to have a coffee at their dinner table. That's literally what you need to do. Sit in their kitchen, walk them through, hold them by hand. That's literally what you need to do. It's not that they're not able to comprehend what you're trying to do, but as a whole country we made them learn that sharing data is dangerous. Now coming there and telling them you should share data even more, to give you an idea how bad data sharing is in Belgium for the fertilizer, and that's just in the local area so in the open fields. Apparently, less than 20% is reported correctly on how much fertilizer they're using. And these are hard statistics so and then here we are telling them that they should share data more. Yeah, we have a bit of a challenge ahead of us.

Interviewer: Could you briefly give just a few key examples, or just one or two, of what have been some of the successful cases or products that are already been produced through DjustConnect?

Respondent B: At DjustConnect we're really in the in the Rampage faze at the moment, there are several that are in the pipeline, which unfortunately are under NDA, which I can talk to you about, but the concepts are still there. So, I talked to you about the ones with potatoes, can’t go really into more detail there but you get the concepts. We have similar ones in certain vegetable areas that we have, we are talking about meat and dairy sector, talked about milk case where you can improve product quality, but also for example the meat pricing - If the farmer has to pay more for the cattle feed, then it wouldn't be natural that in the end, the product prices increase as well, but that's not happening today. Usually, the discounters and retailers provide a discount. Even so, somebody is losing money all the way along. It's usually the farmer, by the way. So, there we try to build some transparency, you see that traceability of information, that's something that's very popular these days. Quite often accompanied with Blockchain technology, where the last person rather of a forum or an animal of a crop is kept throughout the whole value chain and once you get the product in the in the supermarkets, you can trace it back to where it has grown out of cow or cheap or whatever, how it has been raised, etc, etc.

Interviewer: Interesting. But so far, no major projects more with the public sector as the end users, is that correct?

Respondent B: That's the fault of the public sector in Belgium itself, they're brilliantly horrible about privacy. So, we would love to enable that, but they are their own worst enemy. We need data from them, they need data from us and then there's a whole body of privacy regulations and to just people which are obviously doing a brilliant job but making it super complex. So, yeah, they're currently their worst enemy.

Interviewer: Interesting. Well, that was the main questions I had for you but lastly, maybe if you can think of any other measure or factors that are influential in either your data contributors or data users drive to work with you, the advantage they see, what value you give to them that I didn't specifically ask about that you would like to add.

Respondent B: What we see as most, and if we compare to other initiatives abroad, it's really that cost neutrality - that neutrality as a whole, these are things that do it. And we really have the ambition as to stimulate the sector, because the main hurdle, every time is not really the business case, if you starting up quite easily you see where the money is that the main hurdle every time, we need to work on these cases is to build that trust relationship. And it's not really between customer and supplier that work directly, that's fine, you get to know each other, that's all right. It's the next party in the line. So that's for a supplier that has to transfer data to an unknown party later on, and that's really where the challenge is, that it's to get to know them. Wherever all directors who are the other actors in the value chain, and why they should share data and make sure that the data is kept safe and secure etc. So, it's really that yet a trust relationship building that's the most important thing, but unfortunate there's no quick win. The other thing, or a general thing that could be done, and should be done is improving the level of digitization skills, because obviously if you're unfamiliar with a certain technology, a certain application, then the chances of you trusting it, if you're already not that trustee from the beginning are very low, so the more that you're familiar with some technologies and certain capabilities, obviously the better it becomes. So, I think those actions towards meditation are very useful.

Interviewer: Just a very quick clarificatory question, because since you mentioned it as quite an important factor, what is your plan or structure for achieving cost neutrality, apart from the subscription fee?

Respondent B: We don't make profits at all. And the subscription fees are only there to provide the operation costs on the system. If we somehow managed to take the cost down, then the subscriptions fees will drop as well.

Interviewer: And is there any other element, or an important element in achieving this kind of cost neutrality?

Respondent B: Yeah, obviously you can be supported by the public organization, that could see it as a strategic thing but turn and twist it how you ever want somehow adequate so consumers are paying for it, if you do it through public means it's just different, it looks different that money. So, we try to look towards ourselves and try to say - “Look, we are trying to visualize for every actor in the field, the benefits of the platform. And if you get a benefit of a certain activity, then there is some value for you and if it can save you, let's say, two nights a week of administrative burden. Then what is it 50, 100 euros a year?” So that's a bit how we're trying to work.

K Interview with Hight Level Management at Positium

Interviewer: In a previous interview that I've had with a member of your team, I was told that your inception story is kind of loosely connected to Estonia winning Eurovision. But, I was wondering if you could give me a bit more background details, namely what influential factors were there that enabled all the necessary stakeholders to come together around the idea of Positium and to get those first few data partners on board.

Respondent: Hmm, yes, so winning division Song Contest was actually like an instigator of Positium in terms that it was a huge event for Estonia, in terms of very many foreign visitors coming to Estonia. And the idea was to, because it was not very clear, say where the people will come, how long do they stay, are they only going to visit the song festival itself or are they going to stay for longer time, and if they come back again in the future and things like that. So, at that time, the only possibility to measure that sufficiently was basically a survey, and with a survey, we could, we could never have, or Estonia let's say, could never have had the possibility to get this extensive information as with mobile-positioning data. So, the idea then was that the three, basically the three guys who established the company, one was a geographer, one was urban planner, one was an architects, so they kind of came up with the idea "People who come to visit still have mobile phones, mobile phones should be tracing the databases of the of the mobile network operators, and maybe there is a possibility to dig in the database, anonymously, just to understand how many foreigners are there, where they come from, and so on and so on". So, this was actually the first idea, unfortunately, we could not get the data for the sound festival itself, but the work continued, a couple of years later, so the song festival was if I remember 2000 – 2001, something like that, and we got the data first data like 2003. And this was the first date and at that time it was processed and analysed by the University of Tartu, so the Tartu university in Estonia, and Positium as a company actually grew out of this company, so we are a spinoff company of the university, research-based company and now we do have a great connection to university, but we will say, act as a separate business unit now.

Interviewer: So, you started around 2003, but as an independent entity I think you were established around 2006, is that correct?

Respondent: Let's say, company establishment was 2003 and four it's a bit different in terms of what we look at, but the 2006 there were first, let's say - hire people full time employees.

Interviewer: And to go back to that kind of process of getting the first data contributors on board and that initial stage, could you give me more details about legally how the agreements worked? Was there a very project by project basis or was it a more overarching framework of wider purpose data-sharing?

Respondent: Oh, in the beginning it was a project based. The first projects were regarding research, so, again, conducted by university, Positium as a company was basically a technical partner at that time. So, again because one of the establishers, or the founder was Professor of Geography and he worked in university, so most of that came from there, and the access to the data again in the beginning it was just project based. First couple of projects were again regarding the commuting of people who live in suburbs of Tallinn, so this was the first project of that, for research purposes, again in the beginning for research purposes, mostly. The idea came that we need to have a long-term data area to better learn the behaviour of people, and since then we do have a running, we've had contracts with the, in the beginning it was one mobile network operator. I think in 2006 or seven, the second one joined in. And that's how it all began. Legally speaking, it was basically a, let's say, partly commercial partly voluntary agreement between university. Our company Positium and the mobile company cooperated. And at some points, it began so that we were able to commercialize a set of data products for let's say different municipalities or governments ministries, or various organizations mostly government based.

Interviewer: And how does this process work nowadays? Have you since developed a general template that you offer to all data providers or is it still very bespoke in terms of how the data is shared and for what purposes it can be used for?

Respondent: Okay, let's say mostly our customers are still governments, but we've changed our business model, from being a Data Broker, that we used to be in Estonia, to more technical software. Basically, a software company and methodology company who are selling our knowledge, and its software product which processes this kind of data to various governments or mobile network operators.

Interviewer: Interesting, because, I've learned from the previous interview that you're have a very international client base, for whom you do reach out and acquired new private data sets depending on the needs and I'm wondering how this process of finding, and communicating with new data contributors, not just the ones that you have very long-term relationships with work both in terms of reaching out to them, getting them on board and also legally if you could give me a bit more detail on that.

