UNIVERSITY OF TURKU
Turku School of Economics

Data spaces and their potential for Finnish organizations

Information Systems Science

Master's thesis

Author:
Antti Tiekso
1800640

Supervisor:
Ph.D. Jani Koskinen

3.5.2024

Turku

Master's thesis

Subject: Information systems science
Author: Antti Tiekso
Title: Data spaces and their potential for Finnish organizations
Supervisor: Ph.D. Jani Koskinen
Number of pages: 87 pages + appendices 9 pages
Date: 3.5.2024

The European data strategy published in 2020 introduced a new data management solution and platform for data sharing to reach a set of versatile political, ethical, and economic challenges: common European data spaces. Since then, multiple legal initiatives and significant resources have been allocated for developing data spaces and realizing the data strategy both on EU level and in individual member countries. Data space activities are underway in Finland as well, and individual data space solutions have already been carried out. However, it has still been unclear what concrete possibilities data spaces enable. Moreover, even though some data space solutions have already been implemented, their number in Finland is still limited. To discover possibilities and clarify the current status of data spaces in Finland, this thesis strives to answer the following primary and supporting research questions: What opportunities do data spaces provide for Finnish organizations? What restraints or other implications may affect data space utilization in Finnish organizations?

Prior data space literature has covered for instance data space design, structures, and technologies, along with theoretical opportunities and challenges. The closely related concept of data economy has been studied more widely, simultaneously supporting understanding of data spaces as well. Concrete examples of the opportunities and utilization of data spaces have still been lacking. The theoretical framework of this thesis incorporates the ever-evolving data space definitions and structures along with understanding about the importance of certain standardized structures common to all data spaces. These standards have been and are being developed by Gaia-X in Europe and IDSA internationally.

The study utilizes qualitative research methods to answer the research questions. For data collection, semi-structured interviews were conducted for five experts of data spaces and data economy from both public and private Finnish organizations. The data was analyzed through methods of thematic analysis. The interview findings were collected as codes and categorized into opportunities, restraints, and other findings. They were further divided based on if the code applied to public or private, and larger or smaller organizations.

Four themes were created from the opportunities: improved services, environmental sustainability, direct monetary and business benefits, and indirect business and fiscal opportunities. Improved services included for instance data-based developments in the quality of teaching, health, taxation, travel, traffic, and logistics. Opportunities related to sustainability were limited to data-driven traffic control to save fuel, and to the potential to track carbon emissions across value chains more accurately than before. Direct benefits involved both enhanced cost-efficiency and productivity and are easily measured through business metrics. On the other hand, indirect opportunities reflected the potential brought about by data spaces, for instance co-value-creation with subcontractor network, which may generate benefits indirectly.

Restraints for data space activities also arose. Organizations may have trouble breaking existing mindsets about data sharing, and there is a lot of worry and fear around data space operations. On the other hand, another central discovery was that the field of data spaces is wide open and active players now have the chance to affect the future outlooks of European data spaces. This highlights the relevance of this study.

Key words: data space, dataspace, data economy, data economy ecosystem, data strategy, data sharing

Pro gradu -tutkielma

Oppiaine: Tietojärjestelmätiede
Tekijä: Antti Tiekso
Otsikko: Data-avaruudet ja niiden potentiaali suomalaisille organisaatioille
Ohjaaja: FT Jani Koskinen
Sivumäärä: 87 sivua + liitteet 9 sivua
Päivämäärä: 3.5.2024

Vuonna 2020 julkaistussa EU:n datastrategiassa esiteltiin uusi datanhallintaratkaisu ja datanjakoalusta monialaisten poliittisten, eettisten ja taloudellisten tavoitteiden saavuttamiseksi: yleiseurooppalaiset data-avaruudet. Sittemmin useita lakialoitteita ja mittavia resursseja on kohdennettu data-avaruuksien kehittämiseksi ja datastrategian toteuttamiseksi sekä EU-tasolla että yksittäisissä jäsenvaltioissa. Myös Suomessa data-avaruustoiminta on liikkeellä, ja yksittäisiä data-avaruusratkaisuja on jo toteutettu. On kuitenkin edelleen ollut epäselvää, mitä konkreettisia mahdollisuuksia data-avaruustoimintaan liittyy. Lisäksi, vaikka joitain data-avaruusratkaisuja on jo kehitetty, niiden määrä on Suomessa edelleen rajallinen. Tämä tutkielma pyrkii tunnistamaan mahdollisuuksia ja selittämään data-avaruuksien nykytilaa Suomessa vastaamalla seuraaviin pää- ja apututkimuskysymyksiin: Mitä mahdollisuuksia data-avaruudet luovat suomalaisille organisaatioille? Mitkä rajoitteet tai muut tekijät voivat vaikuttaa data-avaruuksien hyödyntämiseen suomalaisissa organisaatioissa?

Aiemmassa kirjallisuudessa data-avaruuksien saralla on tutkittu muun muassa niiden suunnittelua, rakenteita ja teknologiaa sekä teoreettisia mahdollisuuksia ja haasteita. Vahvasti data-avaruuksiin kytkeytyvää datatalouden käsitettä on tutkittu laajemmin, ja se myös tukee ymmärrystä data-avaruuksista. Konkreettiset esimerkit data-avaruuksien mahdollisuuksista ja käytöstä ovat kuitenkin puuttuneet aiemmasta kirjallisuudesta. Tässä työssä teoreettisena viitekehyksenä toimivat alati kehittyvät data-avaruuksien määritelmät ja rakenteet, sekä ymmärrys tiettyjen kaikille data-avaruuksille yhteisten standardirakenteiden tärkeydestä. Näitä standardeja ovat kehittäneet Euroopan tasolla Gaia-X ja kansainvälisesti IDSA.

Tutkielma vastaa tutkimuskysymyksiin laadullisin tutkimusmetodein. Aineistonkeruussa on hyödynnetty teemahaastattelua viidelle data-avaruuksien ja datatalouden asiantuntijalle suomalaisista organisaatioista sekä julkiselta että yksityiseltä sektorilta. Vastaukset analysoitiin temaattisen analyysin menetelmiä hyödyntäen. Haastatteluista saadut löydökset kerättiin koodeiksi, jotka kategorisoitiin mahdollisuuksiksi, rajoitteiksi ja muiksi löydöksiksi sekä listattiin sen mukaan, koskivatko ne julkisia vai yksityisiä sekä pienempiä vai suurempia organisaatioita.

Mahdollisuuksista kehitettiin neljä teemaa: paremmat palvelut, ympäristön kestävyys, suorat rahalliset ja liiketoiminalliset hyödyt sekä epäsuorat liiketoiminnalliset ja kansantaloudelliset mahdollisuudet. Parempiin palveluihin lukeutui muun muassa dataan pohjautuva parempi opetus, terveys, verotus, matkustaminen, liikenne ja logistiikka. Kestävyyteen liittyvät mahdollisuudet rajautuivat datavetoiseen liikenteenohjaukseen, joka säästää polttoainetta, sekä mahdollisuuteen seurata arvoketjun hiilidioksidipäästöjä entistä tarkemmin. Suorat hyödyt sisälsivät niin parempaa kustannustehokkuutta kuin tuottavuuttakin ja ovat helposti mitattavissa liiketoiminnallisin mittarein. Toisaalta epäsuorat mahdollisuudet kuvastivat data-avaruuksien tuomaa potentiaalia, esimerkiksi yhteisarvonluonti alihankkijaverkoston kanssa, joka voi tuottaa hyötyä välillisesti.

Myös rajoitteita data-avaruustoiminnalle nousi esiin. Organisaatioilla on vaikeuksia rikkoa olemassa olevia ajattelutapojaan datanjakamisen suhteen ja data-avaruustoimintaan liittyen esiintyy paljon huolta ja pelkoa. Toisaalta keskeinen löydös oli myös se, että data-avaruuksien pelikenttä on auki ja aktiivisilla toimijoilla on nyt mahdollisuus vaikuttaa eurooppalaisten data-avaruuksien tulevaisuudennäkymiin. Tämä korostaa tutkielman ajankohtaisuutta.

Avainsanat: data-avaruus, datatalous, datatalousekosysteemi, datastrategia, datanjakaminen

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

As the volume of global data continues to soar with no end in sight (Taylor 2023) a new domain of economic and ethical challenges has arisen. Large corporations in possession of masses of data are feverishly attempting to collect more and dig out all the information on their customers imaginable. Simultaneously smaller businesses struggle to rise into competition. Individual people are starting to resemble livestock who consume services and pump out data for the service providers. At the same time, policymakers attempt to get hold of the threads, especially in the EU where most data is flowing directly to foreign giants causing extensive dependency on a few non-European corporations. This centralization, in turn, causes distrust and a perceived risk of potential data misuse.

What the EU is looking for is data sovereignty: for the control and power over data to remain with the companies or individuals who have created it. To actualize this goal, the European Commission published the European data strategy in 2020 and has subsequently brought out regulations to complement the strategy and each other. In the strategy, the European Commission (2020) introduced the concept of a shared European data space as the basis for realizing data sovereignty.

Data spaces are closely related to the concept of data economy. Data economy is a "digital ecosystem where data is gathered, organized and exchanged to create economic value" (Sestino et al. 2023). Koskinen et al. (2019) further specify that data economy ecosystem participators need to be interconnected somehow. Essentially data spaces can be understood as platforms through which members of a data economy ecosystem are connected to each other and where they can share and acquire data from each other according to jointly agreed rules. The motivation for data sharing may vary from maximizing machinery lifespan by sharing their operating data (see e.g., SIX Sustainable Industry X 2023) to determining the carbon balance of food products (Hyyrönmäki 2022). Tardieu (2022) even compares the upcoming role of data spaces to the role of the web back when Internet was initially taking off.

The European data strategy aims to eventually create one single cross-sectoral European data space as a "single market for data" (European Commission 2020). First, however, the European Commission encourages the development of vertical data spaces covering one sector such as mobility, agriculture, or health. Most likely smaller sectoral, possibly

national, data spaces will first be combined into larger sectoral EU-wide data spaces. After that the goal of a single European data space could become tangible.

As the European data spaces are at an early stage and legislative drivers are new or still emerging, there is a chance for high-technology businesses and public actors to leave their mark in the upcoming data space landscape in the EU. Some Finnish organizations have joined as members to the International Data Spaces Association (IDSA) and the Gaia-X European Association for Data and Cloud. A Finnish company, DataSpace Europe, has taken a leap and developed an agriculture data space to enhance data sharing and utilization within agriculture and food production in Finland. Their platform is already in use and being improved.

Finland is seen as a pathfinder in data (1001Lakes 2024) and therefore as a country of high-technology standards Finland has the potential to be a frontrunner in data space development in Europe. Although there are already some Finnish data space initiatives in motion, it is still unclear what concrete opportunities do data spaces pose for Finnish organizations. In order to fill this knowledge gap, the following research question is presented:

- **What opportunities do data spaces provide for Finnish organizations?**

To provide a more comprehensive and reliable image of the current status and future of data spaces operations, a supporting question is also presented:

- **What restraints or other implications may affect data space utilization in Finnish organizations?**

The opportunities are categorized based on the organization sector and size: public vs. private organizations, large (over 250 employees) vs. small or medium-sized (under 250 employees) companies. Additionally, the specific European data space or spaces applicable to each opportunity is presented. Lastly, the opportunities are grouped under connecting themes.

This thesis has the following structure. Chapter 2 introduces data spaces on a general level including their background, relevance in data economy ecosystems, structure, and theoretical opportunities and challenges according to literature. Chapter 3 fixes the aim on European level as it introduces the EU data strategy and its key legislative initiatives:

the Big Five. Additionally, the current status of nine common European data spaces is inspected. Chapter 4 tightens the scope even further and examines Finnish data space activities. Chapter 5 presents the thesis methodology used to provide answers for the research questions: a qualitative semi-structured interview study with five experts specialized in data spaces and data economy from both public and private Finnish organizations. Chapter 6 brings forward the results of the empirical study. Chapter 7 discusses the results in light of prior literature and chapter 8 presents contributions and concludes the thesis.

2 Data spaces

The business needs of today are posing new data use requirements for organizations – in the form of data sharing. Sharing business-related data with other organizations within one's industry or value network is even seen as a requirement for new innovations, business models, services, products, and other activities that enhance organizations' operations. (Jarke et al. 2019; Heikkilä et al. 2023.) Examples include, for instance, "collaborative predictive maintenance" where industries utilizing the same equipment share their machine operation data to gain collectively enhanced maintenance (Jarke et al. 2019) or generally remaining updated on the latest new trends within the business domain (Oliveira et al. 2019).

However, this data sharing is not a given for companies striving to gain a competitive edge over others. Issues include for instance data privacy, ownership, access rights, and data quality (Rantanen et al. 2019; Heikkilä et al. 2023). To provide a solution for a variety of issues, the concept of data spaces has been presented.

The term data space has a plethora of varying definitions (see e.g., Montero-Pascual et al. 2023; Scerri et al. 2022; Green Deal Dataspace 2024b). Perhaps the most comprehensive definition has been provided by the European Commission-funded Data Spaces Support Centre (DSSC). According to their glossary, a data space is

> *"a distributed system defined by a governance framework that enables secure and trustworthy transactions between participants while supporting trust and data sovereignty"* (DSSC 2023b).

The following subchapters dig deeper into data spaces to provide an overview. Chapter 2.1 inspects the background of data spaces further, chapter 2.2 inspects the role of data spaces in data economy ecosystems along with data space components, and chapter 2.3 discusses opportunities and challenges that have been recognized in literature to date.

2.1 Background on data spaces

The data space concept first arose when Franklin, Halevy, and Meier (2005; 2006; 2008) inspected issues in "conventional relational" database management systems (DBMS) and proposed a new concept for managing data: data spaces. The purpose of data spaces was to combine similar functionalities between multiple data sources and data owners,

whereas in DBMSs all data were controlled by just one party (Franklin et al. 2005). The relevance of data spaces has thus increased as the need and opportunity to create business value from data has simultaneously grown.

The issue Franklin et al. (2005) discovered was related to data integration in scenarios where data from multiple sources in many different forms and file formats needs to be available through a query. The authors differentiated data spaces from data integration systems by the fact that data would not need semantic integration beforehand to make it available in data spaces (Franklin et al. 2005; Halevy et al. 2006; Franklin et al. 2008). Semantic integration means that all data points or columns need the same names or formats for them to be discoverable through queries.

The problem with semantic integration is the effort it takes to format the data before being able to share it in a usable form. Solutions for this issue have been sought after in literature broadly (see for example Sarma et al. 2008; Dong et al. 2009; Sarma et al. 2011) for instance by proposing a keyword type search where query results are not exact but approximate (Franklin et al. 2005; Dong et al. 2009) and "probabilistic mediated schemas" (Sarma et al. 2008; Dong et al. 2009) that deal in probabilities. However, it seems that some of the same issues are still present in today's efforts to establish EU-wide data spaces. The European Parliament & Council of the European Union (2023) state that "standardisation and semantic interoperability should play a key role - - to ensure interoperability within and among common European data spaces".

Franklin et al. (2005) proposed Data Space Support Platforms (DSSPs) to aid in issues related to data integration. They describe that "a DSSP must deal with data and applications in a wide variety of formats accessible through many systems with different interfaces." A resembling structure has later been established by the Gaia-X association called Data Spaces Support Centre (DSSC) to support the establishment of European data spaces. The Gaia-X DSSC includes assets to guide organizations in getting into speed with data space operations (DSSC 2023a). The DSSPs introduced by Franklin et al. (2005), however, would have been slightly more technical support tools. There are still similarities in DSSC and DSSPs not only by name but also by function.

In addition to semantic integration, Franklin et al. (2005) stated that data quality cannot be guaranteed in data spaces without agreements between data source owners. Since then, such agreements have been governed by international and European non-profit data space

entities: International Data Spaces Association (IDSA) that was established in 2016 (Pretzsch et al. 2022) and the aforementioned Gaia-X European Association for Data and Cloud established in 2021 (Gaia-X n.d.b). Regarding the agreements, IDSA has introduced the IDSA standard to govern the parties involved in data exchange across international data spaces (IDSA n.d.a). A similar governance system is used in the Gaia-X association in the form of a clearing house that verifies organizations before their entry to the Gaia-X ecosystem (Gaia-X n.d.a).

IDSA was originally established after the German institute Fraunhofer-Gesellschaft introduced the creation of a data space for safely managing data from organizations of different sectors in industry context (Pretzsch et al. 2022). According to Pretzsch et al. (2022) there was so much demand for such structures that the IDSA was grounded a year later. The Gaia-X initiative has similar aspirations, but it operates in the European context.

Hyyrönmäki (2022) presents an interesting statement about data spaces: that such a market transformation could occur where the economy is no longer based on competition, but cooperation. Data spaces and data economy could change the nature of pursuing competitive advantage, as they are strongly based on reciprocity: in order to gain benefits from other organizations' data, one must be ready to distribute their own data as well. In an ideal scenario, these actions could increase innovation rapidly. On the other hand, businesses can still compete in how to make the most out of the acquired data, and who has the best data-analytical tools.

In literature there is no commonly used theoretical framework related to data spaces. In Figure 1 an illustrative depiction is presented to show the connections and participators in data sharing across a data space. The discussed agreements can also be seen by the name of Usage policies.

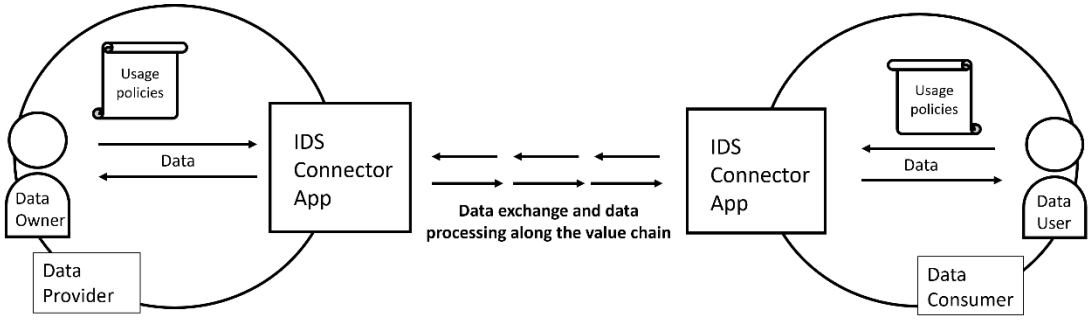


Figure 1. Data sharing in a data space (adapted from IDSA n.d.c)

To conclude on the background, two more concepts are presented: data pools and data lakes. This is done to ensure resembling terminology is understood, and different concepts not confused with each other. According to IBM (2023) “data pool is a centralized repository of data where trading partners - - can obtain, maintain, and exchange information about products in standard format”. On the other hand, data lakes are used to store larger quantities of data but instead of standard format, they are in raw format (IBM n.d.b).

In data pools, the issue of semantical integration emerges again since the data needs to be in standard format in order to use it. In data lakes this is not required. However, this also poses an issue for data lakes: there is no guarantee of data quality as the data could be in any format, which could restrict data usability. Therefore, data pools are more suitable when users search for specific data, whereas data lakes enable broader inspection and machine learning projects, for instance (IBM n.d.b).

Data spaces differ from these concepts in the sense that a data space is only meant to serve as intermediator between organizations, but it does not store data in itself like data pools and lakes. Even if data spaces become a more widely utilized data management structure does not necessarily mean they would replace other solutions, but these different platforms could support one another. Instead, the role of decentralized structures like data spaces, over data pools for instance, is underlined when self-determination over one’s own data is discussed (European commission 2020).

2.2 Data spaces and data economy ecosystems

To better understand the purpose of data spaces, it is worthwhile to lay groundwork by examining them in the context of data economy ecosystems. According to Koskinen et al. (2019) “a data economy ecosystem is a network that is formed by different actors in the ecosystem that are using data as a main source or instance for business.” Additionally, the actors need to be interconnected on some level and follow an agreed set of rules determining conduct within the ecosystem (Koskinen et al. 2019).

The term data ecosystem has been defined for instance by Curry (2016) who elaborates it as a “data environment supported by a community of interacting organisations and individuals.” Moreover, data ecosystems can cover smaller or larger entities, for instance one or multiple business sectors. Various data infrastructures can also be used to reinforce

data ecosystems. (Curry 2016.) To conclude, data economies and data ecosystems share similar descriptions, and they are likely used interchangeably in some contexts.

After discussing the concepts of data economy and data ecosystem it is easier to understand data spaces as well. In the European data strategy, it is stated that "data spaces should foster an ecosystem - - creating new products and services based on more accessible data" (European Commission 2020). Hence, data spaces could also be understood as platforms or supporting infrastructures such as those mentioned by Curry (2016), that are essential for the operation of data economy ecosystems. Scerri et al. (2022) even state that data spaces could broadly stand for ecosystems themselves, if their participants shared data, followed legal contracts, and employed new innovatory modes of business.

Like data ecosystems, data spaces can also cover one or more companies, countries, or business sectors. In the European data strategy, the aim is to eventually create a single European data space, but at first the development of different sectoral data spaces is required. (European Commission 2020.) For instance, the Finnish company DataSpace Europe has established a data space platform for the agriculture and food production sector in Finland with an aim to expand to different sectors in the future (Sitra 2023).

Even though data spaces established among different value networks have their own contextual characteristics, certain standardized structures are also pivotal in all data spaces. IDSA has established a set of these structures as data space requirements. They include a (IDS) connector "as a gateway for data and services", a data owner and data user which are supported by data provider and data consumer devices (see Figure 1), an identity provider to confirm identity information, app stores and apps to perform various functions, a broker to provide data source-related information, a clearing house to process data-related transactions, and vocabularies to "provide standardized descriptors for data based on accepted best practices" (Hillermeier et al. 2021).