Respondent: Okay, so first of all, the customers that we are dealing nowadays, the majority of them are the official statistical offices. So, statistical offices also work together with, for example, telecommunication regulators and different ministries, Ministry of Transportation so on. Usually, we do have our contacts within mobile network operators, but in majority of the cases, the contact is made by our customers, the statistical office or the ministry or whoever requires the data. So, we are basically facilitators, we are the ones who show how the thing should work, and also if there is a possibility to also provide our products to process the data. So the first contact usually, there are exemptions obviously, we have as you mentioned, long-term partners, but let's say if we go into a new country, country X for example, the first contacts, even for Georgia was actually Statistical Office and the telecommunication regulator who already had contacts with

local operators. So, we were introduced, and the basic question is, how they introduce us is - this business Positium, the company who has been working with mobile positioning data, like soon almost for 20 years already, and they know how to process the data, they know how to handle that. So, basic logic is that you don't, they don't have to invent the bicycle.

Interviewer: So, would you say that your business model now completely lacks any kind of an element of data brokerage or it's just has reduced in importance, but maybe still carry a certain element of that?

Respondent: It's still possible, in some cases, theoretically, but the problem is that, because we are a private company, and nowadays for example in Europe, the GDPR and telecommunications regulation e-privacy directive, states very clearly that data cannot be used, can only be used by certain aspects, by certain governments for a specific purpose. This means that it's very difficult for a private company to be able to be a Data Broker. We do have a possibility to work together with the mobile operators directly. So, to say the previous, the default cases for us the customer is mostly the government, who is either we contact them or they contact us, and then we contact operators. There are some cases where basically we either, we contact the mobile operators themselves directly, they are our partners and decide to continue the data brokerage model. However, in this case we are not the ones directly selling the data, it's still the mobile operators. Obviously, it might be so that the data is processed within our infrastructure, our cloud servers for example. It's similar but yes, we are moving towards more like a technological partner and not the brokerage model, simply because business wise it's, it's too complex.

Interviewer: That's interesting. Are there cases when example, your client, the public sector entity approaches you with a particular issue and they help you identify potential privately held data sets that could be valuable to provide insights, because in that sense based on your experience you might not give them concrete access to data but you might guide them towards the kind of right data says to request and seek out?

Respondent: Currently we are focused only on mobile positioning data. So, for us that is a huge data source which provides very wide possibilities to a lot of work with the specific data sets. So, very often we do not include any other type of data directly. Unless it's, there is a specific purpose, so maybe like combining the data, the mobile positioning data with some social media data or something else but mostly we are mobile positioning data specialist and expert sense as our field of expertise, we can suggest in some cases, but it's not our, it's not so that the government comes – "okay, we would like to have, for example, some satellite imagery data or social media data, can you help us?". Our answer is that no we can't, so we can suggest who might help them direct them to other companies.

Interviewer: Let's take a bit of a different direction of your private sector, the public sector clients, how do those relationships and connections form, because I understand, especially in small country such as Estonia where you already have long term

relationships with ministries it's different but how do you outreach, how do you connect with new public sector clients all across the globe?

Respondent: First of all, we do have a very extensive network, international network in terms of different conferences and international organizations. So, we are basically, well, there are only few companies and as a fully private company, I think Positium is the only one company who is participating in a such thing called the United Nations Global Working Group on Data for Official Statistics, you can Google that and find the name of the material and that we do. So that's the biggest, the most international. So, we also work with Eurostat for example, which also combines all the 27 statistical offices of Europe. We have worked with ITU, International Telecommunication Union and so on, and mostly for also for our academic work we do participate in a lot of scientific conferences. That's our main, let's say, access to new customers in different countries, because mostly the conference is wholly held by the UN for example, most of the governments, our potential clients are there, so it's easy for us to access we are known there, most cases we also present in these conferences so that's our main access for that. But obviously there is a lot of direct work in terms of we do have some partners in the countries or in regions, who help us facilitate the contacts with the governments, individuals who need to contact so quite usually quite high-level, it's usually the minister or head of a statistical offices, director generals and so on. But I would say that the most, the biggest work we have done for this are conferences so we are internationally quite well known in this field.

Interviewer: And I think in the previous interviewed it was also mentioned that when it comes to new partnerships you always have also have a partner on the ground, who's more well versed in national context to help you more with the operational side of the partnership.

Respondent: Yes, this is the case. We're still a small company in terms of, globally speaking, also in Estonian case but it's, so for us it's at least in the beginning, it's not an option to open a subsidiary or hire new people specifically for the case. Most of the work we can do from home in terms of providing the software and developing and providing the service, providing the services we need. But in terms of local operations, even in some counties we might, let's say legally in order to sign the contract with the government as a ministry or statistical office, you need to have your own local body. And this majority the case and obviously it's the local relationships also matter, and it's working in different time zones, we need to adjust that. So, we do have mostly partners in these countries.

Interviewer: When it comes to negotiating this new, new collaborations, client relationships, what are in your experience the most common apprehensions if any that are expressed by stakeholders in the process and what kind of approaches do you take to mitigate them?

Respondent: Well, I guess the biggest, always the biggest concern is that because the mobile positioning data is privately owned, data is not open data is not government owned, and usually there are no specific laws in the government in the country to access

this kind of data, so they need to overcome the challenge of getting access to data. That doesn't mean that they can get the data let's say, get all the data from operators to their servers or to Positium's servers. It might also require that maybe, possibly that the data is accessed but it still remains in the infrastructure. The mobile network operators just deploy our software in its process, they aggregated results in terms of number of commuters or number of foreign, so basically indicators, come out to that. But, access in any way it's the first and main issue. And lots of additional European level, for example, almost all countries are still struggling to get even access to the small data, highly regulated. It's very conservative in light of GDPR and the new E-privacy directive it's more and more problematic. So, in European the statistical officers for example, are waiting for a new directive by Eurostat that specifically allows the use of privately owned database, such as mobile positioning data, but there are others also. And other countries, like restrictions are different, the privacy protection is different, and sometimes it's also the case that it's behind some individual ambitions or local conditions. So maybe there's places there, that interpretation of the legislation or someone's wishes that I know for some reason are not allowed to access the data and so on. So, it's an interesting problem but this is the first problem. The second problem usually is regarding more like methodological and technical issues, because it's a huge data set, it's, you know, it's normally an enormous dataset and in order to be able to process the data, and that's where we come in with our experience because we have also seen the cases where the countries have started using the MPD for, or at least researching everybody by themselves and have come up with a conclusion that it's not usable for some reason. Usually what we see is that it's methodologically challenging, and it's quite difficult to do that so that's why we exist and that's why this these problems, give us our bread so to say.

Interviewer: In regards to the first problem, do you have a general approach that you use to help tackle this kind of legal ambiguity on the ground?

Respondent: Usually, it's mostly behind the need for this kind of data. It's not known, in many cases how data can be used and for how many different ministries, different domains, the data can be valuable, very valuable. And if we get, usually if there is a concept or understanding that this is basically a strategically important data source for the country and all necessary players. Let's say, usually the most necessary stakeholders in the country who have to deal with project is the statistical office, telecommunications regulator, Central Bank for example for some specific reasons, and obviously the mobile network operators themselves, and any kind of privacy regulator. Usually, there might be a separate entity that might be under some ministry. If they are all involved in this, they understand the needs and why the data can be used and if they solve a specific problem. Usually, the legislation is not the problem itself because as I mentioned, in most cases, the legislation is just vague, it's not specifically allowing or forbidding the use of the data, it's a matter of interpretation. And in some cases, which just requires the change of law in order to be able to, for everybody to understand that okay we are now allowed, so this has been in some cases where the counties legislation has to be changed in order to be able to access.

Interviewer: So, would it be correct to say that in such situations, past successful use cases that you've developed are very kind of valuable tool, and negotiating the stakeholders on board?

Respondent: Absolutely, absolutely, it's something that we also provide, our value that we can provide examples, and obviously the international community in this level is also quite terms that they can, if we can provide an example of, okay, this country used this kind of access method or access model, we made it, put them together even describe - they find the similarities and potentially use the same example.

Interviewer: Along the same levels, would you say you personally have experienced or leveraged the strong symbolic value of an issue you're trying to tackle as a means of navigating this process? For example, we saw around COVID-19 A lot of private-public data sharing relationship emerge, especially in the mobile sector, quite much more quickly than they would have in normal times. Maybe you have experienced something similar pre COVID around other issues you've worked with.