The European-based DSSC initiative has recently published their own "data space blueprint" including their central data space requirement concepts, or "building blocks". These can be seen in Figure 2. The DSSC building blocks include two categories: business and organizational building blocks along with technical building blocks. (DSSC 2024.) These are slightly more theoretical requirements that consider commercial and

operational motives as well as technological and legal requirements and they are aimed as a support tool for businesses who are in planning or early stages of data space adoption.

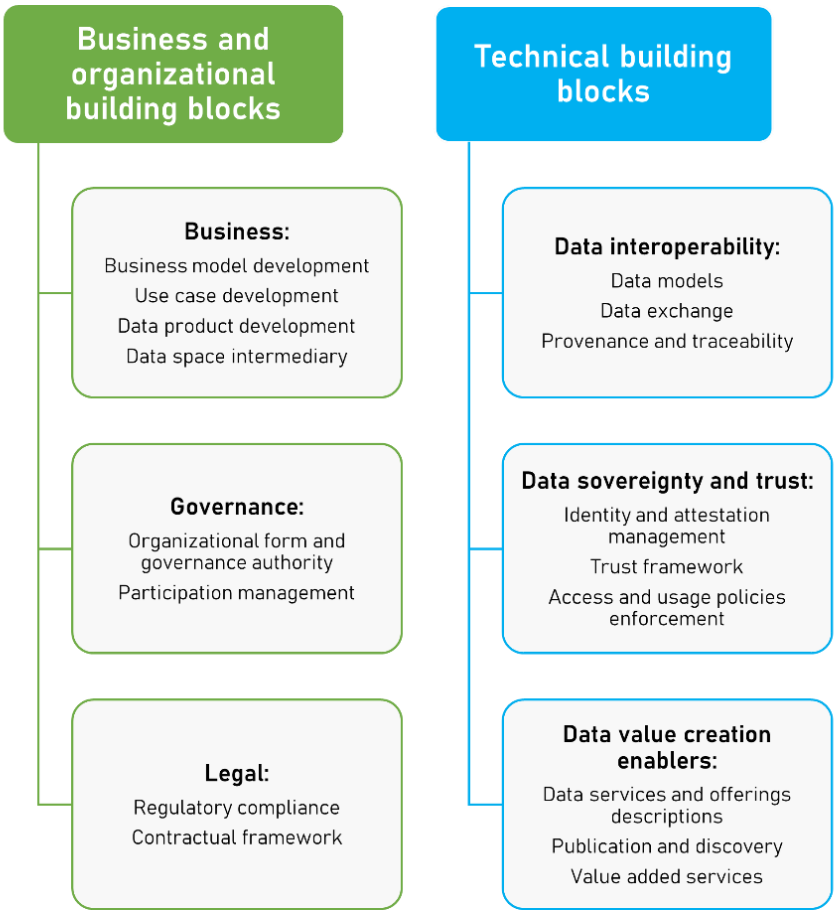


Figure 2. Data spaces building blocks (adapted from DSSC 2024)

2.3 Theoretical opportunities and challenges of data spaces

To make potential effects of data spaces more concrete, Scerri et al. (2022) introduced two frameworks regarding the opportunities and challenges in common European data spaces. Many of the issues are mutual for other international data spaces as well, in addition to European data spaces. In the following subchapters these opportunities and challenges are briefly discussed to give an overview of how literature views future implications.

2.3.1 Opportunities

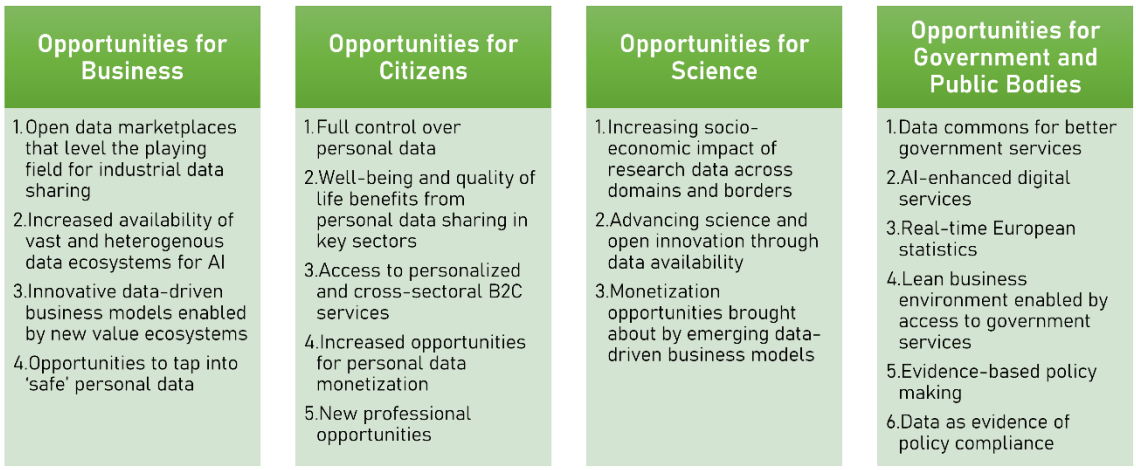


Figure 3. Opportunities for common (European) data spaces (adapted from Scerri et al. 2022)

Scerri et al. (2022) have recognized significant interest in a few areas regarding data sharing and management in the future. These include harmonizing data sharing solutions with each other, enabling the use of data silos in a fair, secure, and decentralized manner, making way for new data-based business models, enhancing the adoption of data technologies, and boosting Europe-wide data analytics in a common ecosystem. To introduce opportunities related to these interest areas, the authors have established four groups of opportunity beneficiaries: business, citizens, science, and government and public bodies (Scerri et al. 2022). These categories and their opportunities are also shown in Figure 3.

First, the opportunities recognized for businesses are discussed. Scerri et al. (2022) present four groups of business opportunities: open marketplaces for data, increased data sources for AI solutions, new "data-driven business models", and new ways to safely use personal data for business. In the open marketplaces, not only the large but smaller businesses are also able to share their data for a "fair compensation" and maintain sovereignty over that data. They can also acquire data from other players for their own business needs, according to an agreed set of rules. (Scerri et al. 2022.) This opportunity largely reflects the key goals of the Data Act initiative (European Parliament & Council of the European Union 2023).

Regarding AI, new accessible data sources could provide high-quality learning material for different AI algorithms, which in turn will support many essential business processes.

However, the Artificial Intelligence Act (European Commission 2021a) is simultaneously governing the safe use of AI-based systems in the EU, which might restrict the use of some datasets. The new "data-driven business models" will move the current competitive markets towards more a collaborative scenario (Scerri et al. 2022; see also Hyyrönmäki 2022), where rapid development of entire business sectors could thrive through data sharing efforts. Using people's personal data for commercial purposes is a complex matter, and Scerri et al. (2022) refer to "explicit consent and true anonymization" as a requirement for enabling better use of personal data for cross-sectoral services between businesses from different fields.

Second, opportunities for individual citizens are presented. The five areas Scerri et al. (2022) present are data sovereignty, increased well-being, personally tailored services, a chance for data monetization, and new job opportunities. Data sovereignty is one of the key drivers in the European data strategy (European Commission 2020) and its importance cannot be overlooked. IDSA determines data sovereignty as "how, when and at what price other may use it [data] across the value chain" (IDSA n.d.b).

Opportunities for increased well-being go hand in hand with personally tailored services, albeit from a slightly different perspective. Using citizens' personal health data for discovering diseases or potentially emerging health issues could increase their well-being and general healthcare (European Commission 2020; Scerri et al. 2022). In Finland, for instance, the "Traffic Data Ecosystem" provides users with "real-time traffic information" (Lautanala 2024). Such solutions could have potential to enhance mobility solutions across borders if implemented as a European-wide dataspace. Simultaneously, using other personal data could present individuals with better personally tailored services, which on the other hand might lead to decreased costs (Scerri et al. 2022) along with increased satisfaction or higher perceived quality of life. The European Commission (2022c), however, recognizes that there might be negative outcomes in personally tailored experiences as well.

Common European or international data spaces could also present direct monetary opportunities in the form of data monetization where users can trade their data to businesses for compensation: this is another focal point in data sovereignty (Scerri et al. 2022; IDSA n.d.b). New job opportunities are also likely to arise regarding these new data space enabled innovations, services, and products (Scerri et al. 2022).

Third, we look at opportunities presented for science and academic world. Scerri et al. (2022) introduce increased cooperation possibilities, data availability, and monetary prospects. Joint standards and harmonization of data sources are sure to facilitate sharing and analyzing datasets for scientific purposes and promote cooperation between researchers since less upfront effort is required for preparing data for analysis. In Europe, the "European Open Science Cloud" (EOSC) is already driven by these motives, enabling researchers with better access to existing datasets (European Commission 2020). The European Commission (2020) is also aiming to widen the EOSC scope beyond research communities all the way to private organizations, potentially enabling aforementioned monetary prospects for research institutions as well. In conclusion, common data spaces could ideally increase cooperation and sources of income in science.

Fourth, and last, we come to opportunities presented for government and public bodies on national and European level. The opportunities are associated with improved public services, security, financial efficacy and predictability, informed decision making, and compliance monitoring. Data is an essentiality for improved public services (European Commission 2020). According to Scerri et al. (2022) free use of public data could aid both public and private organizations, for instance by accelerating the development of new and improved systems and services. AI applications are moving towards more advanced predictive techniques (European Commission 2020) and available public data could facilitate AI-driven service improvements as well (Scerri et al. 2022).

Data spaces could also harness "real-time statistics" to aid in national and international security. For instance, real-time information could be distributed across a data space on spreading diseases, such as COVID-19, or other national security and border control matters (Scerri et al. 2022). The Data Act also recognizes circumstances where such data would need to be transferred between countries to ensure an individual's "right to security and the right to an effective remedy" (European Parliament & Council of the European Union 2023). Scerri et al. (2022) also suggest that real-time data transactions could prove useful for financial efficacy and predictability in the form of potential cost cuts and more accurate budgeting.

The last two opportunities relate to decision making and monitoring compliance. In informed decision making, decision makers could utilize accessible government data to revise policies more flexibly and seamlessly (European Commission 2020; Scerri et al.

2022). Regarding compliance monitoring, some large companies already provide data-driven, automated compliance tools (Deloitte 2020; IBM n.d.a), and similar solutions could potentially be applied to national and international policy compliance monitoring as well. However, this raises the question of why and when policymakers should gain access to individuals and organizations' data, and what are the specific criteria for the statement "where there is an exceptional need", as stated in the European Data Act (European Parliament & Council of the European Union 2023). In conclusion, challenges are sure to arise alongside opportunities in common European and international data spaces. These challenges will be discussed in the following subchapter.

2.3.2 Challenges

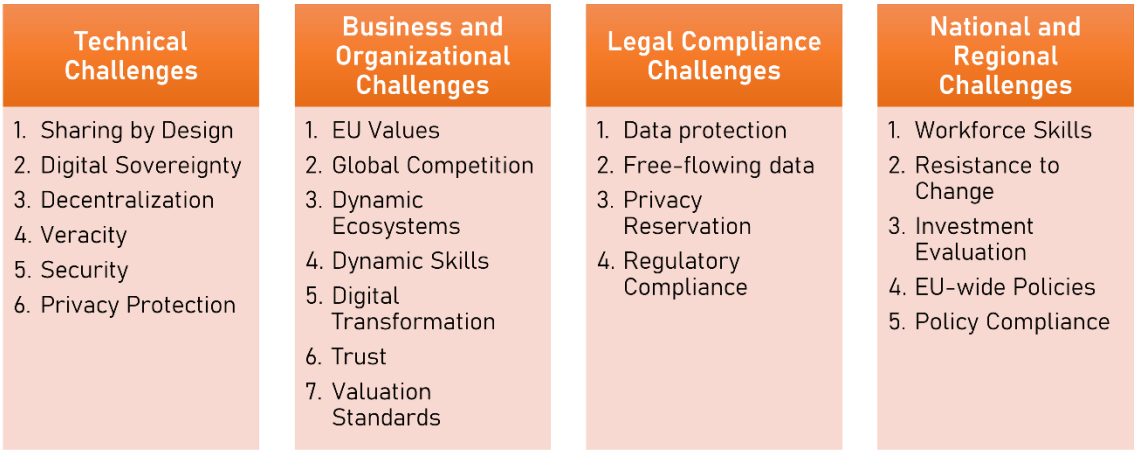


Figure 4. Challenges for common (European) data spaces (adapted from Scerri et al. 2022)

Scerri et al. (2022) also bring forward a variety of recognized challenges related to common European or international data spaces. They first introduce two high level categories: inter- and intra-organizational challenges. A prime example of an inter-organizational challenge is the lack of consensus both in legislation and in trusted data space intermediators. In the intra-organizational sphere, the challenges largely relate to data itself: how should it be shared safely and how can it be used for creating value for business.

Later, Scerri et al. (2022) categorize the challenges into four categories which are also visible in Figure 4. These are technical, business and organizational, legal compliance, and national and regional challenges. Although some challenges are applicable to multiple categories, here they are discussed under their original headings.

First, the technical challenges include topics related to “sharing by design”, data sovereignty, decentralization, veracity, security, and privacy (Scerri et al. 2022). The challenge of “sharing by design” can be seen not only as a technical issue, but an organizational one as well. This means that currently, data management processes in organizations do not consider the prolonged nature of the data lifecycle, as it could be shared for further benefits after its initial utilization (Scerri et al. 2022).

As mentioned, data sovereignty is one of the chief motives behind data spaces and data economy. The challenges Scerri et al. (2022) raise are the technical solutions that ensure the data control remains with its creator. Otherwise, the entire concept would be flawed. For tackling these issues, the “Eclipse Dataspace Connector” (EDC) has been introduced in 2021 as a comprehensive solution for multiple data sovereignty problems (Pampus et al. 2022; Eclipse Foundation n.d.). Related to data sovereignty arises the matter of decentralization: so that no single entity would hold control over all the data. Regarding decentralization, one question is the ultimate nature of a single European data market introduced in the European data strategy (European Commission 2020). It is somewhat unclear if the final common European data space would be controlled by a single governmental body – a centralized approach – or by a cooperative network of private and public data space intermediators – a decentralized approach. One suggestion presented in the data strategy is blockchain technology (European Commission 2020) that relies on decentralization and peer-to-peer surveillance.

The challenges labeled veracity, security, and privacy are closely linked. To ensure veracity, or trustworthiness of the data, it must contain some indication, or metadata, about where it has been created and used. Simultaneously, the privacy of the data creator and other related stakeholders must be ensured. These issues tie together in ensuring the security in a data sharing network so that only authorized people gain access to the data and its metadata. (Scerri et al. 2022; European Parliament & Council of the European Union 2023.)

Second, we will cover business and organizational challenges. Here, the issues involve EU values, global competition, prospective transformations and dynamic nature of ecosystems and organizations, along with trust, and standards for evaluating data quality (Scerri et al. 2022). The challenge of European values is that a European data platform must conform with a set of values many foreign companies do not. This issue has been

recognized in the Digital Markets Act as it imposes obligations for foreign companies as well who wish to operate within the EU (European Commission 2022b). Global competition challenges are closely linked to the prior, and Scerri et al. (2022) emphasize the need to identify globally competitive data service products that originate in the EU.

Transformations and the general dynamic nature of ecosystems, skills, and organizations might also pose challenges among common data spaces. In ecosystems context, defining ownership of data is not necessarily a straightforward process (Scerri et al. 2022). This issue was also brought up after the initial proposal for the EU Data Act since clarification for its data ownership definition was requested (Bräutigam et al. 2022). Simultaneously, organizations and their skill requirements are undergoing digital transformation, and foreseeing all future requirements is difficult (Scerri et al. 2022). What can be deduced is that skillsets and organizational needs are changing and there will be a need to adapt to new requirements. The beneficiaries of these disruptions are likely large high-tech companies who have the necessary resources and know-how, and on the other hand agile and small technology-oriented businesses who might be able to adapt quickly and prosper.

The challenges of trust and standards for evaluating data quality are once again related to one another. Trust can suffer if there are no commonly set criteria for data quality shared across a data space (Scerri et al. 2022), however this issue has already been broadly considered in data space initiatives, (see e.g., European Commission 2020; Eclipse Foundation n.d.). Trust issues are also related to the upcoming legal challenges since legal frameworks are a requirement for inspiring trust within data spaces (Poikola et al. 2023). Non-compliance with agreed legal frameworks could have a trust-extinguishing effect.

Third, we inspect legal compliance challenges closer, especially in European data spaces. The ones Scerri et al. (2022) point out are protection, free flow, and privacy of data, as well as regulatory compliance. The legal challenges are vital in the data space process, and they overlap with other interest areas as well. If data moving across data spaces is not protected and private, not only will there be legal sanctions, but trust will crumble simultaneously. Free flow of data could face both legal and technical constraints. Additionally, without regulatory compliance along with trust and understanding of EU data policies, a common European data space might remain an abstraction.

Fourth, and last, we come to national and regional challenges. Scerri et al. (2022) bring forward challenges related to public employees' digital skills, resistance to change,

investment evaluation, EU-wide policies, and related policy compliance. According to Scerri et al. (2022) the transforming requirements in digital professional skills are most difficult to realize for public organizations and their employees since they lag in progress and are potentially resistant to support digital transformation processes in business.

Scerri et al. (2022) also point out potential challenges in public investments into new data-driven areas. Combined with the former two challenges, slow and reluctant public organizations could in the worst case hinder the development of data-driven societies notably. Additionally, Bräutigam et al. (2022) bring up a concern that new opportunities presented by new European data regulations might not pique the interest of citizens or companies enough, resulting in passive idleness.

Regarding EU-wide policies, Scerri et al. (2022) express their concerns about moving from regional policies to the EU level due to differing requirements between regions. This issue could be intensified by the fact that some EU-wide data regulations are monitored by national authorities and some by the EU. These regulations will be discussed in subchapter 3.1 in more detail. Similar challenges are likely even more significant in the data spaces that comprise not only EU member countries, but individuals, organizations, and public bodies cross-continentally as well.

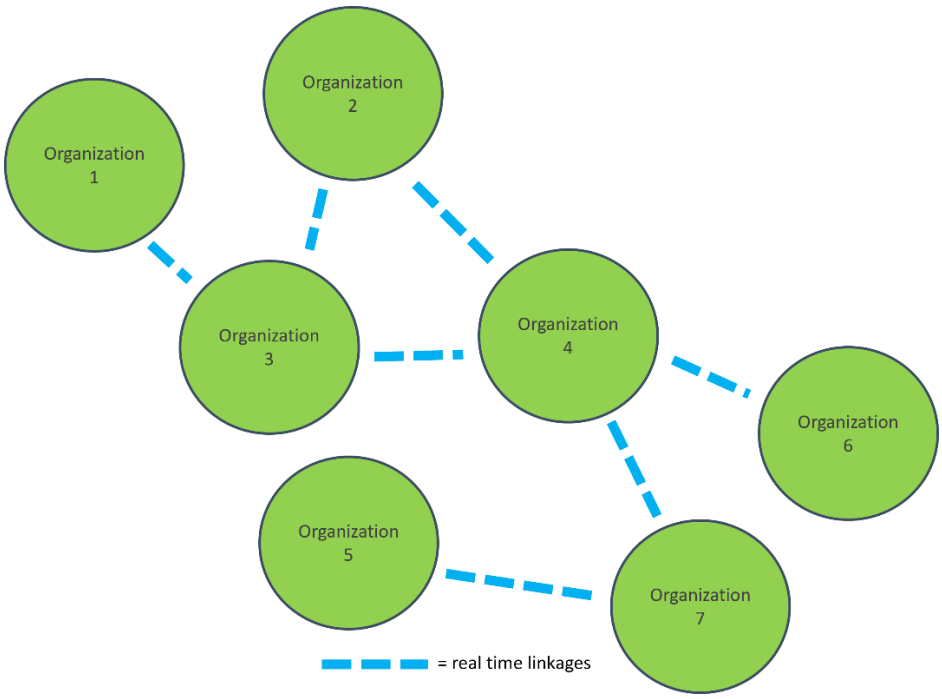


Figure 5. Data space network (adapted from IDSA n.d.b)

In this chapter, we have discussed what data spaces are, why they have emerged, and what theoretical opportunities and challenges they may pose. In the context of this thesis, data spaces are understood as a combination of trust mechanisms, such as legal contracts, frameworks, and technologies to provide a new secure method for sharing data between organizations. They present a wide range of business opportunities and challenges, albeit on a theoretical level for now. Instead of multiple point-to-point integrations between individual companies, data spaces form a network of organizations concurrently providing a lighter and potentially more cost-efficient technological solution suited especially for data transactions between large organization networks. An illustrative example is provided in Figure 5.

3 European data spaces

Over the recent years, the European Union and Commission have published a series of ambitious, data-driven objectives including the European data strategy and a set of legislative instruments to support data economy and data-driven business within the European Union. The data strategy also proposes a "single European data space", defined as "a genuine single market for data, open to data from across the world" (European Commission 2020). Before the establishment of such a single space is possible, nine common European data spaces are to be formed with different focus areas. They have been later defined as "purpose or sector specific or cross-sectoral interoperable frameworks for common standards and practices to share or jointly process data" (European Parliament & Council of the European Union 2023).

This chapter will discuss the European data strategy and its legislative areas and introduce the nine common European data spaces. Chapter 3.1 discusses the strategy and the five related legislative acts, also covering what European data spaces are and why they were originally introduced. Chapter 3.2 inspects the current state of the nine common European sectoral data spaces. Beforehand, it is worthwhile to point out a European Open Science Cloud (EOSC), which can be seen as a tenth European data space, focusing on the field of research. The Open Science Cloud has been introduced separately from the other nine but will be connected to the other data spaces later on. (European Commission 2020.)

3.1 Regulatory initiatives & drivers

3.1.1 European data strategy

In February 2020 the European Commission published "A European strategy for data" (later "European data strategy") introducing the European Union's (EU) goal to become a global frontrunner as a data-reinforced society. The strategy aims to increase data-driven business and innovation in the EU along with creating a more efficient digital public sector. In its core are European values including "open, fair, diverse, democratic, and confident" conduct. (European Commission 2020.) The strategy also proposed the creation of European data spaces for the first time.

In a working document published by Sitra, Bräutigam et al. (2022) discuss key areas of the European data strategy. One of the main issues relates to data economy. The

increasing importance of data economy has been recognized, but related legislation has been too shallow and fragmented between different EU member countries to create clear guidelines. (Bräutigam et al. 2022.) The data strategy thus aims to create a clear, uniform legislative framework to tackle these issues and enable the EU to have a leading position in the world of data (European Commission 2020).

So far, the world's leading technology companies, who also possess the most data, are mainly from outside Europe and the EU (see European Commission 2020; Tardieu 2022). One reason for this is that the large foreign technology companies have fewer legislative restrictions than their European competitors. The European data strategy and its amendments strive to level the playing field by specifically targeting those leading companies' – gatekeepers – operations inside the EU. This is done to return the control of data to its original creators within the EU, but to also permit European technology companies such as cloud service providers to compete with the gatekeeper companies. (Bräutigam et al. 2022.)

The data strategy also introduced the idea of a common European data space (European Commission 2020), a platform supporting a data economy ecosystem where private and public operators from the EU could safely share and trade with data, enabling new business opportunities, innovations, and data sovereignty. For now, businesses lack trust to share their data with others (Bräutigam et al. 2022) which hinders the data economy from taking off. The European Commission (2020) tries to build trust through a ”governance framework” that includes new data-related regulations.

Bräutigam et al. (2022) also discuss these regulations: the Big Five. The Big Five include five sets of EU regulations regarding corporate, private, and public data policies: data governance act, digital markets act, digital services act, artificial intelligence act, and data act. The focal points of each act are discussed in the following subchapters.

3.1.2 Data governance act

The data governance act (DGA) was the first step the European Commission introduced towards actualizing the European data strategy (Bräutigam et al. 2022). The DGA concerns public sector, data intermediation services and individual EU citizens (see European Commission 2022a). The DGA's goals include enabling the reuse of public

sector-held data and providing guidelines about transferring non-personal data within the EU and to third countries (Bräutigam et al. 2022; Kaivola 2023).

This facilitated data sharing provided by the DGA could make it easier for businesses of all sizes to gain access to public data sources. These data include for instance medical records and locational data (Bräutigam et al. 2022). In turn, new data might increase innovation, actualized in "improving personalised medicine, offering new mobility solutions a[nd] promoting the European Green Deal" (Ministry of Finance n.d.). However, some concerns have been raised on whether the DGA motivates individuals and businesses enough for them to actually take advantage of the available data (Bräutigam et al. 2022).

The DGA also promotes the idea of data altruism where people or businesses can offer their data available "to support objectives of general interest" (European Commission 2022a). This is in line with the European data strategy and the idea of data economy, which is largely based on reciprocity: you give some, you get some – albeit indirectly in this case. It will be possible for an organization to register itself as an altruistic organization (Bräutigam et al. 2022; European Commission 2022a) meaning the data they share would be comparable to public data provided by the public sector.

In addition, according to the DGA, a new "Data Innovation Board" will be established to oversee the implementation of a proposed data governance framework. However, the surveillance of the DGA compliance will be given to national authorities in each EU member country. Those authorities are also at liberty to sanction non-compliant data intermediation services accordingly. (European Commission 2022a.)

3.1.3 Digital markets act

The digital markets act (DMA) aims to clear the way for competition inside the EU between large gatekeeper companies and other businesses (see Bräutigam et al. 2022; European Commission 2022b). The gatekeepers are companies who meet the following three criteria: they have a strong influence on the EU market, they provide "a core platform service" that is "an important gateway for business users to reach end users", and they have, or are likely to have, a very stable position in the market (European Commission 2022b; Wiedijk & Roman 2023).

The term “core platform service” includes for instance large search engines, web browsers, social media platforms, operating systems, virtual assistants, and other similar services (European Commission 2022b). Although not specifically named in the regulation, Bräutigam et al. (2022) have recognized at least Amazon, Apple, Google, Meta, and Microsoft as example gatekeeper companies based on the criteria.

The issue of gatekeeper-concentrated power is that because of “network effects” such central entities can end up in “monopolistic or at least oligopolistic” roles (Jarke et al. 2019). This can cause a risk of diminishing competition as other businesses struggle to enter the market. The DMA obliges the gatekeepers regarding areas such as data accessibility, transparency, fair competition, and advertising. The financial sanctions for violating agreed obligations are extremely high to ensure compliance, up to 20 % of the company’s global revenue. These violations are monitored the European Commission. (Bräutigam et al. 2022; European Commission 2022b.)

3.1.4 Digital services act

The digital services act (DSA) aims to clarify what responsibilities online platforms have in regard to what content is visible and what products are sold on said platform. The DSA also provides guidelines on how these contents should be moderated. (See Bräutigam et al. 2022; European Commission 2022c.) The primary tool for reaching these goals is “to impose take-down and transparency requirements” (Lindroos-Hovinheimo 2023) meaning instructions about what content needs to be removed from the platform and why. This can concern for instance “illegal content, online disinformation or other societal risks” (European Commission 2022c).

Unlike the DMA, the DSA can bind smaller online service providers as well as large ones but through different requirements. The DSA is directed to companies providing “intermediary services” in the EU (see European Commission 2022c; Lindroos-Hovinheimo 2023). To summarize the requirements, the larger the platform is, the tighter are the obligations. Instead of gatekeepers, the DSA refers to very large online platforms – VLOPs. The risk for illicit content is perceived largest on VLOPs (Bräutigam et al. 2022).

The DSA is important because there has been no prior legislation about what online platforms can show their users (Lindroos-Hovinheimo 2023). The online platforms in the

DSA include for example social media, e-commerce, app store, traveling, and accommodation platforms (Bräutigam et al. 2022). Like the gatekeepers in the DMA, the VLOPs are not named specifically either, but Bräutigam et al. (2022) mention Alibaba and TikTok from China along with Meta, Apple, Google, and Amazon from the United States. In the light of these estimates, the VLOPs and gatekeepers are likely to include the same businesses to some extent.

The sanctions for violating the DSA are also high, up to 6 % of the company's global turnover. Monitoring DSA compliance is given to national authorities in the EU member countries. (European Commission 2022c.) According to Bräutigam et al. (2022), however, there are inconsistencies in the DSA that could pose issues for the governing authorities. For instance, the definitions separating different company categories from each other can be unclear. Additionally, there are differences between the DMA and the DSA concerning terminology and obligations although the same large companies could be subject to both, either as gatekeepers or as VLOPs.

3.1.5 Artificial intelligence act

The Artificial Intelligence Act (AIA) was published by the European Commission in 2021, making it the first published act that is presented in this thesis. However, according to interviewees of Bräutigam et al. (2022) the AIA has so little in common with the rest that the DGA can be perceived as the first act directly related to the European data strategy. It is worthwhile, however, to introduce the AIA since it is a topical act and shares some commonalities with the DSA regarding user profiling by use of AI tools.

The primary aim of the AIA is to ensure the safety and transparency of artificial intelligence (AI) systems used and brought to market in the EU. It obligates AI service providers operating within the EU and AI service users from EU member countries. The key takeaway in the AIA is a risk-based approach that ranks the AI systems based on the risk they pose to EU values and citizens. (Bräutigam et al. 2022.) AI services within the highest risk classes may be prohibited entirely, whereas services with a lower risk rating will face other regulatory actions (European Commission 2021a).

According to Bräutigam et al. (2022) there have been disputes about whether to allow some AI technologies to be used in the public sector, for instance in law enforcement. According to the AIA, no restriction is set upon such systems when they are used to

“detect, prevent and investigate criminal offences” (European Commission 2022). After recent discussions about the rise of generative AI and its status within the EU, the AIA should be inspected through a particularly critical lens since it does not yet treat the issue of generative AI, such as ChatGPT or Google Bard.

3.1.6 Data act

The data act (DA) is the latest amendment to the data strategy regulation process, being published in November 2023. As the name suggests, the DA focuses on data, especially its accessibility and use. Therefore, it may be the most important act regarding the development of European data spaces and data economy.

The purpose of the DA is to facilitate data accessibility and use “and to remove barriers to a wellfunctioning internal market for data.” According to the DA, data holders should enable their users, whether they are private people or organizations, access to the data generated from their use of a service. The users should thereafter gain sovereignty over their own data and the possibility to use or distribute the data as they see fit. The DA also calls for transparent and fair conduct from the data holders. (European Parliament & Council of the European Union 2023.)

In addition, the DA states that “where there is an exceptional need” data holders should provide public bodies with access to their data if it is needed to carry out something “in the public interest” (European Parliament & Council of the European Union 2023). This statement might stir distrust in some of the more sceptic users although it would seem to primarily obligate businesses over individuals. Furthermore, the weight in the DA lies mostly on facilitating the use of other data, not personal data (Bräutigam et al. 2022).

The DA is important for common European data spaces since it lays out guidelines for the three conditions for data economy ecosystems and data spaces: first, who can use the data, second, what can they use that data for, and third, why or at what cost can they use the data. These guidelines need to be clear to all bodies participating in a data space in order to ensure trust and fruitful cooperation. The DA also enforces the other acts, such as the DMA in the sense that it strives to make it easier for people and organizations to change from a “data processing service” to another while maintaining control over their own data (Bräutigam et al. 2022; European Parliament & Council of the European Union

2023). This could also boost the deployment of European data spaces as data owners seek ways to make the most of their data sovereignty.

Before the final release of the DA there was still obscurity regarding the definition of data. There were also questions related to data ownership and the conditions in defining who really owns the data under which circumstances. If and when these issues are tackled, SMEs are beneficiaries in a data sharing environment where data holders are required to let organizations gain access to their data under fair conditions, something many SMEs may not have been able to do before. (Bräutigam et al. 2022.)

Table 1 provides a summary of the legislative initiatives introduced in subchapters 3.1.2–3.1.6. Each act's primary goals are revised, along with the bodies responsible for its monitoring and the time they become applicable within the EU.

Table 1. Summary of EU data regulation

Act	Primary goals	Monitored by	Applicable
DGA	Reuse of public data Guidelines on transferring non-personal data across borders	National authorities	9/2023
DMA	Facilitating competition between gatekeepers and other businesses by imposing high sanctions	European Commission	5/2023
DSA	Clarifying responsibilities of online platforms in content moderation on their platforms	National authorities	2/2024
AIA	Ensure safety of AI systems in the EU market	National authorities	2024-2027
DA	Enable data sovereignty for individuals and organizations 3 conditions for data use: 1. Who, 2. For what purpose, 3. At what cost	National authorities & European Union bodies	9/2025

3.2 Current state of European data spaces

In the Data Act, the European Parliament & Council of the European Union (2023) define common European data spaces as "purpose of sector specific or cross-sectoral interoperable frameworks for common standards and practices". Originally in the European data strategy, the European Commission (2020) presented the establishment of nine European common data spaces: mobility, agriculture, industrial, Green Deal, health, financial, energy, public administration, and skills data spaces. Each sector has unique business motives and legislative requirements. Scerri et al. (2022) also raise the point that there are commonalities between these data spaces as well – an important factor to consider when planning an eventual single, standardized European cross-sectoral data space. The following subchapters provide an outlook on the current status of the nine

common European data spaces based on scientific literature and other references. The outlook is not exhaustive.

3.2.1 Mobility data space

Mobility data spaces have strong grounds in Germany and France where many public and private organizations of the mobility industry have been acting jointly to strengthen the exchange of mobility data between operators (Drees et al. 2021; Pretzsch et al. 2022, EONA-X n.d.a). Intelligent mobility solutions require lots of data, and the benefits of mobility data spaces include improved traffic safety and flow, collaboration opportunities within multimodal travel where passengers combine various means of transportation, and sustainability (Pretzsch et al. 2022; EONA-X n.d.b). The organizations linked within a mobility data space would include public and private organizations, for instance airlines, trains, buses, taxis, electric car charging stations, and car manufacturers dealing in both autonomous and traditional cars. Additionally, individual citizens and other organizations utilizing these mobility services would be participants in mobility data spaces.

The first mobility data space implementations include a project where data space infrastructure is used to provide a multimodal mobility channel between the cities of Berlin and Hamburg in Germany (Drees et al. 2021), along with a German national non-profit mobility data space, “Datenraum Mobilität” (DRM), that was originally implemented in 2020 (Drees et al. 2021; Pretzsch et al. 2022). The DRM incorporates both public and private organizations with an emphasis on car manufacturers, and it promotes increased safety, fairness, and sustainability, along with new monetization opportunities for different organizations (Mobility Data Space n.d.a). These projects are being supported by large research facilities, the government of Germany and the Gaia-X initiative (Drees et al. 2021; Pretzsch et al. 2022; Mobility Data Space n.d.a).

The key components in a mobility data space include a data marketplace, data app store, vocabulary provider, identity provider, and a clearing house (Drees et al. 2021; Pretzsch et al. 2022). These components seem relatively general and are likely to be exhibited in other sectors’ data spaces as well. The components also resemble those of the IDSA framework, adapted in Figure 1, simultaneously supporting the concept of standardization that is often called for in the data space context (see e.g., European Commission 2020; 2023). Currently, the DRM for example, utilizes a platform called “Mobility Data Marketplace” (MDM) which is not yet compliant with the concept of standardization,

since some of the aforementioned components have not been carried out in "an IDS-compliant way" and some components are missing entirely (Pretzsch et al. 2022; Mobility Data Space n.d.b).

Additionally, in Finland, the traffic control company Fintraffic maintains a "Traffic Data Ecosystem". It does not identify as, or fulfil the requirements of a data space, but it shares many of the same motives. These motives include cost and emission savings, data sharing, and improved business opportunities for organizations within the ecosystem. (Lautanala 2024; Fintraffic n.d.)

In conclusion, the development of mobility data spaces has taken an effective start in central Europe, especially Germany. There are already practical use cases available, and more are likely to emerge since the mobility data spaces have the support of the government, Gaia-X, and several large companies, for instance within the automotive industry. The development of mobility data spaces has been depicted in Figure 6.

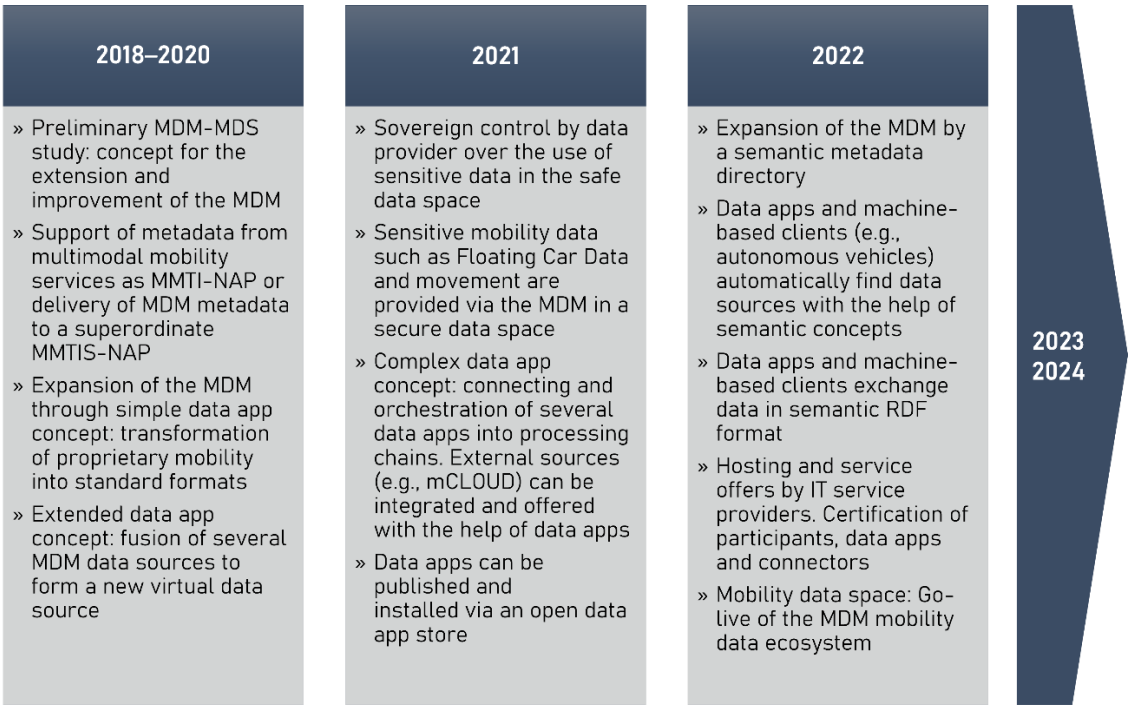


Figure 6. Development of the Mobility Data Space (adapted from Pretzsch et al. 2022)

3.2.2 Agriculture data space

Kalmar et al. (2022) discuss the motives for agricultural data spaces. They point out the need for tackling global hunger through maximizing agricultural yield output. They also

present the issue of sustainability which includes both social and environmental aspects. Agriculture deals with a plethora of variables, such as soil quality and weather conditions, all the way to logistics and cooperation of many parties, machines, and other resources. Overcoming these versatile issues could get easier through a data space accompanied by an “open marketplace for data, digital products, and software services” (Kalmar et al. 2022).

Agriculture data spaces will provide farmers with new value and optimized operations through a multitude of systems collecting and utilizing data from their farms, including “autonomous field robotics” (Fraunhofer Institute 2020). Concrete benefits of an agriculture data space would include discovering the precise carbon balance, origin, and health-related qualities of food products, and improving profitability of farms and even their individual field plots (Hyyrönmäki 2022).

On European level, the EU-funded AgriDataSpace project aims to clear the way for an EU-wide agriculture data space by implementing the familiar EU-values of data sovereignty, trust, security, and standardization (European Commission 2021b; AgriDataSpace 2023). The project has member organizations in ten European countries where different local initiatives are currently developed, including Finland, Germany, France, Poland, Spain, Italy, Greece, Romania, Belgium, and the Netherlands (AgriDataSpace 2023).

Example organizations include both public and private research institutions, and their solutions differ from logistics to crop analyses. For instance, the Finnish company DataSpace Europe provides a private data space platform among other solutions, where farmers can share their data and view other providers’ data. The income model is based on a monthly fee for using the data space, while the platform provider does not gain access to its customers’ data (DataSpace Europe n.d.), hence reinforcing the concept of data sovereignty.

In conclusion, different EU countries have independent agriculture data space initiatives which are following EU-values and are being communicated through AgriDataSpace to some extent. It is likely that a common European agriculture data space might not emerge for a few years still. Meanwhile, the development of these local initiatives is advancing.

3.2.3 Industrial data space

The concept of international data spaces first arose after the Fraunhofer Institute published their whitepaper on an industrial data space in 2014. It was driven from the beginning by the key motives of data sovereignty, decentralization, and data economy. (Otto et al. 2016.) The need for a standardized, secure platform has already been widely recognized for several years (see e.g., Otto et al. 2016; Pullmann et al. 2017; Alonso et al. 2018). After IDSA was established, referring to international instead of industrial data spaces, the industrial data space, also referred to as manufacturing data space, has been recognized as one of the nine central European common data spaces.

In the context of industrial, or manufacturing, data space, the term industry 4.0 arises frequently (see e.g., Mertens et al. 2022; Usländer & Teuscher 2022; EU Data Sp4ce 2024). Industry 4.0 refers to the fourth industrial revolution, where the role of disruptive trends such as data, digitalization, and other technological advancements is underlined (McKinsey & Company 2022). These trends are also recognized drivers in industrial data spaces. Usländer and Teuscher (2022) highlight the need for digitally connected manufacturing equipment to enable better predictive features and maintenance, for example.

Industry 4.0 is also present in the EU-funded initiative Data Space 4.0, or Manufacturing Data Space, which serves as the trendsetter for industrial data space unification (Data Space 4.0 n.d.). It was established by the German-based international Manufacturing-X ecosystem (Platform Industrie 4.0 n.d.). Data Space 4.0 naturally shares many of its goals with the EU-funded initiatives of other sectoral data spaces. It has both public and private member organizations from many European countries (Data Space 4.0 n.d.). Additionally, Data Space 4.0 has already taken the initiative to unite different industrial data space initiatives, which brings it closer to an actualized common European data space than most other sectoral data spaces (Fiware Foundation 2024).

For now, there are still individual data space initiatives in different EU countries. For instance, the Sustainable Industry X (SIX) is a Finnish initiative that promotes industries in Finland in a sustainable way (SIX Sustainable Industry X 2024). In addition, they have established a dataspace that accelerates “the green transition of mobile work machines” (SIX Sustainable Industry X 2023).

In addition, the Catena-X, a German initiative, funded by the EU and German state, operates within the automotive industry with many member companies from a global scale. They describe themselves as a data ecosystem instead of a data space. (Catena-X n.d.) They do, however, lay groundwork for a Catena-X automotive industry data space as well (Catena-X 2023).