Respondent: Well one of our, let's say semi-partner semi-competitor company called Flowminder it's a research company mostly NGO actually. They have done a very fast response project for example in Haiti earthquake. If you Google height the earthquake and mobile positioning data I think you can find a lot of information on that. Same thing for, I don't remember the year but in Nepal earthquake. So natural disasters, absolutely not something which was together more than mobilizes all stakeholders and progress can be very fast. You mentioned pandemic. I think the same thing we've seen here in Estonia in different African countries, not and not only COVID, let's say Ebola a couple of years ago, same thing again, the data was accessed quite quickly. There were some decisions made on this data and I think they were successful and, in the end, saved lives. The problem is that, if it's a normal situation and everybody has time, basically there is no pain to get the data and unfortunately, it's also the same that once it's. I'll just say, it's not very clear for our stakeholders, the users of the data, about the value of data before they can get access to it. Usually when they get access, they can see the value and it's something that they become more and more hungry for and we can see that, what other possibilities can be done in this case. Our experience is that, yes we show them the examples from other countries but they are not 100% clear, they all always want to have, let's say the specificity of the country and how the data can be used in their country and obviously the first question - will the data for the county actually give value? We know from our experience 100% sure every county can benefit a lot from this data. Given that you know the methods are well developed and administrative and legal framework for this access is sufficient.

Interviewer: Interesting, and more on the technical side and the methods I wanted to ask you, because I read on your website about the Positium Data Mediator and I wanted to ask you if you could in more simplistic terms, for someone without much technical background and data analysis, what is the unique value provided by the data mediator?

Respondent: The mediator is basically the software that processes the data from mobile network operators from the beginning from the raw, billions of points on the raw data until the end to very specific statistical indicators describing some specific domain, so let's say the number of tourists visiting specific locations in the country, for how long the tourists are staying there are very different breakdowns - spatial and temporal. So, it includes, I think it was more than 94 Different methodological steps of processing the data itself. So, cleansing of the data, changing the data model from the simple points, which is the original data format into a, we call it continuity data model, which describes the flow of people in terms that you can estimate or extrapolate for all subscribers in the county where they are at specific points. Again, it includes also the anonymization of the data and things like that so it's a, it's a complex methodological software that processes the data from beginning until the basically the end where the data can be used.

Interviewer: Interesting, and yeah so far as I understand, it's been very long-term development in the process. Because in the previous interview it was mentioned that it contains all the experience you had over the years and that's what kind of guarantees the high-quality outcome of the data, every single time.

Respondent: So that's correct. Okay. Software is obviously not ready, it's going to be developed in the future, internally. But basically, it's, I think it's well said that it's combines experience for us, I've been working procedure since 2006, I was one of the first place. So that's, that's my experience in the software and there are many more ideas or areas where we can still develop the software itself, still get new value from the data as presented so it's not, it's still cosmos out there a lot of possibilities.

Interviewer: I had one last closing question, that's kind of a bit more general and you may interpret however you want - what do you believe constitutes most major draw for collaboration or partnership with is it to for your partners or clients as you referred to them?

Respondent: I think it's, it's basically the trust, in both ways, and mostly the customers trust us to be able to take to be the best experts who know the solutions of the problems is a significant data source, any kind of statistical data source or statistical data sources have issues, and our experience, basically guarantees that most of the issues we have already encountered, and we know the solutions for them. And if there are new issues then we are probably the best ones who can find a solution. If there is such solution at all.

And in our experience, it's just, we're from Estonia or businesses to be trustworthy in terms that we have to deliver and we have to be there for our customers because we have seen a lot of cases where, as I mentioned before, somebody tries to build their own system or start from scratch, invent the bike and then they, they're disappointed. They say - okay this data source is not working at all, we're not going to deal with that, then, kind of, we bring in our experience and our software and they can see the value in that, and we all have this background of our research and academic background plus participation international organizations, just gives us the possibility to be trustworthy and that's

something that almost our mode so that you know we have to be able to work with our customers so that they trust us.

Also, simply, even for the model that it's a highly sensitive data source. We haven't had any privacy related security issue in all of those years, we are very strong on our IT security, and the people who work in our company. So far, it's all been that they are very well trained as experts in the area, and obviously the communication with the client is something which is I think one of the most important things. Plus, it's the innovation. It's pure innovation at this point still, although it's 15 years already, we've thought in our plans that all the countries in the world will start using MPT by 25 years ago. It's just so it's innovation. I think the trust is the most important thing right now.

Interviewer: So, to reiterate - it's the reputation, stemming from many years of experience, it's this data security innovation as well as just saving the public sector partners, the resources, time, resources and human resources of doing it all on their own. That was all the questions I had. Thank you so much for taking the time to speak with me I greatly appreciated.

L Interview with Sales Manager at Positium

Interviewer: First of all, could you for the record briefly describe your role and responsibilities at Positium.

Respondent: Sure, so I'm head of sales hare at Positium. The sales and marketing team and I mostly run the international business development of Positium, so that actually entails making the connections to the data providers and also the clients who are waiting for our data to make decisions. So, kind of bridging the topic, private data sharing between the different operators or organizations in the country.

Interviewer: Before we jump into the more particular questions, could can you provide me with some background information on Positium? For example, how the stakeholders first came together around the common interests of its establishment? I would also be interested to know more specificities about the nature of your affiliation with the University of Tartu.

Respondent: Do you know Eurovision Song Contest? Great so you're European enough, and I think George is doing well in it as well. So, the Eurovision Song Contest is actually part of the inception story of Positium. Estonia won Eurovision in 2001, and this was a major event for Estonia. Can you imagine that, we were a new country, I mean by then we had been independent for 10 years, and we were just starting our economy, we had made great strides. We had the Tiger Leap program, which brought internet into schools, and we were making strides in that area of making our next generation more knowledge hungry and able to have the opportunities that the internet provides. And so, when Estonia started to think about having Eurovision Song Contest in Estonia, it was obviously a major decision. We had to host it because we won. We could have also just turned it away but, it was a major investment that Estonian government made, saying that this was our way to put Estonia on the European map. We built a whole new arena to host the Eurovision, so it was a very, very big thing. Now, the question was - is this investment actually beneficial for Estonia? Because, as a small country just coming up, we have to think carefully on this, and do we actually have enough data to understand whether this investment is beneficial, or not. So, that's when researchers in the University of Tartu came up with this idea to gather data on tourism related to Eurovision. And there were no good data sources for this, so they thought, why not use data from mobile phone operators, because people have, especially for traveling, they have mobile phones, and we tap into this data to understand what's the impact of this event. And that's where the idea started. We started talking with mobile operators back in 2002, to understand if we could do this. Of course, they said no, what an interesting idea but no way you're getting our data. So, we didn't manage to do it back then, but the idea stuck and in 2004 Positium was created as a company, to get data from mobile operators on a data sharing agreement. We started with active position data with just volunteers, and then moved on to more passive forms of data which operators already collected for all tourists and for all people in a country.

That's where we are, we are a very research-intensive organization still, we have this background of being a spin-off company at the University of Tartu. And we still have the spin-off contract, so we still collaborate a lot with the university. So, a lot of our methods have some prints of university thoughts also written on them. And many of our employees either came from University of Tartu or are currently even linked with the university itself.

Interviewer: That is very interesting. From your website I knew that you had already 10 years of experience, which was already very impressive but clearly you have more. I had the pleasure of studying in Tallinn at TalTech for a little while so I have had the exposure and some knowledge about the very early digitalization efforts that Estonia has been making. But I still find your example very impressive because in 2000s, this kind of private-public data driven partnerships were very much still uncommon. And they are still relatively uncommon today. So, this must have been very pioneering from your side.

Respondent: Yes, it was, and it was due to Professor Rein Ahas, who led this charge, and his colleagues and friends who came from architecture background and engineering backgrounds. So many different minds came together to start this. And we have Erki and Margus, who are the board of directors of Positium at the moment, and they started very early in the development of Positium, in 2006 already. So, we have people who have been Positium for a long time already.

Interviewer: From the private sector side Positium mainly partners with mobile data providers, and as you said, some of these relationships have existed already for quite a long time. But is this list of private data contributors more or less static or are you and your team trying to continuously expand it?