3.2.4 Green Deal data space

The Green Deal initiative is a project which strives to make Europe the first climate-neutral continent in the world. The role of data in this project is also highlighted as the issues of "climate change, circular economy, zero-pollution, biodiversity, deforestation and compliance assurance" are dealt with. (European Commission 2020.) The Green Deal data space was introduced to ensure the Green Deal goals are achieved. Additionally, the European Commission (2020) introduced related projects "Destination Earth" for creating a digital replica of the Earth, and "GreenData4All" to enable better infrastructures for sharing spatial data within the EU.

To establish a European Green Deal data space, the EU-funded GREAT project was carried out from September 2022 to February 2024 to lay groundwork for a Green Deal data space (European Grid Infrastructure n.d.; The GREAT Project n.d.). The Green Deal Dataspace was established in September 2023 and its key objective is ensuring the attributes of "resilience and sustainability" in both public and private organizations in the long run (Green Deal Dataspace 2024b). The Green Deal dataspace has members from multiple European countries, albeit the number of organizations is smaller than in some other data space initiatives (Green Deal Dataspace 2024a).

Current use cases of the Green Deal Dataspace include resilience-enforcing measures, such as comprehensive risk management and disruption forecasting, mixed with social sustainability programs such as improved inventory planning and supply chain efficiency. Additionally, extensive data is required to calculate total emissions of production lifecycles, and the Green Deal Dataspace helps organizations track down these emissions, thus supporting environmental sustainability goals. (Green Deal Dataspace 2024a.)

The Destination Earth project aims to support the Green Deal initiative by creating a "digital twin" of the Earth by 2030 (Destination Earth n.d.). This digital Earth could be used for simulating and monitoring natural phenomena and human actions to support

sustainability (European Commission 2020). Destination Earth utilizes data lakes as a data storage method (Destination Earth n.d.) and it that sense differs from data spaces.

The ”GreenData4All” initiative is also receiving EU funding under the name “All Data 4 Green Deal” (AD4GD). Its goal is to provide access to data that supports a set of sustainability goals, and the initiative is running from September 2022 until August 2025, coordinated by a Spanish body. (European Commission n.d.a) The AD4GD is co-designing the common Green Deal data space and drives data collaboration pilots to support resolving issues related to pollution, biodiversity, and climate change with a dozen partner organizations from European countries (AD4GD n.d.).

3.2.5 Health data space

The goals of a common European health data space can be roughly divided into two categories: first, to improve individuals’ control over their health data, and second, to enable re-use, or secondary use, of health data for improving healthcare services (European Commission 2020). The health data space is a large undertaking, and there has been criticism about the scale of the endeavor. For instance, Marelli et al. (2023) find the project too large and suggest it should be divided into smaller segments. The authors suggest that the first proposal for the data space was on a course to make matters worse in the healthcare sector, compared to the status quo.

Due to the sensitive nature of individuals’ health data, the health data space is also one of the most difficult data space projects to carry out. The data space faces different legal constraints, for example regarding patient privacy, that need to be thoroughly scrutinized (Horgan et al. 2022; Marcus et al. 2022; Shabani 2022; Marelli et al. 2023). Additionally, there are already existing issues when data is transferred between healthcare providers and organizations, especially within complex systems in large countries (Berlage et al. 2022) or between countries. Marelli et al. (2023) also recognize the importance of digital integration within the healthcare sector, but simultaneously they emphasize the need to learn from past mistakes and to operate on a manageable scale first. A cautionary example in Finland was the case of the Apotti system implementation, with issues leading to severely weakened patient safety at worst (Yle n.d.). In consequence, objectives for the health data space presented in the European data strategy were to complete a framework to govern activities within the data space, such as access to data, and to provide the

required technical infrastructure to facilitate interoperability among the data space participants (European Commission 2020).

If successful, the health data space could present a versatile set of benefits within the EU, the paramount being improved, more efficient treatment for citizens (European Commission 2020; Berlage et al. 2022; Marcus et al. 2022). Additional advantages include opportunities for research and innovation, resulting in more effective medical products and drugs, along with more conscious policymaking, enabling all-round cost savings and better decisions regarding public health (European Commission 2020; Marcus et al. 2022).

The European Health Data Space (EHDS) was launched in May 2022. It is still a work in progress with new governance mechanisms in the planning. The EHDS also introduces some function examples, including healthcare in a different EU country, and data transfer between private and public healthcare providers (European Health Data Space n.d.) Additional health data space initiatives are also active, including the Sphinx project and Health-X, for instance.

The Sphinx is an EU-funded cybersecurity project which aims to increase security within the context of healthcare information technology. The project results have been piloted in three EU countries. (Sphinx Project n.d.) When compared with the original health data space goals, the Sphinx addresses the need for reliable infrastructure to ensure secure data sharing across the data space.

Health-X is a local health data space initiative from Germany. Its goals include data sovereignty and secondary use of health data, similar to the EHDS. For now, however, its partner network consists of only German healthcare and technology organizations. (Health-X n.d.) In addition to Health-X, the Finnish health data system OmaKanta shares some of the same goals as the EHDS, on a national level, focusing on transferring health data between private and public healthcare providers (OmaKanta n.d.).

To conclude, a common health data space is a complex initiative where legislation, privacy, and security play crucial roles. There are implementation projects on national level, such as the Health-X, alongside the established European Health Data Space. Furthermore, there are EU-funded supporting projects, such as the Sphinx Project, that aim to fill the gaps in health data space development.

3.2.6 Financial data space

Like health, the financial industry is also a highly regulated industry with a keen focus on data management (Gaia-X 2021) and special connection to personal data (Chomczyk Penedo & Trigo Kramcsák 2023). This complicates the position of a common European data space whose goals include strengthening "innovation, market transparency, sustainable finance, - - access to finance for European businesses and a more integrated market" (European Commission 2020). Hence, for now, there seems to be no openly published financial data space platform yet, but the construction of regulatory tools is underway.

In June 2023, the European Commission gave its proposal on a "Financial data access and payments package" (FIDA) aiming to guide the field of European finance towards a more technologically advanced era, with focus on electronic payments, open banking, and fraud prevention (European Commission 2023a). Additionally, "the Digital Operational Resilience Act", or DORA regulation, will be applied from January 2025 onwards and is currently in a test phase. DORA strives to improve the cybersecurity of "financial entities such as banks, insurance companies and investment firms" and to strengthen European financial sector's digital resilience against major disruptions. (EIOPA n.d.) These two regulatory tools are laying groundwork for a financial data space.

Bassens and Hendrikse (2022) also estimate that current regulations, such as the DMA, may not be enough for European financial sector organizations to break free from the influence of the large American gatekeeper companies, who seem to have a prevailing control over critical infrastructure. This assessment further underlines the need for additional regulatory measures to enable the emergence of a European financial data space. Furthermore, the European Commission has formed an expert group to oversee the establishment of the financial data space and related issues (European Union 2022).

The specific use cases of the European financial data space could include, for instance, improving the mortgage market for citizens, providing data for creditworthiness assessments of SMEs, making financial advice provision easier, decreasing the costs and footprint of consumers' energy consumption, and increase insurance companies' opportunities to provide better services and safety through, for example, vehicle data (European Union 2022). In addition, Chomczyk Penedo & Trigo Kramcsák (2023) motivate the improved opportunity to use AI tools for informed decision making. In the

financial sector this would include, for instance, credit evaluation and claims management. The Germany-based Financial Big Data Cluster (FBDC) is oriented towards AI optimization and cooperates with Gaia-X. Their chosen data structure, however, is a data pool rather than a data space. (German Federal Ministry for Economic Affairs and Climate 2024.) Vast datasets provided by the data space, utilized by AI tools, are also likely to create new risks given the sensitive nature of the data. Therefore, the need for protective measurements within the financial data space is stressed. (Chomczyk Penedo & Trigo Kramcsák 2023.)

3.2.7 Energy data space

The energy sector is a highly data-intensive and complex industry since demand and supply must always be closely matched (Berkhout et al. 2022). This requires efficient data collection and sharing, both consumption and production-wise. The European energy data space could provide a solution to overcome the complexity and ensure a sustainable and efficient future for the energy sector (Berkhout et al. 2022; Gouriet et al. 2022).

The goals of a European energy data space include better data sharing in the energy industry to enable new innovations, facilitating the change towards renewable energy sources, supporting businesses, producing new services to Europeans, and accelerating the decarbonization of the European energy sector (European Commission 2020, Gouriet et al. 2022). The use cases for reaching these goals comprise, for instance enhanced communication and data sharing between stakeholders and predictive measures like predictive energy supply and maintenance (Berkhout et al. 2022).

The energy data space would incorporate the whole energy value chain as its stakeholders, including operators within all energy networks: gas, electricity, and heat (Gouriet et al. 2022; Omega-X n.d.). These stakeholders vary from consumers to providers, including other partners, contractors, aggregators, open services like maps and public transport, and even electric vehicle stations (Gouriet et al. 2022).

There are several energy data space initiatives in Europe already taking shape. These include, for example, the EU-funded data space projects Omega-X and Enershare. Additional schemes, such as the Synergies project and Platoon project are also supporting data sharing within the European energy market.

Omega-X's motive is to establish a data space within the EU that would support "energy autonomy" and reduction of carbon emissions (Omega-X n.d.) Enershare has similar aspirations. They are creating a reference architecture that would be compliant with IDSA, Gaia-X and FIWARE models. Their focal point is to improve trust within a data sharing ecosystem, and they are implementing the infrastructure through blockchain technology. (Enershare n.d.)

The Synergies project is not a data space project per se, but they are promoting a "data-driven intelligence ecosystem" with similar motives as the data space initiatives (Energydataspaces 2024). The Platoon project, now concluded, strived to improve digitalization within the energy sector through seven pilot projects in five European countries (Platoon project n.d.), simultaneously supporting the goals of a common European energy data space.

In conclusion, the energy sector is at a good pace with its data space initiatives. There are multiple projects with common motives, and potential synergy benefits in the future if a single European energy data market is implemented. The goals are strongly related to sustainability and therefore joint efforts with the Green Deal data space could be advantageous.

3.2.8 Public administration data space

The public administration data space was introduced in the European data strategy to support a multitude of public sector operations: public spending, corruption reduction, law enforcement, and enabling better technology applications for the public sector (European Commission 2020). Generally, smaller public administration bodies, such as towns, have poor prerequisites to utilize data in their decision-making compared to larger ones, for instance due to the scattered nature of data management within their operating environment (Nanni & Napolitano 2024). A common public administration data space could facilitate the use of data for smaller public administration bodies as well, since the resources would be found more easily in one place.

The current implementation of the European public administration data space has been divided into two different spaces: Public Procurement Data Space (PPDS) and Data space for security and law enforcement. The PPDS aims to facilitate the use of public procurement data, which is currently spread across different EU member countries

making it difficult to use. The benefits of this data space for public buyers include bulk benefits for concentrated procurement enterprises, improving the attraction of tendering competitions for businesses, enhanced tracking of potential corrupt actions, supporting sustainability and innovation goals, and reaching significant savings through improved digitalized and subsequently automated tasks. (European Commission 2023b.)

The security and law enforcement data space has different goals albeit both initiatives support public administration. With this data space, the focus is on improving opportunities of using AI algorithms for security purposes. The specific goals involve prevention of hostile third-party activities, enhancing law enforcement's technical abilities within the EU member countries, and setting quality levels within the EU. Due to the nature of the field, special attention is paid to the training of AI algorithms to minimize their bias and prejudice in law enforcement. (European Commission n.d.b.)

The public administration data space has been divided into two subspaces due to the differing nature of data handled on each platform. Reducing corruption and increasing transparency, along with improved technological solutions are common aspirations for both data spaces, however.

3.2.9 Skills data space

"The skills of its people are Europe's strongest asset". To keep up with new and upcoming requirements, lots of data is needed regarding work, learning, skills, and qualifications in European countries. (European Commission 2020.) Zillner et al. (2021) also underline the need to minimize incompatibility between what is taught in education and what is actually needed in organizations.

In the European data strategy, the European Commission (2020) stated to support the establishment of common and digital credential, qualification, and learning opportunity frameworks. One outcome of these efforts is the European Learning Model (ELM) which aims to standardize the classification of learning and skills in European countries (European Union n.d.)

To reach these goals, the 1-year DS4Skills project was launched in October 2022. Its goals included, among others, to enhance European data sharing related to skills, for topical training programs and to provide job and reskilling recommendations (Digital Europe 2022). The DS4Skills ecosystem includes 14 partner organizations from different

European countries (Data Space for Skills n.d.). According to the schedule, the project concluded in October 2023, and there seems to be no further information on the project outcomes yet.

In conclusion, some fields, such as industry, agriculture, and mobility, seem to be at higher maturity levels than the strictly regulated health and finance data spaces. There are already many initiatives that include organizations from multiple countries. However, despite extensive regulations taking effect, no sector-specific European data space is at the point of cross-border deployment yet and therefore a common European single data market is still a distant abstraction for the time being. That being said, Germany and France, who have been trendsetters in the European data space endeavors, are relatively far along in developing some sector-specific data spaces that may stretch across national borders. Additionally, joint efforts between different sectoral data space projects could be beneficial: an example is also presented in the next chapter.

4 Finnish organizations in the field of data spaces

At this point we have discussed what data spaces are, why they were introduced, how they are related to data economy ecosystems, and what potential opportunities and challenges have been anticipated for them. In addition, we have discussed data spaces from the perspective of the European Union and how they tie to the European data strategy. The Big Five regulations that followed the strategy, data governance act, digital services act, digital marketing act, artificial intelligence act, and data act, were discussed, simultaneously introducing the implications they might have for the nine common European data spaces that were originally introduced in the European data strategy. After that we provided an outlook on the current status of the nine common European data spaces.

Now we move on to the core of this thesis: the role and implications of Finnish organizations within these European data spaces. Although many large data space projects discussed in this thesis have their roots in Germany, especially within mobility and manufacturing industry (e.g., DRM and Catena-X), some Finnish projects have also appeared. These include for instance the agriculture data space platform by DataSpace Europe, SIX Mobile Machines cluster, and the IOXIO Dataspace. These current Finnish data space projects will be discussed in the following paragraphs.

In 2022, DataSpace Europe introduced their data intermediation platform Tritom for farmers in the agriculture industry. The goals of their platform included determining the exact carbon footprint of food production and origin of individual food packages. According to the principles of data spaces, Tritom would not store the data in itself but only mediate it across the data space. (DataSpace Europe 2022b.) DataSpace Europe also joined the IDSA and Gaia-X organizations in 2022 (DataSpace Europe 2022a).

To support the establishment of Tritom, DataSpace Europe has initiated a program that engages Finnish farmers to make use of their data (Sinipuro 2024). The company states that farming already produces plenty of data through different software applications, machinery, and other agriculture stakeholders and that sharing this data within an ecosystem could enable new business development opportunities (DataSpace Europe 2022b). The farmer engagement program strives to render itself needless in time with

wishes of these data economy practices becoming conventional within the farming ecosystem (Sinipuro 2023).

In January 2024, the company declared that they have become “Europe’s first registered data intermediation service”. The Finnish Ministry of Transport and Communications acknowledged that the company fulfils the requirements of a data intermediation service as stated in the data governance act (DGA). (Sinipuro 2024.) These requirements are given in DGA article 12 including, for instance, that the provider cannot use the participants’ data for its own purposes, and that the provider strives to facilitate interoperability among the relevant ecosystem through different actions (European Commission 2022a).

Now we present an example of a cross-sectoral data space initiative as stated at the end of last chapter: SIX Mobile Machines cluster is an industry-driven data space initiative that brings together companies from different sectors, such as manufacturing, forestry, and mining to enhance the production of efficient and sustainable work machines. In the initiative data from the batteries of mobile work machines was collected for the purpose of discovering how the usage of those batteries could be maximized. The parties involved in this data space are manufacturers, users, and recyclers of work machines and include businesses such as Kalmar, Ponsse, Sandvik, and Valmet Automotive. (SIX Sustainable Industry X 2023.)

The SIX Mobile Machines project has been supported by the Finnish company IOXIO that also distributes its own data space solution within the industrial and public sectors (Sitra 2023). The goal of the IOXIO Dataspace is to provide a lighter structure for sharing data between organizations instead of previous integration solutions (IOXIO 2024). This aim reflects the goals of Franklin et al. (2005) when they originally proposed data spaces as an alternative for other data integration systems requiring notable upfront effort.

Additionally, there are several projects and initiatives in Finland that accelerate data sharing and data economy but do not necessarily identify as or fulfil the requirements of data spaces. These are led by both private and public organizations and promote commercial and non-commercial goals alike. Example domains include road and marine traffic and mobility, smart manufacturing, sports, retail, sustainability, education, and labor markets. The most prominent schemes have been compiled in the “Most interesting data economy solutions” list. (Sitra n.d.)

As seen in Table 1 in subchapter 3.1.6 the “Big Five” data legislation is being applied presently and in the near future. These legislative drivers will pose consequences for all businesses in terms of data management both by asserting new obligations and by providing novel data-related business models and cost-saving opportunities. Therefore, now is the ideal time to inspect the status of the data space and data economy field, and the implications the data legislation is likely to have on different organization types.

According to a recent study by the Finnish company 1001Lakes (2024) steady business within data economy could still take a few more years to become truly established. However, they point out two important things: first, that the European Union is likely to place significant investments into data economy, including European data spaces introduced in the data strategy; and second, that the development of these data economy projects are at an early stage and for example data spaces and their specifics are still incomplete. They also see that Finland is strong in data and AI domains and that now is the chance to venture into new data-related businesses. In a few years, as these new opportunities are realized, Finnish companies could reap the benefits as pioneers in the field. (1001Lakes 2024.)

Although several data economy projects have already arisen in Finland, data space initiatives still seem to be scarce. In chapter 3 it was discovered that despite extensive efforts in many different fields, concrete use cases of European data spaces are still yet to take off. In addition, the only officially registered data space platform provider in Europe seems to be the Finnish company DataSpace Europe that currently operates within the field of agriculture. These findings lead to an interesting scenario, where regulations are taking effect and contributions have been made to research and funding, but the practical implications are still inadequate.

At this early stage it would be intriguing to discover what potential opportunities lie within this business for both public and private organizations, especially for Finnish actors. Through empirical methods this thesis aims to answer the research question and uncover concrete opportunities data spaces could present for Finnish organizations now or in the future. This is done to bring awareness of the concrete potential of data space activities so organizations can make swift, yet informed decisions about whether or not to join data space networks, produce data space services or participate in some other way.

5 Methodology

5.1 Research approach

The empirical segment of the study uses a qualitative approach to map the opportunities data spaces present for Finnish organizations. Qualitative research was chosen for this thesis because of the subject's novelty, as qualitative methods are used as preliminary research methods for comprehensively inspecting and discovering a topic of interest (Hirsjärvi et al. 2013, 136–161) and for forming an understanding of research areas that have not yet been thoroughly explored (Eriksson & Kovalainen 2008, 4–5).

Instead of utilizing a specific qualitative research method, such as case study, grounded theory, or action research, a combination of semi-structured expert interviews and thematic analysis was implemented in the thesis to collect and analyze data. The following subchapters introduce and justify the chosen methods more closely. First, data collection and analysis methods are introduced through literature. Next, the data collection and data analysis processes implemented in this study are discussed. Last, subjects related to research quality, limitations, and ethics are explored.

5.2 Method

For data collection, a round of semi-structured expert interviews were conducted to interview experts in the field of data spaces and data economy from public and private Finnish organizations. Interviews have been established as the primary data collection method in qualitative research. Semi-structured interviews enable a relatively free-form, but still theme-driven, discussion between researcher and participant which suits this so far modestly examined subject. (Hirsjärvi et al. 2013, 205–209.)

In general, semi-structured interviews have the ability to cover topics that have not yet been objectively tested, and they can provide information on those topics on a universal level (Hirsjärvi & Hurme 1995, 15, 40–41). The findings from semi-structured interviews can also be inductively analyzed to create generalizations and meaning through an iterative process (Galletta & Cross 2013, 18). These qualities match this thesis' objectives well.

Here, the main purpose for choosing a semi-structured interview as the data collection method is that it offers the participants more room to produce interpretations of the

questions and enables opportunities to clarify message meanings both ways (Hirsjärvi & Hurme 1995, 15), especially compared to written surveys. All in all, semi-structured interviews are a flexible method suited for many different scenarios (Hirsjärvi & Hurme 1995, 15; Galletta & Cross 2013, 45).

Expert interviews, on the other hand, may prove beneficial when the time for conducting interviews is limited, as the obtained results can still be good. In addition, motivating experts to participate in interviews is generally easier due to "common scientific background or relevance system" between the researcher and participant. On the other hand, expert interviews may lead to "legitimization of social hierarchies" if expert insights are given invalid, extensive weight. (Bogner et al. 2009, 2–3.) The main purpose for choosing expert interviews as a method for this study was that the data space subject is relatively new, and deeper understanding is still limited to a small number of professionals.

There are still risks in using interviews as a data collection method as well. Although interviews allow clarifying messages throughout the conversation, it is still possible that interpretive mistakes may occur, for instance if either party "plays with words and expressions" or the parties lack common subject-related vocabulary (Hirsjärvi & Hurme 1995, 48). The latter risk was relatively small in this thesis and presented no issues as both the researcher and expert participants were familiar with the necessary vocabulary.

Another issue for collecting empirical data from participants could be their lack of motivation to participate in the study. Gorden (1969) argues that motivating participants is easier in interviews than in form studies, for instance. Bogner et al.'s (2009, 2–3) earlier statement also support this. Given that the subject of data space utilization in Finnish organizations is new, and the participants have professional and personal interests in the subject, motivation was not an issue. To support this argument, of the eight contacted potential participants five responded, all with affirmative answers to participate.