Respondent: We have always tried to expand it. As you might imagine, it is quite difficult to start the partnership with a mobile phone operator, because they are big organizations, in most cases, even if we're talking about a small country, then mobile operators are one of the biggest companies in that country. And it's continuously expanding to make more use of the collaboration that we have going and giving both the mobile operators some value and ourselves as well.

Interviewer: And what are some of the most typical activities you conduct to search and outreach to new data contributors?

Respondent: Our strategy right now is to is to outreach to potential clients. So, once we have set of clients, or one strong client in a country backing this solution as something that they really need, then we outreach to new data contributors and new mobile operators. And sometimes we don't even actually do it ourselves, but the client does it on our behalf.

Interviewer: This is a good segue to the list of your organization's public sector partners. I am familiar that you have some long-term public sector partners like the Estonian

tourism board, but in other cases, do you take active measures to outreach to new clients or do the clients come to you? What is the procedure there?

Respondent: We work with public sector partners mostly as clients. But, of course, they have valuable data themselves. So, they do benefit from our data, mostly from the data that we collect and process from the mobile operators, but also their own data is necessary to understand what we can do. For example, if we go to a country, and we are asked to visualize or process mobile phone data to, let's say we're trying to understand where the tourists are going in a country, we need also the access to proper reference data, like local administrator unit data. And this is usually obtained from public sector organization. So, land use data, land at risk, the unite, the boarders. All of these are valuable datasets that we use as inputs to our processes. We also use a lot of reference data, whatever we can connect to the results. For example, when we work with tourism there is one picture that we get from what the mobile operator sees, based on how many people are doing inbound roaming in different areas, but then also the picture of that we can refer to - for example, how many people are entering the country from a certain point of entry, and this we use to calibrate and actually check if the data that we're processing is of good quality or not. So, we use data as input, and as a comparison source later when we have the output.

Interviewer: That is very interesting, but to circle back a bit to the process of starting new partnerships with public sector organizations, could you give a generalisation of how they come to be? Do they approach you based on knowledge of your previous work or do you actively outreach to them, offering them help with particular issues? Or maybe a mixture of the two.

Respondent: So, our main ways of finding new interested public sector partners are to be present in conferences, be present in global working groups and other forums that are working on new data sources for official statistics for example or new data sources for tourism statistics. And then we do some outreach ourselves. But mostly we do expect to have some kind of general understanding that we are in this space and then the public social partners themselves could turn to us whenever they find the need. We are working in several countries, and they are not grouped together in let's say Eastern Europe or even the Middle East. It's all over the world. And no groups. That's because the public sector organizations, they have a need, and then they find out who can serve that need.

Interviewer: It’s interesting that you bring that up, because I was actually just going to as what is the difference, because of course you operate in Estonia which is a smaller market so it's easier to create a reputation and create awareness about yourself, but I also saw that you're partnering with Ministries in Indonesia, what is the difference in those cases between national international forging of partnerships?

Respondent: Well, it's quite different you could say, but overall, we like to streamline it as much as possible to be the one that provides expertise, and all the administrative aspects and the local aspects will be taken care of by somebody who is very well versed with local cultures and traditions. So, we use a partner in the country. Or we have

somebody who can navigate those at the client's organization itself, so kind of a champion who can do that for us, even if it's in the client's organization. There are some people who are fanatical about bringing in new data sources to their organization. And they'll be able to help us navigate the local situation.

Interviewer: Okay, interesting. Just to give us like a general walkthrough of how a collaboration comes to be - so a public sector partner approaches you with a concrete issue and you analyse it and then maybe on their behalf or through them, you go to the mobile data provider and ask for a particular dataset for this. How does this transaction work on a business level? I am assuming that the mobile operators are reimbursed for the data they provide and, in such case, what is the chain of reimbursement?

Respondent: There are different models for reimbursing for the data. Sometimes it's offered for free. As you might know from Dalberg as well. There are some philanthropic cases of providing the data. So, corporate social responsibility and providing data for free is something that operators do for pilots at least. Of course, they want to get their costs covered. But if they're just provide the data and the data is already there passively collected, then there's not a lot of costs that they have to incur. Unless the processing takes place in their own servers. So, in some countries it's mandated that the mobile operator turns over their data to the statistical office, and they're able to mandate that because they have a physical act that says that they can ask for data even from private organizations that could be useful for statistics. So that's one way. And then there are cases where we work with a mobile operator on a business plan basis and they expect to get paid. And they expect to get paid so that it's big enough to warrant all the investment in making a separate division that handles requests. And this is usually done when there is already a few interested partners in a country or government institutions.

Interviewer: Since you mentioned Dalberg, they have a mixed method whereby sometimes they get reimbursed from the public sector organization that they're working for and sometimes they do pro bono work. Is that the same for Positium?

Respondent: We also do pro bono, sometimes. So, if the question is how we get reimbursed, usually we operate on a commercial model. But then, once some commercial agreements are already in place and we have room to do pro bono work in a country as well. We might be able to do that. Let's say somebody pays for analysing the whole data set in a country - let's say it's a statistical office. Then within that project, there is a small niche of pro bono work that the data has already been processed and cleaned and verified and the pro bono work would be to take out the data for a new use case that could really help the local situation but that there's no current funding for. So, we might be able to help there.

Interviewer: A more of a general question, that kind of overviews what we've already said, but what do you believe constitutes the major draw for collaboration with Positium for your partners?

Respondent: We're one of the only ones who could do it. We're some of the only companies in the world who take mobile data and make it useful and reliable enough for national level statistics that people can trust. So, there are a lot of pilots in the world, and there is a lot of collaboration between organizations that we can think of as our clients or partners, and they collaborate with mobile operators directly or with another organization, and in many cases these pilots do not work out or they don't operationalized for running production. That's where we come in, we actually have the wherewithal, we have the appearance to keep up the regular production of statistics from mobile phone data, whereas others can mostly focus on pilots and research, and trying to do one off projects. So that's, that's where I think the value is.

Interviewer: From your proposal perspective, what would you say the partners see as more of a draw – you long experience of operating in this field, your human capital in term of the expertise you offer, or I've read about your flagship product the Positium Data Mediator? Or maybe that's also a different unique value proposition?

Respondent: Yes, the Positium Data Mediator helps us to do what we do. That's what creates reliable data month in month out, and it has, you know most of the experience that we gained from 15 years already inside there so that we don't have to implement it for every specific project. And the clients can also sleep soundly that most of the issues have already been thought of and that's where the experience comes in. And I think our clients see it. And you have to ask the clients how they see it, but I think that it's because we have all this experience we know how to work with the data, and navigate most of the pitfalls that could lay ahead.

Interviewer: When communicating the potential benefits and the value that can be derived from a partnership with you, what are the main challenges or issues that arise in the communication of this with partners?

Respondent: I think one of the challenges is thinking ahead and making sure that the data is of good quality, even after the first product project. Sometimes this means that we have to work hard on the quality assurance, you know the term garbage in garbage out? In data science, whatever you put in as input data, that's what matters when you take out the results. So, when you blindly apply any of the data science methods on faulty data, then what you get out can be pure garbage. So, we tried to minimize the effect of luck and do as much as possible to understand the quality of the input data before we go and apply all our technology. That makes the output also very reliable. Now clients, because they haven't had experience with mobile phone data, they don't have the experience of failing with data. So, then they don't want to necessarily put a lot of money into the first pilot projects before they can get some kind of value out of it. So maybe that's one of the main things that we need to communicate, and sometimes it's hard for them to understand. And that's why we actually worked with statistical officers, a lot, not only because they might have the mandate to request private data, but also because they speak in the same terms, they know that in order to get quality output you have to have quality input. So, with them we can talk about the quality matters. With the Ministry of Tourism or even a research

organization who is working on a specific project they want output, that's what they need and they don't care what goes in, and sometimes they're taken aback when what goes in, it can be quite resource intensive.

Interviewer: And do you believe that personal professional connections or word of mouth, in your experience, contribute in any significant way to the establishment of new partnerships?

Respondent: Especially in smaller countries yes it does. So that is how we established the first partnerships and professional connection that is also personal to some extent. And we also try to make personal connections to the partners that we already have, because it makes communication that much easier. And word of mouth, of course, happy customer is a referring customer.