The interview also involves aspects of a "focused interview". Its four features, as introduced my Merton et al. (1956) are: first, "interviewees are known to have been involved in a particular situation", here referring to data space research and / or development; second, "investigator has provisionally analyzed situation", here meaning the review of literature and other data space-related materials in chapters 2, 3, and 4; third "this situational analysis provides basis for interview guide" as it determines key themes

for interpreting interview responses; and fourth, “interview focuses on subjective experiences” as it allows the participants to present their personal estimations on data space opportunities based on their own experiences.

The data analysis method chosen for this thesis is thematic analysis (TA). TA has been broadly studied and defined by Virginia Braun and Victoria Clarke (see e.g., Clarke & Braun 2017; Braun & Clarke 2022). They interpret TA as “a method for identifying, analyzing, and interpreting patterns of meaning (‘themes’) within qualitative data” and at its core is a goal to recognize and explain important data properties as directed by the thesis research question (Clarke & Braun 2017). They also point out that in TA, the research aim, formulated in the research question, can change as data analysis progresses (Clarke & Braun 2017; Braun & Clarke 2022, 294).

TA is a flexible analysis method suited for both large and small qualitative datasets with an analytical focus on “text and meaning”, simultaneously incorporating the possibility of partial or multiple truths (Clarke & Braun 2017; Braun & Clarke 2022, 6). A form of thematic analysis, “inductive thematic analysis”, makes generalizations from the data instead of previous theories (Braun & Clarke 2022, 288). Additionally, Braun and Clarke (2022, 294) have paid special attention to “reflexive thematic analysis”. It is distinctive for reflexive TA that the researcher is in an active role and that their subjective views are certain to have an effect on the data analysis. The term reflexivity also incorporates the idea that the researcher must be critical about how their “disciplinary, theoretical and personal assumptions and their design choices” affect the study. (Braun & Clarke 2022, 294.) Reflecting on the effects of one’s choices and how they shape research findings is also highlighted by Eriksson and Kovalainen (2008, 4–6).

5.3 Data collection and analysis

5.3.1 Collection

Five participants took part in the interviews. They were carefully selected from leading experts in the field of data spaces and data economy in Finland. The participants work for Finnish research institutions, public bodies, and private businesses. They were contacted through email messages based on their publicly available contact information.

One of the interviews was held in English and the other four in Finnish. The interview duration was 60 minutes apart from one interview, which lasted approximately 35 minutes due to the schedule of the participant. Hirsjärvi et al. (2013, 206) argue that a half-hour interview is generally too short and should be replaced with a questionnaire. In this interview, however, the subject was in the participants' field of expertise, they had an opportunity to familiarize themselves with the thesis scope beforehand, and the discussion required very little personal communication between interviewer and participant. Therefore, I found even short interviews to be the best choice as they allow further clarification and mitigate misunderstandings.

The interview structure varied slightly in each interview based on the participant's area of expertise. As proposed by Hirsjärvi and Hurme (1995, 41), instead of a detailed set of questions, a list of thematic areas was compiled beforehand. These thematic areas included the types and sizes of organizations and the different business sectors based on the nine common European data spaces, and they were used to categorize responses. Additionally, some general questions were presented to each participant to map similarities and differences in their views. After covering a specific thematic area, I closed by asking if the participants had anything to add, as suggested by Galletta and Cross (2013, 52). This worked well as most participants had something further to tell that had not been covered earlier. An indicating interview structure can be seen in Appendix 5.

After the interviews, the recordings were transcribed using an AI assisted transcription tool provided by the University of Turku. After the transcription, each interview was manually verified. There were some sound-quality issues with one interview which required thorough manual verification. Some parts remained unclear but most of the interview was understandable. Appendix 4 gives more information on the processing of audio and video recordings.

5.3.2 Analysis

As discussed previously, reflexive thematic analysis incorporates a level of subjectivity to data analysis. In this thesis, subjectivity prevails in two ways: first, in the data itself, as each participant represents the ideologies and expertise of their own person, organization, or field; and second, in the interpretations and generalizations made from the results by the researcher. In addition, inductive TA also suits this thesis as the goal is to discover and concurrently classify previously unidentified opportunities related to data spaces.

Since reflexive and inductive TA are not mutually exclusive, they can both be viewed as applied analysis methods in the thesis.

Braun and Clarke (2022, 35–36) also introduce six phases of reflexive TA. First, the author must become familiar with the dataset. Second, the dataset is coded by recognizing "potentially interesting, relevant or meaningful" parts and labeling them descriptively. Third, preliminary themes are created by grouping together codes with some shared features driven by the research question. Fourth and fifth stage include revisiting the entire data to ensure the themes are well justified and accurately named. The sixth and last phase is writing the analysis based on the developed themes. (Braun & Clarke 2022, 35–36.) This structure has driven the analysis process of this thesis.

First, the transcribed text files were read thoroughly, and each interesting notion was marked with an attached comment. Next, the comments were transferred into a spreadsheet, where they were paired with keywords. After that, driven by the main research questions, the codes showing potential data space-related opportunities were labeled, filtered, and moved to a new, separate spreadsheet.

Braun and Clare (2022, 4, 77) highlight that themes are the fundamental goals of TA and that in reflexive TA, "a theme has to capture a wide range of data that are united by, and evidence, a shared idea", in other words a "central organizing concept". Theme development also requires organized coding of data. (Braun & Clarke 2022, 4, 77.) Based on these notions, the opportunity spreadsheet was color-coded, and resembling opportunities were divided under descriptive themes based on the benefits of the opportunity, the applicable data space or spaces, and whether the organizations affected were public, private, or both. After this, the codes were grouped together and connected with other codes under the discovered themes, as seen in Table 2 in chapter 6.

After the initial framework, the focus was moved back to the full set of codes. This time the purpose was to discover insights from the other codes as well, not only opportunities. The remaining codes were then labeled as either restraints or other important findings related to data space activities. The opportunities depicted in Table 2 along with recognized restraints and other findings were then written open in the results chapter, as instructed in the sixth phase of TA.

5.4 Research quality, limitations, and ethics

5.4.1 Quality and limitations

Eriksson and Kovalainen (2008, 294) refer to the work of Lincoln and Guba where they introduced a revised set of trustworthiness criteria for qualitative research in 1985. There are four criteria: dependability, transferability, credibility, and conformability. These criteria are inspected in the following to help evaluate the quality of this thesis.

First, dependability means providing the reader with information regarding the entire research process and that it is "logical, traceable and documented" (Eriksson & Kovalainen 2008, 294). The previous subchapters provide information regarding the data collection and analysis processes, supported by the data management plan in the appendices. Regarding previous literature and other references, systematic APA style references have been used throughout the thesis to ensure dependable reference management.

Second, transferability depicts how well the study connects with other, previous research. The key is not to recreate previous work but to find some level of comparability. (Eriksson & Kovalainen 2008, 294.) The entirety of chapters 2, 3, and 4 lay groundwork for this thesis based on previous research and current endeavors. Chapter 7 discusses this work's empirical results with prior research, such as the findings discussed in earlier chapters.

Third, credibility evaluates the logic of interpretations made in the thesis. In practice, credibility assesses if some other researcher could come to same or resembling conclusions by using the same materials. (Eriksson & Kovalainen 2008, 294.) In an effort to facilitate credibility, the stages of data analysis are discussed in chapter 5.3.2. Additionally, Table 2 in chapter 6 depicts clear codes derived from the interview responses to provide a level of transparency to the analysis.

Fourth, and last, conformability shows that the interpretations made from the data have truly been made based on the data, here the interview responses, and not invented for example during the analysis process (Eriksson & Kovalainen 2008, 294). The aforementioned codes in Table 2 are all indirect quotes from research participants. The themes developed from those codes, however, show researcher subjectivity as is distinctive for reflexive thematic analysis. To endorse conformability, some direct interview citations are also provided in the results chapter.

This thesis has limitations that should be considered in making generalizations and planning future research. The literature review depicts a point of time when data space use cases remain rare and frameworks and regulations, such as European data legislation, are still incomplete or at least their full implications unclear. Additionally, the interviews focus on the role of Finnish organizations and although international and European data space players and initiatives were also discussed, no empirical material was produced on them. Furthermore, the inspected data spaces are limited to the nine that were introduced in the European data strategy (European Commission 2020). Those nine were mobility, agriculture, industrial (manufacturing), Green Deal, health, financial, energy, public administration, and skills. What is worth noting, is that this is not an exhaustive list of domains that have potential for data spaces but an initial goal for the European Union on their road towards a single European data market.

5.4.2 Ethics

In the study, I strive for objective and ethical research practices. This chapter addresses that statement and discusses the ethics of the empirical segment of the thesis. Each participant took part in the study voluntarily. The interview responses have been protected by pseudonymity so that no individual participant can be recognized by their response. As the participant target group is limited, the term pseudonymity was used instead of complete anonymity. The responses have been stored on the author's personal computer protected by a password and backed up on University of Turku's server and transcribed on an AI tool that also operates on the server, ensuring that the data remained secure throughout the process of data collection and analysis.

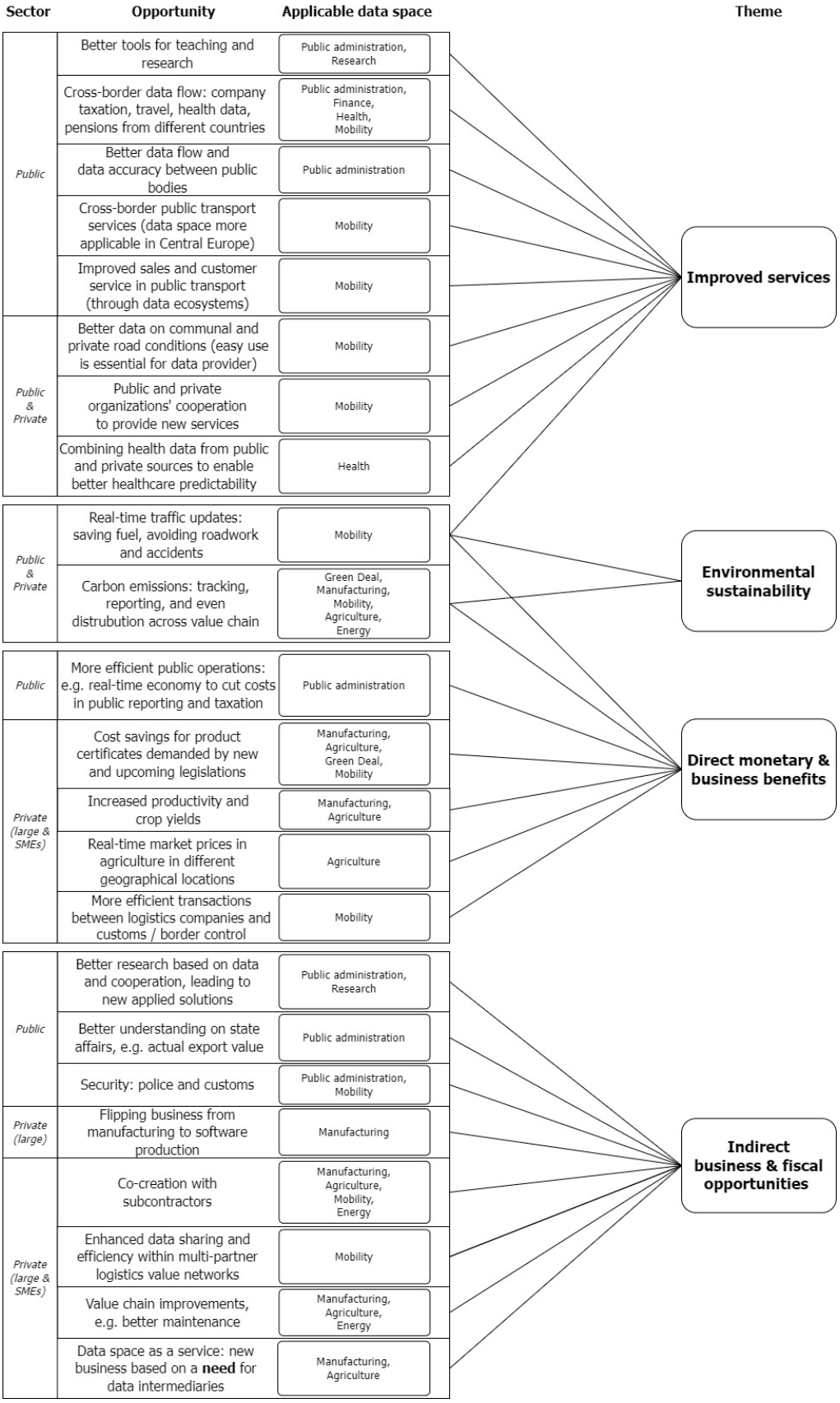
To ensure ethical data processing and reporting, most indirect identification data has been removed to maximize the participants' privacy. The participants were provided with a privacy notice and consent form via email before the interview, which they all accepted either in writing before the interview or verbally in the beginning of the interview (see Appendix 1, Appendix 2, and Appendix 3). They were also informed of their liberty to withdraw their participation at any time. A data management plan was also developed according to university guidelines, marking how each piece of data is handled and destroyed (see Appendix 4). A potential issue worth noting is that all participants were men. This was due to the fact that the majority of the experts in the target organizations were men, yet it is an issue that should potentially be addressed closer in the future.

6 Results

This chapter presents the results of the interview study. Regarding discovered opportunities, four themes were formed: improved services, environmental sustainability, direct monetary & business benefits, and indirect business & fiscal opportunities. The opportunities and corresponding themes are depicted in Table 2. Additionally, a lot of discussion arose in the interviews regarding other data space-related implications, such as restraints and other important findings. All these findings are discussed in the following subchapters.

When depicting interview results, the five participants are referred to as P1, P2, P3, P4, and P5. Apart from one participant, all direct citations presented in the chapter have been translated to English with the help of Google Translate and manually verified. Another thing worth noting is that the mobility data space includes logistics in this context, and that the nine sectoral data spaces are used as a categorizing tool but are not exhaustive and may not depict all domains where data spaces and their implications could be applicable. Moreover, the tenth additional data space, research, has been considered in the results since its potentialities were many and data space activities are still largely on research level.

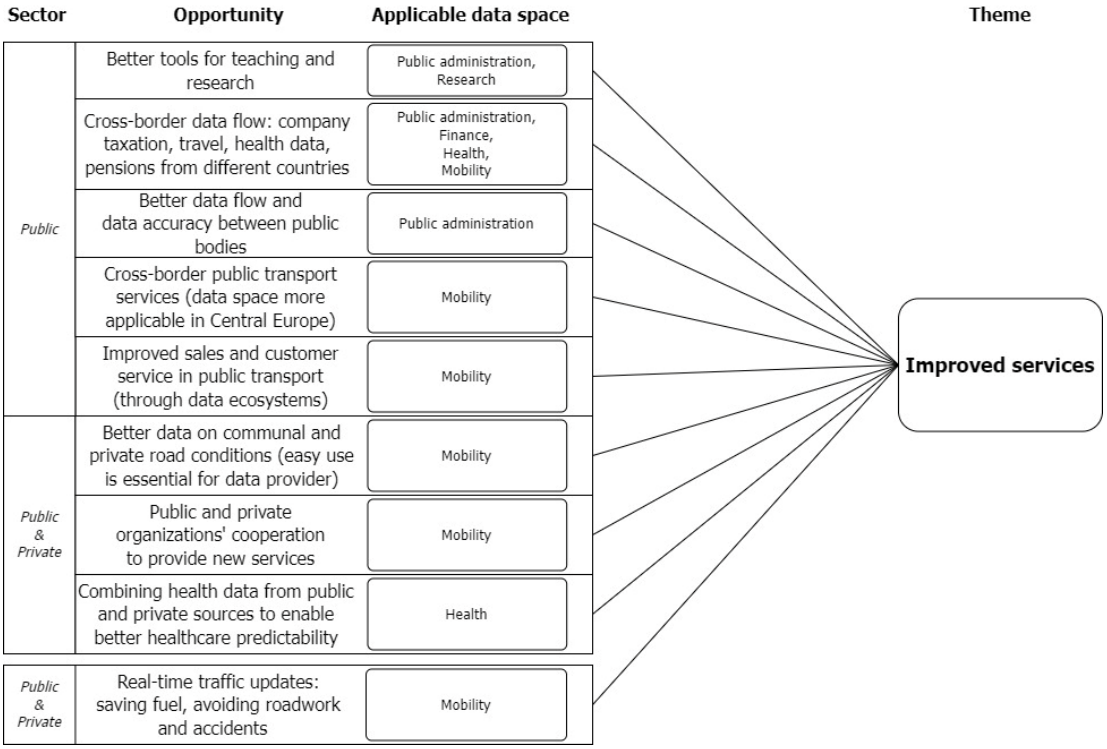
Table 2. Opportunities of data spaces for Finnish organizations



6.1 Improved services

Improved services incorporate both public and private organizations and the applicable data spaces include at least mobility (including logistics), public administration, finance, health, and research. Even if the relevant sector in the framework was public, the beneficiaries of these data space-enhanced services could still be from the private sector. In the following, we will first inspect public sector opportunities and then look at the ones combining organizations from both public and private sectors.

Table 3. Improved services



The first service-related opportunity within the public sector is better teaching tools and quality. As a response to the question: "Who benefits if public bodies get into it [data space activities]", P4 answered:

> "Well - - the researchers definitely. Students. And then if we think about what that data could be used for. Both for research, but also for education. In other words, teaching tools."

The next opportunity is improved cross-border data flow between public bodies. This is sure to have a plethora of applications but the examples that arose in the interviews included company taxation, traveling, health data, and pensions data regarding work carried out in different countries. Also, data flow between different Finnish public bodies

could improve, simultaneously enhancing data accuracy. This issue and solution were formulated by P2 as follows:

> "…each actor builds these kinds of own registers. And the data exists many times, where it almost inevitably happens that the data is wrong somewhere. In Finland, there are excellent basic registers to know the latest information and dates of birth and such things. But then this related information - - it would be much smarter if we had, for example, a data space for the state administration or the entire public administration. Even so that the personal information would then be shared through some kind of personal fact wallet through consents."

An important observation when inspecting these services is that the benefits related to cross-border public transport services between Finland and neighboring countries may be achievable through a data economy ecosystem with point-to-point integrations instead of a data space structure as there are not many countries at play. P2 depicted the subject as follows:

> "In Finland, this is actually not a particular problem, because there is very little cross-border public transport…But in Central Europe this is a real problem. The borders are quite close and there is a lot of cross-border traffic."

According to P2, a single national access point within mobility can be enough when there are only a few participants that share data with each other, but in Central Europe a data space between many countries could be more meaningful. The same was found to be relevant in public transport inside Finland where public bodies share data one-way, but the end-users do not share data back. These opportunities were added to the framework as they represent the potential benefits of data sharing, even if the most suitable structures in Finland were point-to-point integrations instead of a data space. Additionally, the opportunity labeled "real-time traffic updates" presents potential for many different organizations, but from the perspective of public bodies these real-time updates are also a service improvement enabled by data sharing, whatever the chosen method.

Next, the remaining improved service opportunities are covered that consider both public and private organizations. The first one is also related to mobility and logistics: better data on road conditions. According to P2, Finland's roads are owned by three sets of entities: the state, towns, and private owners. Especially the privately-owned roads became under focus in P2's statements:

> "…so now there is no clue of what is happening there, even though they are extremely important for example for forest companies. That they know whether a log truck can go on that road or not."

P2 also stated that there is a need for such a platform where information on road quality is shared by all three road owner entities. Current restraints for this include, for instance, data capture and funding. When discussing data capture in another context, P2 stated that sufficient incentives for the data providers are currently lacking and that potential incentive options include either rewards for providing data or sanctions if they fail to do so. Therefore, as is mentioned in Table 2, ease of use of the data sharing service was found vital so that the inconvenience for the data providers would be as small as possible.

An improvement opportunity on health services also came up in the interviews, where clinical data would be combined with health data that is collected personally, for example through smart watches, phones, rings, and others, to increase healthcare predictability. This presents opportunities for both public and private healthcare providers, but also for companies offering health data tracking through smart devices, for instance. P4 demonstrated:

> "Someone has died of a heart attack. On the basis of clinical data, it is possible to back up and say 'Yeah, that caused it'. If that clinical data was expanded, for example with self-collected lifestyle, health, or other information, then we could combine these data and say that if you continue this, it will happen to you that this brisk exercise hobby of yours will defeat these symptoms behind them."

In conclusion, the opportunities regarding improved services are largely in the hands of public organizations. The domains whose importance was highlighted were mobility, health, and public administration. Especially within mobility, for instance with public transport, the importance of data spaces as platforms was downplayed in Finland, where the number of operating organizations is small. Instead, potential for data spaces within mobility was perceived larger in Central Europe, for instance.

6.2 Environmental sustainability

The theme of environmental sustainability arose in data analysis, as two opportunities gave out potential benefits in multiple directions. The opportunities discussed under this theme, real-time traffic updates and carbon emissions, both have other implications as well. Environmental sustainability is an underlying construct that has indirectly affected

businesses far and wide. In this case, the concrete sustainability benefits are saving fuel based on real-time traffic updates, and motivation to cut down carbon emissions due to their increasingly accurate monitoring and distribution enabled by data spaces. The environmental sustainability opportunities are presented in Table 4.