Interviewer: My next question we already touched upon briefly when we were discussing business models and reimbursement. Do you find that financial resources meaningfully enable or restrict your partnership establishment efforts and if yes, in what way?

Respondent: Yes, of course. So, if it's the client, they have a specific need and there is a value attached this need. If they really need the data and if there is a budget, if there's a will there's a way, if there's a budget there is a way. And in such cases getting a new partnership up and running is much easier, rather than hoping for social responsibility aspect. Usually, those corporate social responsibility projects they end up as one-of pilot projects, and they have research benefits and can provide some insights on how to continue, but sometimes we're talking about the graveyard of pilots, and sometimes this graveyard of pilots is filled with under-resourced projects.

Interviewer: And lastly on a slightly different note, have you found that strong symbolic or societal importance of a targeted issue, the issue you are trying to address through the data analysis, meaningfully enables or restrict partnership establishment efforts? For example, we saw that in response to Covid-19 a lot of data sharing partnerships emerged quite faster than they maybe usually would have. So, do you find that when the thing you're trying to address has very big societal resonance at that moment that it makes partnership development, client acquisition or especially the process of negotiating with data contributors easier in any way?

Respondent: Yeah, actually that's a great question. It really does. Showing how the data links to input, input to the society, how it benefits society and people in it is actually very, very important. Not only to us, and not only to clients, because our clients are mostly government organizations. But also, the mobile operators who are torturing us for data sharing, because if they can see the impact that they're having it not only enables the partnership establishment effort, but also might make the budgets smaller as well. Meaning their budgets that they require to share the data, and they might be able to survive with just covering their own costs, because it has such, big importance. And we saw that

with COVID-19, where most of the operators in European Union, at least had some kind of project in the world with health mysteries or other organizations that needed data on how people are moving about, if they're adhering to mobility restrictions, where people actually staying during this, during this pandemic. And most of this was because of societal importance, it was volunteer work.

Interviewer: Just a quick additional question, this wasn't on the list but - apart from the social responsibility, social philanthropy and the reimbursement, do you think there's any other benefits or advantages for a private company to contribute and to cooperate with you and to provide you with data?

Respondent: Well, we like to think that they get the benefit from the processed data as well. So, the monetary benefit could be either paid directly or through a business model. We have similar goals to make sure that data creates value in the market. But there might be also the effect of us going in and putting a lot of experience, and the software in place, which generates data, which they can, in turn, use in their own organization. But usually, I feel that this reasoning does not resonate too well because operators work in silos - the one that is providing the data, they might not work, it could go towards the goals of another organization or silo that actually could use this data. For example, if we analyse roaming in a country, then we could provide this data back to the roaming department saying that you have issues with roaming connections here or here or that on some days somebody is actually hacking into your network and scamming you for roaming charges here. But because we're not actually talking with the roaming department, then this might not go in. There might not be a connection there.

Interviewer: And in the process of forming this collaboration, or acquiring clients as you would say, what are the most common apprehensions, if any, expressed by potential clients? What are their fears, around participating in the data exchange? And what are the approaches you take to mitigate such concerns? For example, of course, the primary one that always comes up is the issue of privacy.

Respondent: Yes, privacy is very important. Security is the solution to privacy concerns, to deliver things secure. Even if the data is not moved out of the mobile operator's office, it's still very much a concern. But I would say that there are technical solutions around it and as long as we don't talk in legalese, there are always solutions to that.

The other barrier is the budget - whether we can get enough budget to do good job, you know, for the first project, from the outset. I already talked about these issues. They might want to see a win value from the data before they actually invest a lot of money into it, making sure that to quality is also okay. And there we provide a case studies and descriptions of how we actually know how to make sure that you can get value out of this, but in order to reach those milestones, when you can actually make decisions, you have to invest and invest in data quality as well. Otherwise, the value might be short lived. So, using case studies and experiences from other countries. It really helps.

Interviewer: Thank you, that was very interesting and those were my main questions. Just two quick concluding questions, so that I don't take up too much of your time. Firstly, apart from what we just discussed, are there any other aspects that you believe are relevant to your stakeholder linking or engagement practices and processes that I didn't ask about, that you would like to speak about additionally?

Respondent: Yeah, there are, there are many, many aspects to having stakeholder engagement. And I would say that it's, it's somehow, you can you can formalize it and somehow, we can make those things work in most countries but not in all countries. Making new data partnerships is still quite difficult for us, even with all of our experience. And we'd like to think that it's, you know, it's times that are changing, and we can move to a new world where data is becoming socially more important, and we can discuss about the value that the data can bring, rather than the privacy aspects or the threats. There are two aspects of privacy that I want to mention, it's not only about personal privacy but also business interests of telecom operators as well. So, any data that private operators have, it has some business interest included. And whenever you make outputs out of this data, then you have to make sure that not only is it anonymized, but you're not, for example, giving away how many subscribers they have in different areas of the country, which could be used by competitors.

Interviewer: Yes, that is a good point, that has been brought up in other interviews as well. Just briefly, because I just recalled that I wanted to ask you this as well - I saw on your website that you're affiliated with the World Tourism Organization. Do you find that the participation in this kind of institutionalized way of networking has helped you, maybe particularly in establishing partnerships outside of Estonia? How valuable do you think this type of memberships have been for your client acquisition efforts?

Respondent: Yes, it is. Being part of these big international organizations is very valuable for us with UN WTO. Of course, we also have some projects, there are also other organizations where we work together with, including TSM illustration. Actually, can I share my screen? Okay, so these are all the countries where we worked in and analyse data. So, there are quite a number of them and some of them are also pilot projects I have to say, but there are a couple of countries where we regularly operate even to this day and because we have a good bunch of partnerships there we hope to operate there for a long time.

Interviewer: Interesting. Can I find this information online on your website or maybe I could kindly ask you to forward me the information?

Respondent: Yes, I can send you this. I'm just saying that UN WTO is not the only, in all of these organisations

what they want is to teach and give guidance to countries that new data sources are very relevant and that they should be used and put as much effort as possible to get input from

experts, just like us. So, we work together to enable countries to have access to new data sources.

Interviewer: And I see that you're referencing to the Sustainable Development Goals. Probably that's a tool you used for communicating the societal value of your work and maybe also to link to the corporate responsibility, I would presume.

Respondent: Yes, well, and it's not only for marketing purposes. This is actually something that our clients care about, especially National Statistics Offices, but also Ministries. We're all working towards these goals, and the more we can contribute in measuring what we still need to do, we reach these goals better. And especially, you know, number 17 - partnerships for the goals. So, we are just making sure that we also contribute what we can to reaching those goals.

M Interview with Senior Data Scientist at UBDFC

Interviewer: Could I briefly ask you to, for the record, describe your position at the UBDC and your responsibilities?

Respondent: Yeah, of course. So, I work at Urban Big Data Centre, we're based in University of Glasgow. It's perhaps easiest to describe my role by first talking very briefly about what UBDC does. We have two primary functions – the first is to form a pretty traditional kind of research centre function. We're looking at issues around cities and at all ranges of different issues through the lens of big data, particularly through some new and emerging forms of data. The second function, which was the original function of the centre, was to provide on behalf of the Economic and Social Research Council a national Data Service, principally aimed at academic users. The goal being to, through a variety of means, to reduce the barrier to entry for many social scientists interested in exploring issues around cities with the use of data and data driven methods. So, I'm the senior data science manager within UBDC, I lead a team of data scientists and technology and information professionals within the team. We have a responsibility to support with our own research agenda, but also to support the needs of end users. So, in leading that team some of the things that keep me particularly busy include talking with prospective users of our data services, talking to data owners, and seeking to negotiate licensing agreements to get data, as well as working with researchers to develop new methods and identify interesting novel uses of data to support their work.

Interviewer: I wanted to ask you for some background information of how the UBDC first came to be and how the stakeholders first gathered around this kind of shared idea of its creation, to the extent of your knowledge.

Respondent: Yeah, well, the origins of the centre preceded my involvement. But I certainly know that the UK government investment in this area, there was an obviously prioritization of the use of technology and data driven methods across the board. And I think there was historically less culture and less capacity within the social sciences to do so. So, the ESRC were keen to support our drive and interest and capacity in that area through a number of different investments. And there was an original kind of program of work, can't remember what the acronym stands for precisely but essentially, they also invested in a couple of other national data services. So, we're principally focused, obviously, on urban and city issues, and each have their own sort of respective thematic communities of interest and whatnot. So that took place, obviously, within Glasgow University, where I'm based, there was an enthusiasm in responding to that call and establishing a centre, as well as a kind of an appreciation to think of the value of having an associated research agenda. So whereas we are now formally support from ESRC funding to do this research function, there was always a kind of an aligned and very

complementary research agenda going on within the university and the various different schools and disciplines that were affiliated with the centre at Glasgow University. So yeah, that's kind of my understand of how that came about. The current story is we've had a period of initial funding, it was a five-year period, from 2014-2019. Then midway through that we got confirmation of this center funding, and subsequently, we've been awarded some continuation funding for our data service function. At this point that continues until just April of next year, but we're hopeful that we'll be able to sort of sustain that for even longer.

Interviewer: Okay, great. And then to move on, I think, believe you've briefly mentioned your expertise and more dealing with private sector data contributors. So, in that regard, before we jumping to more specific questions could you give me a general walkthrough of how the process of collaboration comes to be?

Respondent: Yeah, sure. The thing I probably always end up saying up front is thar there's no single kind of model here. It's often quite dynamic. But, in broad terms, we describe ourselves as, as far as our data service function is that this is a research driven or research led data service. And that means that we don't seek to be exhaustive consolidators of data, that's not really our goal. Rather, we appeal to our wider research community of users, to our own internal researchers, and by reference to some of the strategic partnerships that we're developing to as a means to identify what data sets may be potentially of interest. And sometimes that can be very specific data sets that are sought by particular communities of researchers that communicate that to us, sometimes it can be in the form of insights or indicators that are of interest that we explore - where might we be able to produce some of those, what data might be available to produce those. But generally speaking, this leads us to conversations with individual data owners, and we have an initial conversation about what are the data products that are available, and how might those be accessible, and sometimes we have ------- purchasers and owners that are, their businesses are very much aligned, and driven by this whole model of providing access to data.

So, there it is in particular, increasingly, there are companies that will sell your data, and that is their main kind of raison d'etre.

Meanwhile, there are other companies who are doing things and as a by-product of that activity, they are producing data that will be of potential interest to us. And they may be more or less mature in terms of being able to value that data and to provide a business service for its delivery. But it’s based on this, these back-and-forth conversations. Our needs are maybe not typical, that's quite clear, we have to be able to, as a minimum, be able to offer a value for a potentially widespread community of users. And our goal is essentially to say we can’t come up with a licensing agreement, can we come up with a deal, which provides not only access to the University of Glasgow or to a single user, but rather, it gives us the means to effectively sublicense that data so that a wider community of users can enjoy accessing it? And obviously specifying the limits and the permitted purposes and usage, to ensure that our primarily academic audience is able to use that

data, to access it, to use it to publish, to do all the other kinds of impact activities that are rolled into academic work. And then that leads to licensing agreements being agreed, sub licensing agreements, and protocols and frameworks being agreed. And then we have an internal governance process too, that runs concurrently, where we have to sign off and approve of any proposed acquisitions, not just in terms of value for money and occurrences, but also the likelihood of being able to offer scientific benefits, being able to offer more widespread public impact benefits, these kinds of things, we have a whole set of criteria we apply when reviewing and approving any acquisitions.

Interviewer: And as far as I'm aware, the list of the data donors that the UBDC, works with it's not static, it's evolving. I think that recently I saw on your website, a call for researchers wanting to work with a new data set that was just recently made available to you. And as I understand you're constantly trying to recruit new data donors. And can you explain to me how this process of attracting new data donors or new datasets works? Do you primarily go to already established partners or if you seek out new partners on what basis this happens? And what kind of recruitment methods you employ?

Respondent: Sure. Well, I think there's, you know, continuity of supplies is a very, very attractive thing. And certainly, there has been a number of occasions in the past where we have established new relationships. And companies like Strava, for example, that we've worked with, through several years of the centre's life, who provide active travel data based on the usage of their app that athletes and cyclists choose. Yeah, continuity is very good. And then when we can develop a sort of longer-term partnership, it provides value - every new data set is complemented by previous data and as you can imagine a whole range of work. But researchers who wish to benefit from having a kind of longitudinal data set - that's not always possible, though. And certainly, companies will sometimes change their business priorities or business models, or new products that they're developing will emerge, which means that they're perhaps less enthusiastic about maintaining that kind of relationship they've already established. So, we're very conscious of the dynamic nature of these markets, and of the importance of not being too reliant on any individual data providers. And we are very keen to work with different kinds of providers, often one of the biggest challenges is just starting off that conversation and finding an entry point into someone within these organizations, who's able to have the conversation. And that's especially challenging when you have a company or a data set or a service that is producing data you intuitively think might be useful, but they don't have a provision to make that available. So, a recent example would be, we'd be looking with a range of different academic and also public sector partners about mobility. And then the broader sense of how do people move around cities and narrowly located and dispersed within cities. So, we, last year was to undertake a pretty wide-ranging exercise to identify datasets that would potentially be of interest. We eventually settled on two providers, called Huck and Tomoko, and they both work on similar kind of models whereby they, they've partnered with a number of mobile phone app publishers and developers, and they're able to get from those apps that have location services enabled details of users of those apps locations into them, translate that into a dataset that covers

a whole range of different periods. But, before we got there, we were trying to speak to a whole variety of different providers and the starting point of that conversation – we piggybacked onto some work that already been started with in Glasgow City Council. I've been speaking to various vendors, I don't know if I can see who they were specifically, but we were looking into the sort of companies that operate at the kind of intersection between mobile app publishers and advertising agencies. So, there's these companies that are basically called mobile app exchanges. And we tried to reach out, we identified several and tried to reach out directly and they were not at all receptive to our request. We knew, they were basically geared up, their whole communications platforms are geared up to accept business from mobile app operators or advertisers, and for academics who are interested in accessing these data, it was quite, quite tricky. So, get that we're always on placement in very practical terms, places like LinkedIn, you know, trying to find people who appear to be placed within their organization to have the means and the to be able to understand the request, the new opportunity. Or in other cases, as it happened with how Huck and Tomoko, there was already, I think we found our way to Huck through other researchers elsewhere that we were aware of that had used their data set. With Tomoko, it was a recommendation from someone else. But actually, both those companies are very much established and geared up towards the provision of data. And that includes for academic purposes. And in this case, for kind of government operational policy.

Interviewer: It brings up a great point, and I'll return back to it in a second. But before that, I just wanted to clarify something. As far as I understand, the way this your centre works is that there is research agenda and concrete issue you're working on. And then maybe you reach out to tried to get the data sets that will be useful for those purposes, rather than acquiring the data sets first, and then analysing them and seeing what can be gained from that? Or is it a mixture of the two?

Respondent: So that's probably where it's a bit blurry, this line between. There are these kinds of dual functions, and obviously, within our centre, we do have a research agenda, but that's I said, our kind of research centre mode, we have an agenda that part of the research proposal to ESRC, in that respect that we're committed to. And that had alongside a number of data needs that we could either satisfy through some of these kind of licensing agreements or indeed through other forms of collection of online data, for instance. With the data service, we're not bound by limits as far as specific agenda that we can support. Rather we are a service, we are a part of the research infrastructure that is available to reflect the needs of that wider community of researchers. So, if some researcher approaches us and says - listen, I'm really interested in something about the agility, about labour markets or skills, for instance, then we will be very responsive and seek a data set that supports that use. Now, in some cases, that means we can get a license agreement, which means that we can provide data and providing data, and trying to specify the data to be as broadly useful as possible. That's sometimes a challenge. Obviously, as you can imagine that a researcher will come to us with a very specific needed, perhaps a location of interest, or a time period of interest. We will typically

maybe broaden that or try to generalize it just to enhance its broader usefulness. And whereas that researcher will then be able to do some work, which is based on questions they've already had, that they have a good idea that data set will help them to answer, it will also free up that data for more potential users whose needs are more about "let's have a little play with the data and then see what that can tell me about whatever the questions may be". Sometimes, it can it can be the data that informs the question just as much as the question of one's choice of data. Obviously, as I said, we don't seek to provide an exhaustive or definitive collection of city data. But where we do have collections, that have been prompted by specific needs, we try and configure those and specify them to make them as widely useful. And it's certainly important that we have a kind of good balance within the collections we have available across the areas that we either intuitively or with our experience of communities of users think will be reasonably balanced - a sort of spread of insights across the piece.