Table 4. Environmental sustainability

Sector	Opportunity	Applicable data space	Theme
Public & Private	Real-time traffic updates: saving fuel, avoiding roadwork and accidents	Mobility	**Environmental sustainability**
	Carbon emissions: tracking, reporting, and even distrubution across value chain	Green Deal, Manufacturing, Mobility, Agriculture, Energy	

Real-time traffic updates through data sharing enable logistics companies to re-evaluate their routes in case of accidents or roadwork. From an environmental perspective, this saves fuel and therefore reduces emissions. Regarding emissions, an entire value chain linked through a data space could each view and share their own exact emissions, which would make it easier to report and distribute them to all members. P5's statement on the subject was:

> "And one thing that immediately comes to mind is this carbon dioxide, CO2 emissions. If they are distributed to logistics operators, for example. Then we can estimate where this emission comes from and who pays the most. You'd think there's a motive. Then there will be emission fees or others. Then, if you don't produce those emissions yourself, then you have a big motivation to provide that data."

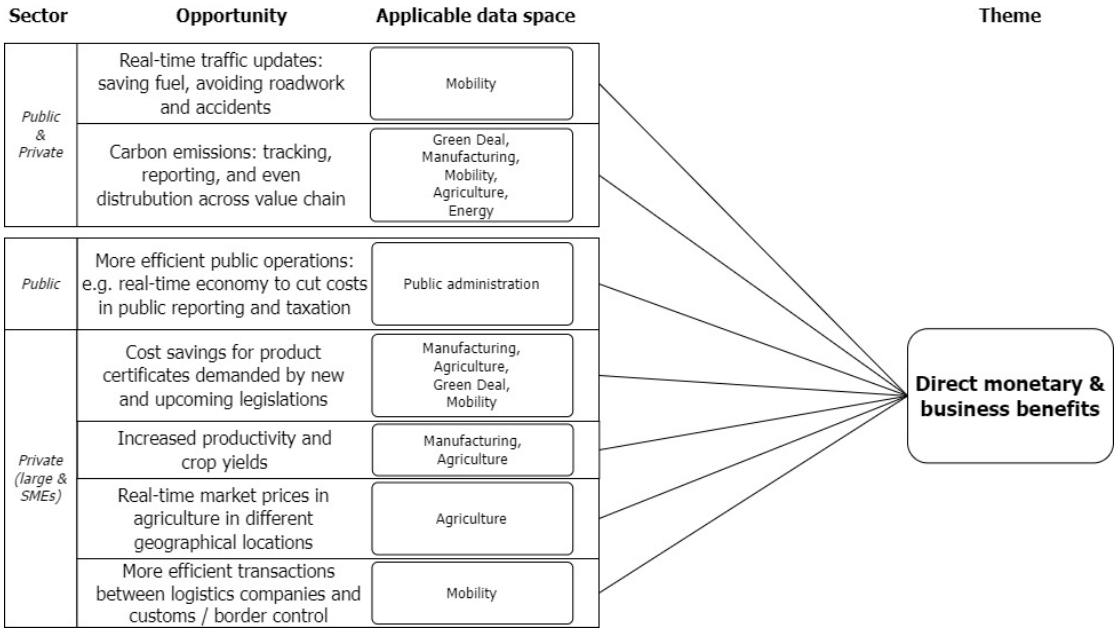
The motivations may be financial savings, moral values, or something else, but the benefit of such tracking and distribution would be a reduction in total carbon emissions. The applicable data spaces would be at least the Green Deal data space along with data spaces in all sectors where carbon emissions are generated, and where value chains or networks contain multiple organizations. As the sustainability benefits may have financial outcomes as well, they are covered in the next theme too.

6.3 Direct monetary & business benefits

When discussing companies' motivation to get into data space business, the biggest affecting factors are likely direct business benefits, i.e. how to make or save money by incorporating data spaces into one's business. Naturally, this theme of opportunities

affects private companies, both large and small, but there is potential for the public sector as well. These opportunities are summarized in Table 5.

Table 5. Direct monetary & business benefits



Revisiting the sustainability-related opportunities of real-time traffic updates and carbon emissions tracking, reporting, and distribution, they may have potential financial possibilities as well. First of all, as logistics companies save gas by using data to avoid traffic disruptions, they also save time and money. Not only are they saving on gas costs, but making deliveries on time mitigates the risk of delay sanctions and maximizes the use of transport equipment, for instance. In regard to carbon emissions tracking, the prime motivation to accurately report and reduce discharges for many companies regardless of their sector, is cost-savings in emission fees.

Direct monetary benefits for the public sector first arose in the interview with P1. These benefits include cutting costs in public processes, as P1 elaborated:

> "There's an initiative called real-time economy within the Ministry of Economic Affairs, where they would like companies to be able to report their accounting and everything, invoices, all in real time, right? So, then they also cut down the resources needed for reporting and taxes and all that stuff."

The remaining direct monetary and business benefits concern the private sector. Firstly, P4 pointed out that upcoming legislations demand new sustainability certificates on products to prove the methods and materials used in production are acceptable. They pointed out that the cost efficiency of data spaces and data economy is a way to produce

these certificates without them becoming too expensive in relation to the actual product price.

An opportunity shared at least by manufacturing and agriculture is enhanced productivity, or crop yield in agriculture. These benefits are reachable through efficient data collection and analysis enabled by data spaces, for instance in mining, as P1 illustrated. Another agriculture-related opportunity brought up by P2 is knowing real-time market prices in different geographical areas to optimize cultivation accordingly.

The last direct monetary benefit was related to the logistics sector presented by P2 and is here discussed under mobility data space. Logistics chains with many organizations and plenty of cross-border operations could save a lot of time on borders by providing freight data to border authorities and customs in electronic form. As mentioned before, in a field with a large number of organizations, such as logistics, a data space platform could be beneficial even for Finnish organizations, unlike in public transport.

6.4 Indirect business & fiscal opportunities

The last theme labeled indirect business and fiscal opportunities depicted in Table 6 discusses implications that may apply for both public and private organizations with long-term potential. In a sense these opportunities can be seen as strategic or risky, as the benefits are harder to measure or may take longer to realize. Relevant data spaces involve at least public administration, research, mobility, manufacturing, agriculture, and energy.

Table 6. Indirect business & fiscal opportunities

Sector	Opportunity	Applicable data space	Theme
Public	Better research based on data and cooperation, leading to new applied solutions	Public administration, Research	**Indirect business & fiscal opportunities**
	Better understanding on state affairs, e.g. actual export value	Public administration	
	Security: police and customs	Public administration, Mobility	
Private (large)	Flipping business from manufacturing to software production	Manufacturing	
Private (large & SMEs)	Co-creation with subcontractors	Manufacturing, Agriculture, Mobility, Energy	
	Enhanced data sharing and efficiency within multi-partner logistics value networks	Mobility	
	Value chain improvements, e.g. better maintenance	Manufacturing, Agriculture, Energy	
	Data space as a service: new business based on a **need** for data intermediaries	Manufacturing, Agriculture	

Starting with public sector opportunities, top-level research was one thing that came up in the discussions. As mentioned in the context of improved services, research can be tied to teaching, but also research itself could benefit from data sharing across data spaces, eventually providing potential applications to different business domains as well. P3 stated:

> "And of course, along with the research actors, one thing that should be paid attention to is the teamwork of these research actors. That they don't just work in their own silos, but that there would be profiling along with the discussion and that cutting-edge research would be created through this."

P5 brought up a fiscal opportunity that could eventually be reached through broad data space adoption: better understanding of the state economic situation. For instance, they pointed out that currently "we don't even know, for example, the value of Finnish exports". By real-time data tracking and reporting across data spaces, this goal could possibly become more achievable. P5 also pointed out that security is another issue that could benefit from data spaces, particularly the activities of police and customs if data on travelers and cargo was always available and updated in real-time.

As we move on to the private sector opportunities, we first inspect one that is most likely only applicable to large manufacturing companies: "flipping" business from manufacturing to software production, for instance data space platform provider. In analysis this was depicted as a large-company-only opportunity due to the resources such a transition would most likely require. P1 presented the subject like this:

> "…the market or sector where Finland is the market leader again, I come back to these big machineries. There's huge potential for these manufacturers to flip over to become a software provider too"

P1 also offered the elevator company Kone as an example that if profit margins in their original field drop, a company can take steps towards becoming a service or soft infrastructure provider along with or instead of manufacturing. But as P1 continued: "it's a slow painful road".

The remaining opportunities are more likely applicable to both large businesses and SMEs. The first one, co-creation with subcontractors, involves processes of combining forces with a company's own value network, and cooperating to create new business opportunities. A network of subcontractors is also a good example of a data economy ecosystem. According to P4:

> "For some time, large companies in Europe have been doing development together with a network of subcontractors, and then there is shared data that everyone uses."

Data spaces could present a new, efficient data sharing solution for such networks and enable more fruitful innovation. Co-creation is applicable especially within manufacturing, but also likely in other domains such as agriculture, mobility, and energy, for instance. The next opportunity of enhanced data sharing and efficiency within logistics networks has resembling qualities, as there are inevitably many different organizations within the same value network who benefit from effective data sharing amongst each other.

The opportunity of value chain improvements is also based on a data economy ecosystem of cooperating parties to create more business value. P1 introduces the opportunity in the context of large manufacturing companies:

> "They built those types of big machines, they invest billions of money in developing their machines and so through that machine they create a lot of data stream. Collection of data, now, for them it's very hard to - - implement, why would they openly share this data now to other people because they invest hundreds of millions in that. - - when you look at the value chain of that company you realize that services is also a part of it, maintenance, right? And today, the opportunity there is to improve maintenance and maintenance services through data tracking."

Again, efficient data tracking enabled by data spaces could improve the maintenance procedures of companies within a data economy ecosystem. Although P1 discusses large manufacturing companies, I argue that this opportunity applies to both large and small organizations who use machinery and require maintenance services, for instance within agriculture and energy as well. SIX Mobile Machines cluster introduced in chapter 4 is an example of such an ecosystem where machine manufacturers and utilizers share their data and analyze it to maximize the lifespan of the machines' batteries.

The last opportunity is in this case definitely not the least. The phrase that came up in many different contexts was *need*. According to experts, there seems to be a need for a data sharing platform, service, or service provider that businesses can take advantage of in their business. P5 referred to DataSpace Europe, the first registered data intermediary in Europe, and phrased it like this:

> "There are actors who are capable of that. And others benefit from the fact that it is such a *data space as a service*, which is the model that works best,

> because if the manufacturing industry has equipment and suppliers, they don't necessarily want to make a data space."

The key discovery was that companies may not want to steer from their key processes to create their own data spaces. P1 continues:

> "And an example, in the agriculture, farmers have no time to fiddle around with data. They don't have time to fiddle around with the computers. That's not the business, right? So, they just want to be able to sign up and go. And this is why the work that DataSpace Europe to build this platform is important for the farmers to get access to all that. And this effort is needed across different sectors."

P4 further adds that upcoming legislation on B2B data processing is likely to increase the need for a data sharing structure that does not require heavy integrations. Hence, the need for data space platform providers in a variety of domains may rise. P2 on the other hand reasons that what is actually needed is:

> "…to promote the sharing of data in a cost-effective, interoperable form. - - And data spaces can be the tool to get more comprehensive data more cost-effectively, so that all this can be done. But that's not necessarily the solution."

The applicable domains in this context are likely many but only a couple have been covered. Agriculture in Finland is represented by DataSpace Europe. Manufacturing was brought up in the interviews as a key example. The statements of P2, along with the small number of applied data space use cases, prove it is vital to examine what restrains these progressions. What underlying factors are there that affect data spaces in Finland or in general? These subjects also came up in the interviews and provide additional information to support the original research question.

6.5 Further findings

6.5.1 Restraints

It is safe to say there are plenty of opportunities in data spaces for a variety of organizations. However, the organizations must feel that the benefits will outweigh the required resources and emerging risks. In this subchapter we discuss restraints the expert participants brought forward, that might impede data space development and deployment.

Especially within the private sector companies worry about losing their data to competitors. P1 also stated that companies fear the upcoming Data Act will force them to

open up their data. P3 mentioned that some companies may fear losing potential business opportunities they have not yet recognized, by sharing their data with others.

A parallel issue is the difficulty of changing the business mindset when it comes to data sharing. P1 phrased it like this:

> "The point here is the mindset. Now, when for those companies who are already doing it, - - they understood very quickly they can create much more bigger value when working within an ecosystem instead of working alone isolated in their own corner, and these often are the people who want to be early adopters for data space - - those more forward-looking and more risk-taking companies that are actually jumping in first, and all these big companies, established companies, have really hard time trying to break this mindset thinking."

P1 also referred to "status quo", particularly in the case of large companies, as a possible restraint. P4 also stated Finnish companies are finding it difficult to change their mentality about co-creation with subcontractors, for example.

Larger, structural issues also came up as restraints. P1 and P3 both mentioned that in Finland there is no single large sector that would guide development, for example when compared to aviation in France and automotive industry in Germany, including the Catena-X automotive industry data ecosystem initiative. Additionally, P3 expressed a need for a governmental agenda for industry renewal within the manufacturing sector to reduce strategic discontinuation. P5 further added that Finland is lacking proper innovation funding that could also support broader data space adoption.

A couple of issues arose that apply to SME companies in particular. Firstly, P3 said:

> "When we talk about manufacturing SMEs, the challenge in Finland is that there is a lack of passion for growth, a lack of development resources, and quite often these SME operators follow in the wake of these big companies, following the actions of the big companies."

P4 further added that in SME companies, the owner motivation is central. If the owner only wishes to focus on the main business and work for that, the company will utilize its data very little, if at all. In these cases, the role of an external data space service provider becomes essential and the ease of use of their service equally important.

A legislative issue presented by P4 is that current legislation enables organizations the possibility to stay idle and do nothing regarding data processing. This causes a risk that when new EU data legislations become applicable, the business of many companies will

suffer as they attempt to meet the legal minimum requirements or face consequences in the form of monetary sanctions.

The last issue is related to the earlier opportunity of improved maintenance services and can potentially extend to other contexts as well. P1 stated that “it's very hard for small part of business to drive the big part.” This means that if a large company can benefit from data spaces in one of their operations, for instance maintenance, the company might still not be motivated to make large data space developments or investments if they see that only a small part of their business would benefit from those efforts. P2 summarized on a general level, that organizations will evaluate if the effort they put in is in balance with the achieved benefits.

6.5.2 Other implications

The semi-structured interviews enabled a relatively free-form discussion guided by the thesis key themes. In the interviews, a variety of other subjects related to data spaces and data economy were covered, besides the opportunities and restraints. These implications are covered in this subchapter.

The first point is the ease of use of data platform services, which has been mentioned previously already. I found this important, as multiple related examples came up in the interviews. In addition to users not willing to steer away from their core business in agriculture, for instance, it has other implications too. For instance, to enable real-time roadwork data sharing, be it across a data space or other integrated system, the roadwork contractor on site must be able to share information about an upcoming or ongoing roadwork so easily that no extensive application or platform training is needed. Additionally, in the real-time traffic data example, a new incentive would be required to motivate roadwork contractors to share the data. Either a reward of some kind or a legislative obligation were suggested.

According to P4, data sharing efforts in Finland are furthest specifically in traffic and that behind this development is a mandate given to the relevant organizations. P3 suggest that mandates should be given to other organizations as well to raise awareness of data spaces. P3 also suggests that the organizations already providing data space services should be brought forward to increase understanding even further.

In Finland, the agriculture sector is also far along in data space utilization. The company DataSpace Europe operates a data space service for Finnish agricultural operators. According to P5, this is because they realized the opportunity three years ago. Like DataSpace Europe, other service providers could also utilize a license fee as their income model. P4 compares this to the operations of telecommunications operators. P5 further added:

> "There is a pretty good probability that there are many of these actors in Finland who, just by changing the vocabulary, can start marketing what they do [in data space context]."

It was unclear, however, if the actors in question were telecommunications operators or something else. P1 also anticipated that getting into data space business is easier for companies who already have software as their core business.

Some cross-domain and cross-sectoral cooperation subjects were also covered. P3 highlighted the opportunity of Finland becoming an agile implementer in Europe through close cooperation between researchers and manufacturing and by utilizing the extensively funded Continental European initiatives and upcoming regulation. Additionally, P3 continued that in Finnish ministries the practical implementation of their projects should be taken under tighter focus instead of just making strategies.

Regarding startups and small and medium-sized companies, P3 and P4 suggested these smaller companies should be used as an innovation source within data solutions. P4 further added that innovations in SMEs would develop even more in data space cooperation. This could create a positive innovation cycle around data spaces. Additionally, P4 discussed that for smaller businesses participating in a data space could also provide an in-built ecosystem. P3 also introduced that another way to build ecosystems is through establishing non-governmental organizations between companies, for instance. In these NGOs the goals would be in business development, not data spaces.

Lastly, P5 made it clear that the playing field within data spaces is open at the moment and P1 supplemented that this is an opportunity for Finnish organizations to shape the standards of European data spaces. P5 elaborated:

> "If you're shaping and presenting your own views and then bringing what we offer - - then there's really only to be won."

Additionally, scalability and contextuality need to be considered: scalability in the sense that sectoral data spaces are easier at first than covering multiple sectors; contextuality in the sense that all organizations and data spaces may be different. Lastly, P4 noted that Finnish legislation is quite ready for data intermediaries compared to some other countries. This may give Finnish organizations a competitive edge for the time being.

7 Discussion

The empirical results of this thesis achieved through the semi-structured expert interviews strive to answer the research question "What opportunities do data spaces provide for Finnish organizations?" Through thematic analysis methods, four themes of opportunities were created: improved services, environmental sustainability, direct monetary and business benefits, and indirect business and fiscal opportunities. Additionally, in the interviews, several restraints and other important data space-related implications arose that supplement the original research question. Key findings are discussed in the following.

An important finding discovered in the literature review and supported by the interview study was that despite extensive regulation, funding, and effort, concrete data space use cases in Europe and Finland are minimal. The first registered data space platform provider, DataSpace Europe, is a small Finnish company operating within the agriculture sector in Finland. Especially in Europe but also in Finland, there are multiple data economy ecosystem initiatives and data space development projects, but those do not seem to be mature enough yet to implement data space operations in practice. A participating expert mentioned in the interview that the standardization required by data spaces in Europe is not yet ready or in accordance with Gaia-X and IDSA standards. These associations guide data space development on European and global levels and are responsible for creating common standards, which are important in data space contexts.

In previous literature, concrete data space opportunities and challenges are also scarce. The greatest contribution has been provided by Scerri et al. (2022) and their findings have been discussed in chapter 2.3 mainly with the support of European Union and Commission publications. Next, we will compare these prior findings depicted for instance in Figure 3 and Figure 4 with the findings of this study. As the main goal of this study was to discover concrete opportunities for organizations, the opportunities for citizens are not covered. Moreover, the opportunities are highlighted here above the challenges.

Starting from the private sector, the opportunities presented in this thesis provide more specific opportunities than Scerri et al.'s (2022) work. For organizations interested in data spaces, this thesis provides examples in a more understandable form. On the other hand,

the opportunities discussed by Scerri et al. (2022) provide broader, overarching themes that may incorporate this thesis's findings. For instance, "Innovative data-driven business models enabled by new value ecosystems" (Scerri et al. 2022) covers at least the following findings from Table 2: co-creation with subcontractors, enhanced data sharing within multi-partner logistics value networks, value chain improvements including maintenance, and data space as a service.

Scerri et al. (2022) provide an equal assessment of science-related opportunities as well. This thesis touches opportunities for research organizations but does not cover them so comprehensively. European data strategy also mentioned research-related data space opportunities but paid greater attention to the other nine sectoral data spaces (European Commission 2020). In Table 2 there are two research-related opportunities: better tools for teaching and research, and better research based on data and cooperation, leading to new applied solutions. These go under Scerri et al.'s (2022) heading of "Advancing science and open innovation through data availability" and are linked with the motives of the European Open Science Cloud (European Commission 2020).

Lastly, we compare the public sector opportunities discovered in this thesis with those of previous research. Largely these new findings complement those of Scerri et al. (2022), again providing some more tangible potentials. For instance, effective real-time public data processing can support a variety of goals, including updates on road security, cost-cuts in public operations and efficiency and safety at border control. Concrete examples on improved public digital services were also discovered, for instance in teaching, traveling, health, customer service, and sustainability.

In general, the results of this study offer a more concrete depiction of data space opportunities than previous research and the level of detail is higher. This thesis, however, has considered the potential for organizations and not individual citizens. In addition, the findings depicted here represent the knowhow and opinions of five researchers working in Finnish organizations and are necessarily not applicable on European or global level. The findings also emphasize some domains over others based on the participants' areas of expertise.

Regarding the restraints discovered in the interviews, the issues of trust and resistance to change also mentioned by Scerri et al. (2022) came up. Technology was not seen as a problem here, provided that sufficient funding is available. New findings were that in

Finland, or any individual country, the lack of a large guiding industry sector, like the automotive industry in Germany, is an issue that hinders large-scale development. Additionally, in smaller companies' context the motivation of owner was found pivotal. In larger companies, the data space benefits may have to extend throughout the organization to motivate funding an effort, as benefits for only small parts of business might be seen insufficient. Lastly, a new legal restraint was that current status quo of legislation may discourage organizations from putting effort into data utilization, including data space development, and therefore cause issues when new legislation comes into force in the upcoming years.

As we discuss the landscape of new and upcoming EU data legislation discussed in chapter 3.1, there are many interesting subjects whose surface has only been scraped in this thesis. With the European data strategy and Big 5 progressively taking place, businesses will need to consider massive amounts of new information that will affect their operations, and knowing all the end-results of this legislation may be very complex. The legislation may accelerate data space operations in Europe, but through its strict sanctions, there is also a risk that all companies may not keep up.

In the interviews, it also came up that many companies have said they are interested in the upcoming legislation, but only some have started preparing. P4 gave some thoughts:

> "But if I may now say a little as my own opinion, the smart ones are on the move. Most are waiting. And then there are also like 'Let's see then [when legislation comes into force], let's pay the fines, if there are any.'"