Interviewer: And what do you personally, based on your experiences, believe constitutes a major draw for collaboration with you for private data donors?

Respondent: Well, crudely, I think that historically the thing that we've relied on most, in relative terms, is our deep pockets. I think there has been a lot of, we have as part of our funding a fairly significant budget available to go and acquire data and to pay money for licensing agreements, that are obviously based on usage not just by ourselves, but by a much larger community of users. Now, that said, as we become more experienced in this, and we are fairly experienced now in these kinds of conversations, I think there are definitely missed opportunities, if the only relationship that an academic centre or service like ours has with the data owners is a vendor customer type relationship. I think there's opportunities for much more than that. The most recent licensing agreement that we've concluded, while there is still money changing hands, and I think that's an important part - I mean, these are businesses, and particularly where they are businesses making data available - we've been able to negotiate considerable savings against certainly commercial market rates, and indeed, even against some of the initial prices quoted for the kinds of academic uses we require. And I think we can do that by providing some pretty tangible benefits for these companies. The thing that I typically will present initially is, certainly when we're using data, a big part of it for academia is to obviously - validate the data to determine its representativeness, to determine its value, to determine its kind of credibility as part of a robust evidence base. And that comes with some weight, I think, from the academic community. So, companies that we've worked with, I think they do see the value in having this stamp of endorsement or a validation from academic users. In terms of their data, we can find an audience for that data, which in turn will enhance the perceived perception of value within the data set as well. And we can drive interest in again, well, whereas we're primarily dealing with academic users, it's not exclusively so and we do have a number of licensing agreements that happen there are examples where it's not limited just to academic use - government use is permitted as well. But again, we don't have the means to have a comprehensive geographical coverage of these data, so, we have just the Glasgow City region for those datasets, which is the city of Glasgow and the

seven neighbouring local authorities, but with a promise of access to other government agencies as well, so departments and local authorities, they can access these data, they can be convinced of the value in the benefits and that might build up enthusiasm for other organizations to seek to license these kinds of data. So, I think, really, enhancing the value of the data is a big one. Meanwhile, but there's also, in those kind of terms, where you're giving it prominence and endorse fairly weighty endorsement, but also, in very practical terms, we can feed back suggestions and methods around how the data can be curated more effectively, how there may be specified more effectively how organizations can respond to questions around privacy or more elegantly and more effectively, kind of balance the tension between, on one hand the data as utility, and on the other hand kind of protecting privacy or other sensitivities within the data. And I think that they see, yeah, they see value in some of those things we do the data as well.

Interviewer: You kind of brought up the aspects that I was going to touch upon the next - partnerships with more government and public sector users. But before that, you said an interesting point that I haven't heard anybody else bring up an interview so far of how private companies could be interested in having public sector organizations exposed to their data, and that they could see that as a value proposition possibly. Could you elaborate slightly on that?

Respondent: Oh, well, just I mean, again, from the point of view of them being businesses, principally just in terms of them having a potential audience that will be willing to pay them for data or for associated services. I think, obviously sometimes there is a misunderstanding about how much money a lot of these government organizations in particular have to invest in data. But in our experience, we're seeing that there is an appetite, and never more so than after the last year, I think, and enthusiasm for the role and the value of data. So, it feels like the mobility data, there's almost a bit of an arms race going on. And there's a lot of companies that will try and tell you about how people move around and how people are behaving within cities, and so on. And there's a lot of interest in there, not just in terms of things like social distancing, but just what is going to happen to our cities, what's going to happen to the way that we live and configure ourselves spatially, and so on, in the coming months and years, as a result of the pandemic and other factors as well. So, I think that's intuitive learning. I think, in the conversations I've had that would be what I would say was one of the benefits I've intuited.

Interviewer: Going back to, as you said, the fact that you're quite heavily academia and research focused. However, as you also mention you do have partnerships with local governments or public sector users. And could you elaborate a bit more and tell me details about how those partnerships come to be? And what are some of their more general characteristics?

Respondent: Well, yeah, we are mainly academic focus. But I think, and indeed, we have to remember that our remit is provide a national service within the UK as a whole, not just limited to Scotland, for example. But inevitably, you kind of strike up partnerships, conversations, relationships, often with more local organizations and so on. So, we have

been working recently, much more extensively with Glasgow City Council, for example. So, that might be a useful one to describe how that collaboration looks and what it entails. And it's been a kind of a multi-pronged thing. We've been obviously, through location, we've been involved with Glasgow for a long time. And there's been a history in the city of Glasgow with things like, the various different investments that the city had - Smart Cities, demonstrator investment, around about the time that our Centre started off. And there's been various different commitments around data that we've been keen to support them with. In the last two or three years in particular, something that's developed has been, some important to have individuals within these organizations that are enthusiastic about the value of data and about the value of engaging with academia. And there's certainly several individuals now in Glasgow city council that do see the benefits with the reduced resources available in house, academia can meet a lot of those demands, or certainly allow them to do things that they would otherwise not be able to do. And we developed a partnership project with Glasgow, sort-of start 2019 we started talking about it, which was very mutually beneficial. We were interested in doing some evaluation of an ongoing capital investment program within the city of Glasgow. And we wanted to evaluate the effectiveness of those investments, which is called the "Avenues Project" - to upgrade a whole range of different parts of the city. Assessed on a range of fronts, but how does it change how people move, does it change the adoption of active travel, how people engage with the local economy, ect. We've been developing methods around computer vision for some time, and we proposed to colleagues at Glasgow the development of a project to use the redundant infrastructure within the CCTV community safety and crime prevention system within the city. To use that as a means to actually do some object detection and count pedestrians and track behaviours. Essentially, as a ready-made sensor network, the CCTV infrastructure would be redeployed, while ensuring no interference or disruption with this main operational use. And that was a really successful piece of work, which led to other conversations in strengthening of individual relationships between people within our Centre and the local authority. More recently, we've developed that - we're doing other works around them, the liveable neighbourhoods' program that's developing there, we're taking a look at the data and indicators there. We've partnered on this mobile data acquisition program whereby drawing on our experience and our expertise in our resources to be able to acquire data, we relicensed these data, explicitly providing the means for local authority and city region colleagues to get their hands on these data and specifying the coverage of the data to reflect some of their needs, as well, as well as obviously been driven very much by an academic research agenda.

Interviewer: In these partnerships, do public sector organisations pay fees as well?

Respondent: No, no, no. Nobody pays for the services at the point of use. However, in cases like this, there was some additional investment that Glasgow provided to us, this is a whole piece that is in addition to just the acquisition of the data, a commitment to do various different analyses on the data as well, which is aligned to some of these programs that were engaged with them on. And indeed, we've sort of more very recently we've

sought to kind of add, or we're dipping our toe rather, a third service agenda, which is much more firmly aimed at local authorities. We had a very modest sum of money invested by ESRC, just towards the tail end of last year, for us to build some demonstrator dashboards and visual interfaces. To make our data under analyses more accessible to local government and other kinds of government audiences. A lot of academics appreciate the value and the potential impact of government adoption of their outputs, but if you just give someone in local government, for example, a copy of research publication that's unlikely to absorb or be particularly accessible, so we're trying to find ways to make these things more accessible. And that means having conversations determining needs and aligning those with our own work.

Interviewer: To reiterate if I understood correctly - you're saying for public sector partnerships, making the data analysis that you offer and the transformation data into more accessible takeaways could be kind of their main draw for them in collaborating with you?

Respondent: I think it's is perhaps not so much the main draw, but it's an essential precursor to working, I think they are interested in the analyses themselves and our capacity to do the, sort of the methodological work. But there's a real, there's a missing piece of the puzzle of anything, which is that communication in order to make those analyses meaningful and aligned with and attuned to their specific needs and the capabilities of people working on that other side of the fence. So, it's not necessarily that, they don't just want accessibility, it isn't the end goal, but it's an essential part for the collaboration to be successful.

Interviewer: And of course, the actual seeking out the useful data sets and negotiating with our private sector partners is another valuable service you give to them, as they myself not even know where to start looking? Or they'll have the capacity, as you said, to kind of build those partnerships themselves.