The implications of EU data laws for individuals and organizations, both public and private, is a vital research topic for researchers of legislation, particularly within business law. Also inspecting the specific details that the new legal data processing requirements pose for data spaces, particularly service providers, would be interesting.

The current status of the nine European data spaces was discussed in chapter 3.2. This landscape is also ever changing and keeping up with it is an important goal for data space and data economy research. In our empirical results, some data spaces were highlighted above others, and this may be due to either of two reasons. On the one hand, the interest and expertise area of the participants may have had such a strong effect on the focus. On the other hand, the findings may be generalizable enough to depict the current state of

Finnish data economy rather realistically and show that some fields are ahead of others in data utilization and data spaces.

When inspecting the results summarized in Table 2, the data spaces with best coverage were at least mobility (including logistics), public administration, manufacturing (industry), and agriculture. For instance, energy, health, finance, and Green Deal remained with less attention in the interview but some opportunities were extended to them as well during the analysis. Skills data space did not come up in the interviews and due its more abstract nature, the discussed opportunities were not recognized in skills data space context.

In the context of the nine European common data spaces, it would be interesting to understand the motives for making the specific limitation to those nine domains. It would be worthwhile to examine if other fields should be taken into account as well and if so, which fields. On the other hand, it could also be comprehensively examined if there are European data space initiatives in other specific fields that may not be included in those nine, or on the other hand, if there are sub-sectoral data spaces like there is in automotive manufacturing industry. For instance, if there are European data space projects in mining and forestry in addition to industry (manufacturing), or if logistics and traveling should have their own designated development programs outside the concept of mobility.

When reflecting the empirical results by the side of chapter 4 where Finnish organizations in the field of data spaces were discussed, it is particularly highlighted that practical data space use cases and concrete examples are few and far between, and the ones presented in chapter 4 were mostly covered in the interviews as well. In Finland there are cooperative data utilization projects, but data spaces are necessarily not the utilized or even the most suitable method.

An important point discovered in the interviews was that data space projects are not that different from other data economy and data sharing-related projects. For instance, there seem to be many more data economy initiatives in Finland besides specific data space initiatives. The contextuality brought up at the end of chapter 6 is a particularly central discovery, as not only data space projects have different contextual requirements, but the choice of data sharing or data integration method is also contextual. For instance, let us consider the example of public transport in Finland versus Central Europe. If there are only two or three entities sharing transport data – like Finland with its neighbors Estonia

and Sweden – a data space may not be necessary as point-to-point integrations would suffice. However, in Central Europe there are many more bordering neighbors that travelers may want to visit, and therefore a data space solution might be more efficient than a network of individual point-to-point connections of five, ten, or more countries.

This leads us to one more interesting area of future research. In the context of either Finnish or European organizations, it could be examined in which situations and organizational environments would a data space solution be most suitable, and when are point-to-point integrations or other data management methods more fitting. After providing this last future research implication, the next chapter concludes the thesis by summarizing its key findings.

8 Conclusions

This thesis strived to inspect the opportunities Finnish organizations may gain through a recently emerged data management solution: data spaces. In the European data strategy, common European data spaces were introduced as a tool for achieving a variety of political, ethical, and economic goals in the EU. Since then, the EU has published new legislation and allocated significant resources to the development of data spaces within the EU.

Finland is an example of an EU country where data space development is gradually emerging with a few concrete use cases already taking place. It has still, however, been unclear what concrete opportunities data spaces have for Finnish organizations. On the other hand, as the development has been only moderate, questions about potential restraints and other important implications have also arisen. To inspect the current status and future of data spaces in Finland, the following research questions were presented:

- What opportunities do data spaces provide for Finnish organizations?
    - What restraints or other implications may affect data space utilization in Finnish organizations?

Data space-related research and development has taken big leaps since the concept's introduction almost 20 years ago. In scientific literature the adjacent subject of data economy has been covered more extensively, for instance from ethical and ecosystem perspectives which are also central in European data spaces. Data space development and theoretical opportunities and challenges have also been covered in research. What has been lacking so far are concrete examples and implications of data space utilization.

To fill this knowledge gap and answer the research questions, a qualitative study through semi-structured expert interviews was conducted. The subject's novelty accounted for both the qualitative nature of the study as well as choosing experts of data spaces and data economy from Finnish organizations as participants. The transcribed interviews were then analyzed through thematic analysis methods. The coded transcriptions were categorized in a spreadsheet file and data space opportunities, restraints, and other findings were separated from each other. As the research focal point was on the opportunities, those were then inspected more closely and divided into four themes.

The four discovered themes of opportunities were improved services, environmental sustainability, direct monetary and business benefits, and indirect business and fiscal opportunities. These findings have been summarized in Table 2. Improved services concerned, for instance better teaching, health, taxation, travel, transport, and logistics. The beneficiaries of these services may be anyone from public or private organizations to individual citizens. One opportunity overlapped with the next theme: environmental sustainability. For instance, better traffic data services may help road users avoid roadwork and accidents, which simultaneously saves gas. Moreover, data spaces can support sustainability by enabling accurate tracking of organizations' carbon emissions.

The last two themes of direct benefits and indirect opportunities differ in the sense that direct benefits are easily quantifiable and can either help save costs in public or private operations or on the other hand increase productivity in for example manufacturing and agriculture. The indirect opportunities, however, require more strategic planning and long-term effort. For example, data spaces enable co-creation with partners or subcontractors, but here reaching measurable benefits may be prolonged.

In the interviews, a variety of restraints hindering data space implementation also arose. These include the difficulty of breaking current mindsets along with fears and worries about the risks of data sharing. Additionally in Finland, there seems to be no specific field of business to guide data space development. It was also discovered that at this point in time, there is a real chance to affect the future European data space landscape and that for companies providing data space as a service, the easy use of their service is pivotal.

The scope of this thesis included only Finnish organizations, as the participants worked in Finnish public and private organizations. To ensure reliability, the results are not generalized beyond Finnish organizations. Regarding future research, it would be important to understand the full effects new EU laws will pose regarding data spaces. Also, comparing data spaces with other data integration solutions would be interesting.

This thesis contributes to scientific literature by providing an image of the current status of Finnish data space operations and its outlooks. In practice, organizations may become more informed about the concrete opportunities and restraints of data spaces as their coverage in prior research has been minor. On a societal level, this thesis provides a concise depiction of the European data strategy along with new and upcoming data legislations and examples of how they may affect European organizations and citizens.

References

1001Lakes. (2024). *European networks in the data economy*. https://mediabank.businessfinland.fi/l/QwHfMwnvsX7P/f/PRRg, retrieved 1.4.2024.

Alonso, Á., Pozo, A., Cantera, J. M., De la Vega, F., & Hierro, J. J. (2018). Industrial Data Space Architecture Implementation Using FIWARE. *Sensors*, *18*(7), Article 7. https://doi.org/10.3390/s18072226.

Bassens, D., & Hendrikse, R. (2022). Asserting Europe's technological sovereignty amid American platform finance: Countering financial sector dependence on Big Tech? *Political Geography*, 97. https://doi.org/10.1016/j.polgeo.2022.102648.

Berkhout, V., Frey, C., Hertweck, P., Nestle, D., & Wickert, M. (2022). Energy Data Space. In B. Otto, M. ten Hompel, & S. Wrobel (Eds.), *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage* (pp. 329–341). Springer International Publishing. https://doi.org/10.1007/978-3-030-93975-5_20.

Berlage, T., Claussen, C., Geisler, S., Velasco, C. A., & Decker, S. (2022). Medical Data Spaces in Healthcare Data Ecosystems. In B. Otto, M. ten Hompel, & S. Wrobel (Eds.), *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage* (pp. 291–311). Springer International Publishing. https://doi.org/10.1007/978-3-030-93975-5_18.

Bogner, A., Littig, B., & Menz, W. (Eds.). (2009). *Interviewing Experts*. Palgrave Macmillan UK. https://doi.org/10.1057/9780230244276.

Braun, V., & Clarke, V. (2022). *Thematic analysis—A practical guide*. SAGE.

Bräutigam, T., Cunningham, F., Aholainen, M., Geus, M., & Toivanen, M. (2022). *EU-sääntely rakentaa reilumpaa datataloutta*. Sitra. https://www.sitra.fi/app/uploads/2022/06/sitra-eu-saantely-rakentaa-reilumpaa-datataloutta.pdf, retrieved 3.1.2024.

Catena-X. (2023). *Catena-X Operating Model*. https://catena-x.net/fileadmin/user_upload/Publikationen_und_WhitePaper_des_Vereins/CX_Operating_Modelv2.1_final.pdf, retrieved 28.2.2024.

Catena-X. (n.d.). *Who we are*. https://catena-x.net/en/about-us, retrieved 28.2.2024.

Chomczyk Penedo, A., & Trigo Kramcsák, P. (2023). Can the European Financial Data Space remove bias in financial AI development? Opportunities and regulatory

challenges. *International Journal of Law and Information Technology*, *31*(3), 253–275. https://doi.org/10.1093/ijlit/eaad020.

Clarke, V., & Braun, V. (2017). Thematic analysis. *The Journal of Positive Psychology*, *12*(3), 297–298. https://doi.org/10.1080/17439760.2016.1262613.

Curry, E. (2016). The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches. In J. M. Cavanillas, E. Curry, & W. Wahlster (Eds.), *New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe* (pp. 29–37). Springer International Publishing. https://doi.org/10.1007/978-3-319-21569-3_3.

Data Space 4.0. (n.d.). https://manufacturingdataspace-csa.eu/, retrieved 28.2.2024.

DataSpace Europe. (2022a). *DataSpace Europe liittyy Gaia-X- ja IDSA -organisaatioihin*. https://www.dataspace.fi/fi/uutinen/dataspace-europe-liittyy-gaia-x-ja-idsa-organisaatioihin, retrieved 16.3.2024.

DataSpace Europe. (2022b). *DataSpace Europe Oy perusti datanvälityspalvelu Tritomin*. https://www.dataspace.fi/fi/uutinen/dataspace-europe-oy-perusti-datanvalityspalvelu-tritomin-laittaa-datan-liikkumaan-ruokajarjestelman-toimijoiden-valilla, retrieved 16.3.2024.

DataSpace Europe. (n.d.). *Who are we?* https://www.dataspace.fi/en/company, retrieved 27.11.2023.

Data Space for Skills. (n.d.). *Welcome to the data space for* skills. https://www.skillsdataspace.eu/, retrieved 16.3.2024.

Deloitte. (2020). *A dynamic data driven approach to compliance monitoring*. https://www2.deloitte.com/sg/en/pages/energy-and-resources/articles/a-dynamic-data-driven-approach-to-compliance-monitoring.html, retrieved 23.2.2024.

Destination Earth. (n.d.). *Building a highly accurate digital twin of the Earth*. https://destination-earth.eu/, retrieved 5.3.2024.

Dong, X. L., Halevy, A., & Yu, C. (2009). Data integration with uncertainty. *The VLDB Journal*, *18*(2), 469–500. https://doi.org/10.1007/s00778-008-0119-9.

Drees, H., Kubitza, D., Theissen-Lipp, J., Pretzsch, S., & Langdon, C. (2021). *Mobility Data Space—First Implementation and Business Opportunities*. 27th ITS World Congress.

DSSC. (2023a). *DSSC Delivery Plan - Summary of assets publication*. https://dssc.eu/space/DDP/117211137/DSSC+Delivery+Plan+-+Summary+of+assets+publication, retrieved 25.1.2024.

DSSC. (2023b). *Glossary – 2. Core Concepts*. https://dssc.eu/space/Glossary/176554052/2.+Core+Concepts, retrieved 26.3.2024.

DSSC. (2024). *Knowledge base*. https://dssc.eu/page/knowledge-base, retrieved 26.3.2024.

Eclipse Foundation. (n.d.). *Eclipse Dataspace Components*. https://projects.eclipse.org/projects/technology.edc, retrieved 20.2.2024.

EIOPA. (n.d.). *Digital Operational Resilience Act (DORA)*. https://www.eiopa.europa.eu/digital-operational-resilience-act-dora_en, retrieved 5.3.2024.

Energydataspaces. (2024). *Synergies*. https://energydataspaces.eu/, retrieved 5.3.2024.

Enershare. (n.d.). *About*. https://enershare.eu/, retrieved 5.3.2024.

EONA-X. (n.d.a). *Who We Are.* https://eona-x.eu/page-about-1/, retrieved 19.2.2024.

EONA-X. (n.d.b). *Vision*. https://eona-x.eu/vision/, retrieved 19.2.2024.

Eriksson, I., & Kovalainen, A. (2008). *Qualitative Methods in Business Research*. SAGE Publications Ltd. https://doi.org/10.4135/9780857028044.

European Commission. (2020). *A European strategy for data*. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066, retrieved 27.11.2023.

European Commission. (2021a). *Artificial Intelligence Act*. https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF, retrieved 4.1.2024.

European Commission. (2021b). *Common European data spaces for agriculture and mobility.* https://digital-strategy.ec.europa.eu/en/library/common-european-data-spaces-agriculture-and-mobility, retrieved 28.2.2024.

European Commission. (2022a). Data Governance Act. *Official Journal of the European Union*. http://data.europa.eu/eli/reg/2022/868/oj.

European Commission. (2022b). Digital Markets Act. *Official Journal of the European Union*. https://doi.org/10.5040/9781782258674.

European Commission. (2022c). Digital Services Act. *Official Journal of the European Union*. https://doi.org/10.5040/9781782258674.

European Commission. (2023a). *Financial data access and payments package*. https://finance.ec.europa.eu/publications/financial-data-access-and-payments-package_en, retrieved 5.3.2024.

European Commission. (2023b). *Public procurement data space*. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52023XC0316(02), retrieved 16.3.2024.

European Commission. (n.d.a). *All Data 4 Green Deal - An Integrated, FAIR Approach for the Common European Data Space*. https://cordis.europa.eu/project/id/101061001, retrieved 5.3.2024.

European Commission. (n.d.b). *Data space for security and law enforcement*. https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/digital-2022-cloud-ai-02-sec-law, retrieved 16.3.2024.

European Grid Infrastructure. (n.d.). *GREAT*. https://www.egi.eu/project/great-project/, retrieved 5.3.2024.

European Health Data Space. (n.d.). *The European Health Data Space (EHDS)*. https://www.european-health-data-space.com/, retrieved 1.3.2024.

European Parliament, & Council of the European Union. (2023). *Data Act*. https://data.consilium.europa.eu/doc/document/PE-49-2023-INIT/en/pdf, retrieved 4.1.2024.

European Union. (2022). *Report on open finance*. https://finance.ec.europa.eu/document/download/226bcd0a-fff7-4fbd-9664-a84f50122101_en?filename=2022-10-24-report-on-open-finance_en.pdf, retrieved 5.3.2024.

European Union. (n.d.). *Common European Skills Data Space*. https://op.europa.eu/en/web/eu-vocabularies/europass, retrieved 16.3.2024.

Fintraffic. (n.d.). *Liikenteen dataekosysteemi*. https://www.fintraffic.fi/fi/liikenteenekosysteemi, retrieved 28.3.2024.

Fiware Foundation. (2024). *EU Data Sp4ce*. https://www.fiware.org/project/eu-data-sp4ce.html, retrieved 28.2.2024.

Franklin, M., Halevy, A., & Maier, D. (2005). From databases to dataspaces: A new abstraction for information management. *ACM SIGMOD Record*, *34*(4), 27–33. https://doi.org/10.1145/1107499.1107502.

Franklin, M., Halevy, A., & Maier, D. (2008). A first tutorial on dataspaces. *Proceedings of the VLDB Endowment*, *1*(2), 1516–1517. https://doi.org/10.14778/1454159.1454217.

Fraunhofer Institute. (2020). *Agricultural Data Space (ADS)*. Fraunhofer. https://www.iese.fraunhofer.de/content/dam/iese/publication/agricultural-data-space-cognac-fraunhofer-iese.pdf, retrieved 28.2.2024.

Gaia-X. (2021). *Gaia-X Domain Finance*. https://www.bmwk.de/Redaktion/DE/Publikationen/Digitale-Welt/211116-pp-finance.pdf?__blob=publicationFile&v=10, retrieved 5.3.2024.

Gaia-X. (n.d.a). *GXDCH.* https://gaia-x.eu/gxdch/, retrieved 27.11.2023.

Gaia-X. (n.d.b). *Who are we?* https://gaia-x.eu/who-we-are/association/, retrieved 24.11.2023.

Galletta, A., & Cross, W. E. (2013). *Mastering the Semi-Structured Interview and Beyond: From Research Design to Analysis and Publication*. New York University Press. http://ebookcentral.proquest.com/lib/kutu/detail.action?docID=1187368.

German Federal Ministry for Economic Affairs and Climate. (2024). *Financial Big Data Cluster (FBDC)*. https://www.bmwk.de/Redaktion/EN/Artikel/Digital-World/GAIA-X-Use-Cases/financial-big-data-cluster-fbdc.html, retrieved 28.3.2024.

Gouriet, M., Barancourt, H., Boust, M., Calvez, P., Laskowski, M., Taillandier, A.-S., Tilman, L., Uslar, M., & Warweg, O. (2022). The Energy Data Space: The Path to a European Approach for Energy. In B. Otto, M. ten Hompel, & S. Wrobel (Eds.), *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage* (pp. 535–575). Springer International Publishing. https://doi.org/10.1007/978-3-030-93975-5_33.

GREAT Project. (n.d.). *The GREAT project – Green Deal Data Space*. https://www.greatproject.eu/, retrieved 5.3.2024.

Green Deal Dataspace. (2024a). https://green-deal-dataspace.eu/, retrieved 5.3.2024.

Green Deal Dataspace. (2024b). *About us*. https://green-deal-dataspace.eu/about-us/, retrieved 5.3.2024.

Halevy, A., Franklin, M., & Maier, D. (2006). Principles of dataspace systems. *Proceedings of the Twenty-Fifth ACM SIGMOD-SIGACT-SIGART Symposium*

on Principles of Database Systems, 1–9. https://doi.org/10.1145/1142351.1142352.

Health-X. (n.d.). *The future of healthcare*. https://www.health-x.org/en/home, retrieved 1.3.2024.

Heikkilä, M., Ahmad, F., & Heikkilä, J. (2023). Unlocking the Potential of Data-Driven Business Models: An Empirical Investigation into the Role of Ecosystems and Fair Data Use. 85–106. https://doi.org/10.18690/um.fov.4.2023.6.

Hillermeier, O., Punter, M., Schweichhart, K., & Usländer, T. (2021). *Data Sovereignty – Critical Success Factor for the Manufacturing Industry* (Version 1.0). https://doi.org/10.5281/ZENODO.5675873.

Hirsjärvi, S., & Hurme, H. (1995). *Teemahaastattelu* (7th ed.). Yliopistopaino.

Hirsjärvi, S., Remes, P., & Sajavaara, P. (2013). *Tutki ja kirjoita* (15th–17th ed.). Tammi.

Horgan, D., Hajduch, M., Vrana, M., Soderberg, J., Hughes, N., Omar, M. I., Lal, J. A., Kozaric, M., Cascini, F., Thaler, V., Solà-Morales, O., Romão, M., Destrebecq, F., & Sky Gross, E. (2022). European Health Data Space—An Opportunity Now to Grasp the Future of Data-Driven Healthcare. *Healthcare*, *10*(9), Article 9. https://doi.org/10.3390/healthcare10091629.

Hyyrönmäki, J. (2022). *Datan välityspalvelut panevat ruoka-alan datan liikkumaan ja tuottamaan*. https://www.data-avaruus.fi/fi/artikkelit/datan-valityspalvelut-panevat-ruoka-alan-datan-liikkumaan-ja-tuottamaan, retrieved 27.11.2023.

IBM. (2023.) *Data pool.* https://www.ibm.com/docs/en/product-master/12.0.0?topic=gds-data-pool, retrieved 12.12.2023.

IBM. (n.d.a). *Data compliance software and solutions*. https://www.ibm.com/products/guardium-insights/data-compliance, retrieved 23.2.2024.

IBM. (n.d.b) *What is a data lake?* https://www.ibm.com/topics/data-lake, retrieved 12.12.2023.

IDSA. (n.d.a.). *Our mission and vision.* https://internationaldataspaces.org/why/, retrieved 23.11.2023.

IDSA. (n.d.b.). *Data sovereignty*. https://internationaldataspaces.org/why/data-sovereignty/, retrieved 23.11.2023.

IDSA. (n.d.c.). *Data spaces. Where the future of data happens.* https://internationaldataspaces.org/why/data-spaces/, retrieved 26.11.2023.

IOXIO. (2024). *Dataspace*. https://ioxio.com/dataspace, retrieved 28.3.2024.

Jarke, M., Otto, B., & Ram, S. (2019). Data Sovereignty and Data Space Ecosystems. *Business & Information Systems Engineering*, *61*(5), 549–550. https://doi.org/10.1007/s12599-019-00614-2.

Kaivola, T. (2023). *Data Governance Act, DGA*. Lost in Data Laws? EU's Big 5 and Beyond lecture series, Turku. https://ipr.mrooms.net/course/view.php?id=73#section-5, retrieved 9.1.2024.

Kalmar, R., Rauch, B., Dörr, J., & Liggesmeyer, P. (2022). Agricultural Data Space. In B. Otto, M. ten Hompel, & S. Wrobel (Eds.), *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage* (pp. 279–290). Springer International Publishing. https://doi.org/10.1007/978-3-030-93975-5_17.

Koskinen, J., Knaapi-Junnila, S., & Rantanen, M. M. (2019). What if we Had Fair, People-Centred Data Economy Ecosystems? *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation*, 329–334. https://doi.org/10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00100.