Respondent: Yeah, absolutely. I mean, but it is generally the case, though, that we will mean, certainly we would be if we can do something for an academic audience. And the data will throw in access for local authorities at no additional cost, then that's a no brainer, of course, we'll do that. If we are having to pay some substantial additional sums to be able to support those kinds of audiences, I think that would be, maybe it would become a harder sell, in terms of our own governance, in terms of our own approval of that with regard to our specific remit. And then, but we would look creatively and think, well, there are economies of scale here, if you have 10 organizations, licensing data from a single provider, then why don't just do through a single collaborative approach or, you know, there will be savings on offer, and also a tremendous amount of value. If you've got more people tapping into the scene, well, then everybody benefits because you get a much better picture of the opportunities within that data.

Interviewer: Yeah, thank you. Well, I had just a few small clarification questions, because I know I'm going over my time here. You kind of alluded to this fact, but do you

believe that personal professional connections, networking, word-of-mouth contributes in any significant way to the expansion of the network and the establishment of new collaborations?

Respondent: I think it's really big, I mean, at every stage whether it's in the point where you're trying to find an interested audience and building up a relationship to make it very clear what the needs are, that you're actually trying to address. The Glasgow example is a really good one of where we've never been working more effectively, with Glasgow than right now, and a lot of that can be attributed to the positivity of the individual relationships. But then also, when you're dealing with the data owners as well, I think it's really big as well. A lot of these companies are often quiet, I mean some of them are the Twitter's and Google's, very inaccessible. And that's been our experiences to date, but I think that if you were going to get access, it would probably be through a well-placed individual within these organizations. But within the smaller companies that we have typically worked with, you often not talking about a large number of people working there and, and you can have a tremendous amount of success by building personal relationships. But conversely, sometimes when individuals change, and individuals in leadership roles change, and organizational philosophies change as a result - it can be quite detrimental to this. I always talk about sustainability of these data sources as being a really integral part of any kind of quality consideration. So, while this data is great, it's got so much value in detail, but you know, the company may turn it around and say, well, we don't want to provide this anywhere we want to change the specification for business reasons, or for another reason. I think trust and individual relationships is really important. And it never works as effectively, as I mentioned before, if you're just a customer, whereas if you have a feeling of a collaboration, and it's more than just the feeling, it can just be as easy to lip service to this idea that we're collaborating, and maybe universities like to be told that we're collaborators and maybe businesses can use that quite cynically I don't know, but we do have a number of relationships, though considered to be much more collaborative, and we're developing business models for the companies themselves. They’re perhaps not well versed in these kinds of deals where they make data available to widespread audiences, that we that we require them to. And so, there's a real feeling of the importance of trust in that.

Interviewer: In that vein of additional value, do you believe, in your experience, partnerships sometimes easier to negotiate if the target issue that you're trying to tackle has strong societal or symbolic importance? Maybe we can see him could potentially seeing the fact of how many data sharing initiatives quite quickly formed around dressing challenges of COVID-19, for example?

Respondent: Yeah, well, COVID was a really good example. And there was a lot of data out there. But I think a lot of the limits of what was available kind of revealed themselves quite quickly as well. I wouldn't want to be too pessimistic about it, but I think that it will very much depend on the individual company, this is the one thing that's really clear, that should come through here really clearly, is that some companies have a real appetite, the

avid sort of sense of corporate social responsibility, where you can leverage that to some extent. But I think ultimately, you have to also, you know, you got to acknowledge that these are companies at the end of the day as well. And, and I would never be dismissive of a company that was unwilling to for fear of cannibalizing a business model were unenthusiastic or unwilling to work with us, we said, well, this is great, we're going to make cities better, and then people are going to have a better life and things. And these are, these are true statements, but they are not necessarily compelling to these companies. But that doesn't make them bad people.

Interviewer: But that's a relevant factor. You bring up within your experience, what has been kind of the most common fears apprehensions expressed by partners, whether public or private, when it comes to forming collaborations, their deterrence?

Respondent: Yeah, well, I mean, that was one I mentioned there. If they're providing data services, in the past, for example, we had a data supplier that provided us with data under a license that allowed us to make it available for non-commercial research use. We relied on that, and we made it available to public sector organization, and then the data provider expressed dismay, because they'd hoped to be able to provide data at cost to that to that same organization. So that was a really good example of a company not knowing exactly what they were signing up to. So, these days, we are trying to manage expectations a lot more and really outline and articulate the implications of these agreements. That's one thing, another one is where companies are less well established to sell data, they may be very cautious about any kind of openness agenda. Maybe their concern is that these are commercially sensitive data, they may fall into the hands of competitors, or rivals, or they may be used as the basis to establish rival companies as well, if people are so minded. So, we need to give these companies tremendously strong reassurances, we have a number of sanctions available that we can sort of, you know, we can we have a stick we can wave at users of data, we ensure that people are appropriately trained, and offered and properly kind of liabilities.

Interviewer: As far as I understand, when it comes to accessing just the data, people have to submit concrete applications with what purposes they'll be using for?

Respondent: That’s a huge, huge, process, the whole governance of actually of licensing itself is massive. And again, we agree that with the data owner in question, and that that can look, it can look like a whole range of different things from something that's fairly open, we actually have a couple of examples with perhaps non-commercial companies, but we've agreed to produce open datasets with. But at the point where we agree licensing, we also agree this sub licensing framework, which will, for example, be very strict limits on the type of work that can be done where data can be accessed, eligibility and potentially in a very granular sense. Again, we would start off with what our goal is to licenses as unencumbered as possible with restrictions. But you can calibrate that and nuance that to come up with something that meets their kind of expectations and offers them the reassurances you need, whilst not limiting the usability of the utility of data for f academic purposes. Well typically, we invite researchers to apply for our data that they'll then have

to summarize, their proposed usage is the very barest minimum. And sometimes it's a much more involved process, but summarizer uses that mainly be an eligibility check. And, then lastly, they'll be made aware of what the limits are, in terms of the access to data. Sometimes we can send the data to them to use within their own environment. Other times, they'll have to access it within a controlled Remote Desktop type environment. We don't have any cases where someone has to physically travel somewhere, we can always offer people remote access, but various different limits associated with that.

Declaration of Authorship

I hereby declare that, to the best of my knowledge and belief, this Master Thesis titled "Matchmaking: Trusted Data Intermediaries and the Connection of Private Data to Public Needs " is my own work. I confirm that each significant contribution to and quotation in this thesis that originates from the work or works of others is indicated by proper use of citation and references.

Brussels, 05 August 2021

Nia Chigogidze

Consent Form

for the use of plagiarism detection software to check my thesis

Name: Chigogidze
Given Name: Nia
Student number: 509672
Course of Study: Public Sector Innovation and eGovernance
Address: Schlossplatz 2, 48149 Münster
Title of the thesis: Matchmaking: Trusted Data Intermediaries and the Connection of Private Data to Public Needs

What is plagiarism? Plagiarism is defined as submitting someone else's work or ideas as your own without a complete indication of the source. It is hereby irrelevant whether the work of others is copied word by word without acknowledgment of the source, text structures (e.g. line of argumentation or outline) are borrowed or texts are translated from a foreign language.

Use of plagiarism detection software. The examination office uses plagiarism software to check each submitted bachelor and master thesis for plagiarism. For that purpose the thesis is electronically forwarded to a software service provider where the software checks for potential matches between the submitted work and work from other sources. For future comparisons with other theses, your thesis will be permanently stored in a database. Only the School of Business and Economics of the University of Münster is allowed to access your stored thesis. The student agrees that his or her thesis may be stored and reproduced only for the purpose of plagiarism assessment. The first examiner of the thesis will be advised on the outcome of the plagiarism assessment.

Sanctions. Each case of plagiarism constitutes an attempt to deceive in terms of the examination regulations and will lead to the thesis being graded as "failed". This will be communicated to the examination office where your case will be documented. In the event of a serious case of deception the examinee can be generally excluded from any further examination. This can lead to the exmatriculation of the student. Even after completion of the examination procedure and graduation from university, plagiarism can result in a withdrawal of the awarded academic degree.

I confirm that I have read and understood the information in this document. I agree to the outlined procedure for plagiarism assessment and potential sanctioning.

Brussels, 04.08. 2021

Nia Chigogidze