Lautanala, J. (2024). *Fintraffic & Data Economy in EU*. https://mediabank.businessfinland.fi/l/QwHfMwnvsX7P/f/PRRg, retrieved 28.3.2024.

Lindroos-Hovinheimo, S. (2023). *DSA in General*. Lost in Data Laws? EU's Big 5 and Beyond lecture series, Turku. https://ipr.mrooms.net/course/view.php?id=73#section-3, retrieved 8.1.2024.

Marcus, J. S., Martens, B., Carugati, C., Bucher, A., & Godlovitch, I. (2022). *The European Health Data Space* (SSRN Scholarly Paper 4300393). https://doi.org/10.2139/ssrn.4300393.

Marelli, L., Stevens, M., Sharon, T., Van Hoyweghen, I., Boeckhout, M., Colussi, I., Degelsegger-Márquez, A., El-Sayed, S., Hoeyer, K., van Kessel, R., Zając, D. K., Matei, M., Roda, S., Prainsack, B., Schlünder, I., Shabani, M., & Southerington, T. (2023). The European health data space: Too big to succeed? *Health Policy*, *135*, 104861. https://doi.org/10.1016/j.healthpol.2023.104861.

McKinsey & Company. (2022). *What are Industry 4.0, the Fourth Industrial Revolution, and 4IR?* https://www.mckinsey.com/featured-insights/mckinsey-

explainers/what-are-industry-4-0-the-fourth-industrial-revolution-and-4ir#/, retrieved 28.2.2024.

Mertens, C., Alonso, J., Lázaro, O., Palansuriya, C., Böge, G., Nizamis, A., Rousopoulou, V., Ioannidis, D., Tzovaras, D., Touma, R., Tarzán, M., Mallada, D., Figueiras, P., Costa, R., Graça, D., Garcia, G., Laibarra, B., Celaya, A., Sobonski, P., … Poulakidas, A. (2022). A Framework for Big Data Sovereignty: The European Industrial Data Space (EIDS). In E. Curry, S. Scerri, & T. Tuikka (Eds.), *Data Spaces: Design, Deployment and Future Directions* (pp. 201–226). Springer International Publishing. https://doi.org/10.1007/978-3-030-98636-0_10.

Merton, R. K., Fiske, M., & Kendall, P. L. (1956). *The Focused Interview*. The Free Press.

Ministry of Finance. (n.d.). *EU digital statutes lay the ground rules for operating in the digital age.* https://vm.fi/en/digital-statutes-in-the-eu, retrieved 9.1.2024.

Mobility Data Space. (n.d.a). *Mobility Data Space*. https://mobility-dataspace.eu/, retrieved 20.2.2024.

Mobility Data Space. (n.d.b). *The MDM in the Mobility Data Space*. https://www.mobility-data-space.de/en/the-mdm-in-the-mobility-data-space.html, retrieved 20.2.2024.

Montero-Pascual, J. J., Finger, M., & De Abreu Duarte, F. M. (2023). *Creating a common European mobility data space*. European University Institute. https://doi.org/10.2870/521152.

Nanni, R., & Napolitano, M. (2024). Dataspaces, public administration and collective rationality: Opportunities and limits for data-driven policy-making. *Public Money & Management*, 1–7. https://doi.org/10.1080/09540962.2023.2298260.

Oliveira, M. I. S., Barros Lima, G. de F., & Farias Lóscio, B. (2019). Investigations into Data Ecosystems: A systematic mapping study. *Knowledge and Information Systems*, *61*(2), 589–630. https://doi.org/10.1007/s10115-018-1323-6.

OmaKanta. (n.d.). *OmaKanta*. https://www.kanta.fi/omakanta, retrieved 1.3.2024.

Omega-X. (n.d.). *Home*. https://omega-x.eu/, retrieved 5.3.2024.

Otto, B., Jürjens, J., Schon, J., Auer, S., Menz, N., Wenzel, S., & Cirullies, J. (2016). *White paper Industrial Data Space*. Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.

https://www.fraunhofer.de/content/dam/zv/en/fields-of-research/industrial-data-space/whitepaper-industrial-data-space-eng.pdf, retrieved 28.2.2024.

Pampus, J., Jahnke, B.-F., & Quensel, R. (2022). Evolving Data Space Technologies: Lessons Learned from an IDS Connector Reference Implementation. In T. Margaria & B. Steffen (Eds.), *Leveraging Applications of Formal Methods, Verification and Validation. Practice* (pp. 366–381). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-19762-8_27.

Platform Industrie 4.0. (n.d.). *Manufacturing-X*. https://www.plattform-i40.de/IP/Navigation/EN/Manufacturing-X/Manufacturing-X.html, retrieved 29.2.2024.

Platoon project. (n.d.). *Home*. https://platoon-project.eu/, retrieved 5.3.2024.

Pretzsch, S., Drees, H., & Rittershaus, L. (2022). Mobility Data Space. In B. Otto, M. ten Hompel, & S. Wrobel (Eds.), *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage* (pp. 343–361). Springer International Publishing. https://doi.org/10.1007/978-3-030-93975-5_21.

Pullmann, J., Petersen, N., Mader, C., Lohmann, S., & Kemeny, Z. (2017). Ontology-based information modelling in the industrial data space. *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 1–8. https://doi.org/10.1109/ETFA.2017.8247688.

Rantanen, M. M., Hyrynsalmi, S., & Hyrynsalmi, S. M. (2019). Towards Ethical Data Ecosystems: A Literature Study. *2019 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, 1–9. https://doi.org/10.1109/ICE.2019.8792599.

Sarma, A. D., Dong, X., & Halevy, A. (2008). Bootstrapping pay-as-you-go data integration systems. *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, 861–874. https://doi.org/10.1145/1376616.1376702.

Sarma, A. D., Dong, X. L., & Halevy, A. Y. (2011). Uncertainty in Data Integration and Dataspace Support Platforms. In Z. Bellahsene, A. Bonifati, & E. Rahm (Eds.), *Schema Matching and Mapping* (pp. 75–108). Springer. https://doi.org/10.1007/978-3-642-16518-4_4.

Scerri, S., Tuikka, T., Lopez de Vallejo, I., & Curry, E. (2022). Common European Data Spaces: Challenges and Opportunities. In *Data Spaces: Design, Deployment and Future Directions* (pp. 337–357). Springer International Publishing AG.

http://ebookcentral.proquest.com/lib/kutu/detail.action?docID=7083155, retrieved 26.1.2024.

Sestino, A., Kahlawi, A., & De Mauro, A. (2023). Decoding the data economy: A literature review of its impact on business, society and digital transformation. *European Journal of Innovation Management*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/EJIM-01-2023-0078.

Shabani, M. (2022). Will the European Health Data Space change data sharing rules? *Science*, *375*(6587), 1357–1359. https://doi.org/10.1126/science.abn4874.

Sinipuro, J. (2023). *Datapellon uudisraivaajat*. https://www.dataspace.fi/fi/uutinen/datapellon-uudisraivaajat, retrieved 16.3.2024.

Sinipuro, J. (2024). *Euroopan ensimmäinen rekisteröity datan välityspalvelu*. https://www.dataspace.fi/fi/uutinen/euroopan-ensimmainen-rekisteroity-datan-valityspalvelu, retrieved 16.3.2024.

Sitra. (2023). *IOXIO ja DataSpace Europe tarjoavat työkalut datan jakamiseen, jotta syntyisi uutta liiketoimintaa*. https://www.sitra.fi/caset/ioxio-ja-dataspace-europe-tarjoavat-tyokalut-datan-jakamiseen-jotta-syntyisi-uutta-liiketoimintaa/, retrieved 3.1.2024.

Sitra. (n.d.). *Datatalouden kiinnostavimmat*. https://www.sitra.fi/hankkeet/datatalouden-kiinnostavimmat/#valitut, retrieved 28.3.2024.

SIX Sustainable Industry X. (2023). *SIX Mobile Work Machines nousi Sitran Datatalouden kiinnostavimmat -listalle*. https://www.six.fi/post/six-mobile-work-machines-nousi-sitran-datatalouden-kiinnostavimmat-listalle, retrieved 28.2.2024.

SIX Sustainable Industry X. (2024). *Sustainable Industry X*. https://www.six.fi/, retrieved 28.2.2024.

Sphinx Project. (n.d.). *Overview*. https://sphinx-project.eu/about-overview/, retrieved 1.3.2024.

Tardieu, H. (2022). Role of Gaia-X in the European Data Space Ecosystem. In B. Otto, M. ten Hompel, & S. Wrobel (Eds.), *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage* (pp. 41–59). Springer International Publishing. https://doi.org/10.1007/978-3-030-93975-5_4.

Taylor, P. (2023). *Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025.*

https://www.statista.com/statistics/871513/worldwide-data-created/, retrieved 23.11.2023.

Usländer, T., & Teuscher, A. (2022). Industrial Data Spaces. In B. Otto, M. ten Hompel, & S. Wrobel (Eds.), *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage* (pp. 313–327). Springer International Publishing. https://doi.org/10.1007/978-3-030-93975-5_19.

Wiedijk, S., & Roman, V. D. (2023). *Intro to Big 5 and DMA in general*. Lost in Data Laws? EU's Big 5 and Beyond lecture series, Turku. https://ipr.mrooms.net/course/view.php?id=73#section-1, retrieved 4.1.2024.

Yle. (n.d.). *Apotti-hanke*. https://yle.fi/t/18-273882/fi, retrieved 1.3.2024.

Zillner, S., Gomez, J. A., García Robles, A., Hahn, T., Le Bars, L., Petkovic, M., & Curry, E. (2021). Data Economy 2.0: From Big Data Value to AI Value and a European Data Space. In E. Curry, A. Metzger, S. Zillner, J.-C. Pazzaglia, & A. García Robles (Eds.), *The Elements of Big Data Value: Foundations of the Research and Innovation Ecosystem* (pp. 379–399). Springer International Publishing. https://doi.org/10.1007/978-3-030-68176-0_16.

Appendices

Appendix 1 Consent form (English)

European data spaces and their potential for Finnish organizations

Master's thesis, Antti Tiekso, Turku School of Economics

Consent to take part in research

- I voluntarily agree to participate in this research study.
- I understand that even if I agree to participate now, I can withdraw at any time or refuse to answer any question without any consequences of any kind
- I understand that I can withdraw permission to use data from my interview **within two weeks after the interview, in which case the material will be deleted.**
- I have had the purpose and nature of the study explained to me in writing and I have had the opportunity to ask questions about the study.
- I agree with my interview being **audio and video recorded.**
- I understand that all information I provide for this study will be treated confidentially.
- I understand that in reporting the results of this research my identity will remain **pseudonymous.** This will be done by coding personal and other information so that they cannot be connected to a certain person.
- I understand that disguised extracts from my interview may be quoted in the thesis.
- I understand that the **original audio and video recordings** will be retained on the author's personal computer (password-protected) and University of Turku's server **until they are transcribed into a text file.**
- I understand that **my written / verbal consent and a transcript of my interview** in which all identifying information has been altered will be retained on the author's personal computer (password-protected) and University of Turku's server **until the thesis has been accepted by the university.**
- I understand that I am entitled to access the information I have provided at any time while it is in storage as specified above.
- I understand that I am free to contact the author of the thesis to seek further clarification and information.

Consent of participant

Participant will give their consent verbally at the beginning of the interview and / or in writing via email.

Signature of researcher

I believe the participant is giving informed consent to participate in this study.

_____ Antti Tiekso
Signature of researcher

_____
Date

Appendix 2 Consent form (Finnish)

Eurooppalaiset data-avaruudet ja niiden potentiaali suomalaisille organisaatioille

pro gradu -tutkielma, Antti Tiekso, Turun Kauppakorkeakoulu

Suostumus tutkimukseen osallistumisesta

- Osallistun tähän tutkimukseen vapaaehtoisesti.
- Ymmärrän, että vaikka päätän osallistua nyt, voin vetäytyä missä tahansa vaiheessa tutkimusta ja kieltäytyä vastaamasta mihin tahansa kysymykseen ilman minkäänlaisia seuraamuksia.
- Ymmärrän, että voin perua luvan käyttää haastattelussa antamiani tietojani **kahden viikon kuluessa** haastattelusta, missä tapauksessa materiaali tuhotaan.
- Tutkimuksen tarkoitus ja luonne on selitetty minulle kirjallisesti, ja minulla on ollut mahdollisuus kysyä kysymyksiä tutkimuksesta.
- Suostun siihen, että haastatteluni **ääni ja videokuva tallennetaan**.
- Ymmärrän, että kaikki tutkimuksessani tuottama informaatio käsitellään luottamuksellisesti.
- Ymmärrän, että tutkimuksen tuloksia raportoidessa henkilöllisyyteni säilyy **pseudonymisoituna**. Tämä toteutetaan koodaamalla henkilö- ja muut tiedot niin, ettei niitä voida yhdistää tiettyyn henkilöön.
- Ymmärrän, että tunnistamattomaksi muunnettuja poimintoja haastattelustani voidaan lainata tutkimuksessa.
- Ymmärrän, että ja alkuperäisiä ääni- ja videotallenteita säilytetään **tutkimuksen tekijän henkilökohtaisella tietokoneella (salasanalla suojattu) ja Turun yliopiston palvelimella, kunnes ne on litteroitu tekstitiedostoiksi.**
- Ymmärrän, että suullista / kirjallista suostumustani sekä litteroitua versiota haastattelustani, josta kaikki minut tunnistava tieto on muutettu, säilytetään tutkimuksen tekijän henkilökohtaisella tietokoneella (salasanalla suojattu) ja Turun yliopiston palvelimella, **kunnes yliopisto on hyväksynyt tutkimuksen.**
- Ymmärrän, että minulla on oikeus päästä käsiksi tuottamaani informaatioon milloin tahansa niin kauan kuin sitä säilytetään yllä kuvatun mukaan.
- Ymmärrän, että voin ottaa yhteyttä tutkimuksen tekijään pyytääkseni tarkennuksia ja lisätietoja.

Osallistujan suostumus

Osallistuja antaa suostumuksensa suullisesti haastattelun alussa ja / tai kirjallisesti sähköpostitse.

Tutkijan allekirjoitus

Uskon, että osallistuja antaa tietoisen suostumuksen osallistua tähän tutkimukseen.

_____ Antti Tiekso

Tutkimuksen tekijän allekirjoitus

_____

Päivämäärä

Appendix 3 Privacy notice

Privacy notice

1. Name of the register:

Participants' verbal consents for data use

2. Data Controller:

Antti Tiekso, +358** *** ****, **********@utu.fi

********** ** * **, Turku, Finland

3. Contact information of the responsible person:

Antti Tiekso, +358** *** ****, **********@utu.fi

4. Purpose and legal basis for the processing of personal data:

The research collects experts' views on the potentials of Finnish organizations within the European data spaces with interviews. Email addresses are used when sending out invitations to interviews. The interviews involve collecting personal data for recording consent and this data will be stored separately from the interview.

The legal basis for processing personal data in the Article 6 of the EU General Data Protection Regulation is:

☒ Processing is necessary for scientific research (public interest, Point 1a of the Article 6)

☒ Data subject has given their consent to processing personal data (consent, Point 1e of the Article 6)

5. Processed personal data:

The following information of the data subjects is stored in the register: **Name, organisation, title, email address**

6. Recipients and recipient groups of personal data:

The data will not be transferred or disclosed to third parties.

7. Information on transferring data to third countries:

Personal data will not be disclosed to parties outside the EU or the European Economic Area.

8. Retention period of personal data or criteria for its determination:

The recorded interviews will be transcribed into text files and the recordings will be destroyed apart from the verbal consent given at the beginning of the interview. Simultaneously, the transcripts will be pseudonymised by erasing identifiable personal data. Verbal consent including remaining personal data is destroyed as soon as the thesis is accepted.

9. Rights of the data subject:

The data subject has the right to access their personal data retained by the the Data Controller, the right to rectification or erasure of data, and the right to restrict or object the processing of data. The right to erasure is not applied in scientific or historic research purposes in so far as the right to erasure is likely to render impossible or seriously impair the achievement of the objectives of that processing.

The realisation of the right to erasure is assessed on a case-by-case basis.

The data subject has the right to lodge a complaint with the supervisory authority.

10. Information on the source of personal data:

In order to send the invitations to the interview, email addresses are collected from public websites. The other data is collected directly from those who participate in the interviews for the study.

11. Information on the existence of automatic decision-making, including profiling:

The data will not be used for automatic decision-making or profiling.

Appendix 4 Data management plan

UNIVERSITY OF TURKU
Library

Research data management plan for students

This document will help you plan how to manage your research data. More detailed instructions for each section are available online in the Research Data Management Guide for Students.

1. Research data

Research data refers to all the material with which the analysis and results of the research can be verified and reproduced. It may be, for example, various measurement results, data from surveys or interviews, recordings or videos, notes, software, source codes, biological samples, text samples, or collection data.

In the table below, list all the research data you use in your research. Note that the data may consist of several different types of data, so please remember to list all the different data types. List both digital and physical research data.

Research data type	**Contains personal details/information***	**I will gather/produce the data myself**	**Someone else has gathered/produced the data**	**Other notes**
Data type 1: *Interview video recordings*		x		Stored on my personal computer and backed up on the university's server and deleted after transcription
Data type 2: *Interview audio recordings*		x		Stored on my personal computer and backed up on the university's server and deleted after transcription
Data type 3: *Verbal consent*	x	x		Deleted after thesis accepted
Data type 4: *Transcribed interview data*		x		Deleted after thesis accepted
Data type 5: *Analyzed data*		x		Deleted after thesis accepted

* Personal details/information are all information based on which a person can be identified directly or indirectly, for example by connecting a specific piece of data to another, which makes identification possible. For more information about what data is considered personal go to the Office of the Finnish Data Protection Ombudsman's website

UNIVERSITY OF TURKU
Library

2. Processing personal data in research

If your data contains personal details/information, you are obliged to comply with the EU's General Data Protection Regulation (GDPR) and the Finnish Data Protection Act. For data that contains personal details, you must prepare a Data Protection Notice for your research participants and determine who is the controller for the research data.

I will prepare a Data Protection Notice** and give it to the research participants before collecting data ☒

The controller** for the personal details is the student themself ☒ the university ☐

My data does not contain any personal data ☐

** More information at the university's intranet page, Data Protection Guideline for Thesis Research

3. Permissions and rights related to the use of data

Find out what permissions and rights are involved in the use of the data. Consult your thesis supervisor, if necessary. Describe the use permissions and rights for each data type. You can add more data types to the list, if necessary.

3.1. Self-collected data

You may need separate permissions to use the data you collect or produce, both in research and in publishing the results. If you are archiving your data, remember to ask the research participants for the necessary permissions for archiving and further use of the data. Also, find out if the repository/archive you have selected requires written permissions from the participants.

Necessary permissions and how they are acquired

Data types 1–5: A consent form and privacy notice will be sent for the respondents before the interviews.

4. Storing the data during the research process

Where will you store your data during the research process?

In the university's network drive ☒
In the university-provided Seafile Cloud Service ☐
Other location, please specify: On personal computer protected by password / fingerprint ☒

The university's data storage services will take care of data security and backup files automatically. If you choose to store your data somewhere other than in the services provided by the university, please specify how you will ensure data security and file backups. Remember to make sure you know every time where you are saving the edited/modified data.

If you are using a smartphone to record anything, please check in advance where the audio or video will be saved. If you are using commercial cloud services (iCloud, Dropbox, Google Drive, etc.) and your data contains personal data, make sure the information you provide in the Data Protection Notice about data

migration matches your device settings. The use of commercial cloud services means the data will be transferred to third countries outside the EU.

5. Documenting the data and metadata

How would you describe your research data so that even an outsider or a person unfamiliar with it will understand what the data is? How would you help yourself recall years later what your data consists of?

5.1 Data documentation

Can you describe what has happened to your research data during the research process? Data documentation is essential when you try to track any changes made to the data.

To document the data, I will use:

A field/research journal ☐
A separate document where I will record the main points of the data, such as changes made, phases of analysis, and significance of variables ☒
A readme file linked to the data that describes the main points of the data ☐
Other, please specify: ☐

5.2 Data arrangement and integrity

How will you keep your data in order and intact, as well as prevent any accidental changes to it?

I will keep the original data files separate from the data I am using in the research process, so that I can always revert back to the original, if need be. ☒

Version control: I will plan before starting the research how I will name the different data versions and I will adhere to the plan consistently. ☒

I recognise the life span of the data from the beginning of the research and am already prepared for situations, where the data can alter unnoticed, for example while recording, transcribing, downloading, or in data conversions from one file format to another, etc. ☒

5.3 Metadata

Metadata is a description of you research data. Based on metadata someone unfamiliar with your data will understand what it consists of. Metadata should include, among others, the file name, location, file size, and information about the producer of the data. Will you require metadata?

I will save my data into an archive or a repository that will take care of the metadata for me. ☐

I will have to create the metadata myself, because the archive/repository where I am uploading the data requires it. ☐

I will not store my data into a public archive/repository, and therefore I will not need to create any metadata. ☒

UNIVERSITY OF TURKU
Library

6. Data after completing the research

You are responsible for the data even after the research process has ended. Make sure you will handle the data according to the agreements you have made. The university recommends a general retention period of five (5) years, with an exception for medical research data, where the retention period is 15 years. Personal data can only be stored as long as it is necessary. If you have agreed to destroy the data after a set time period, you are responsible for destroying the data, even if you no longer are a student at the university. Likewise, when using the university's online storage services, destroying the data is your responsibility.

What happens to your research data, when the research is completed?

I will destroy all data immediately after completion, because: After the thesis in complete and accepted, there will be no further need for the data. This is also promised in the consent form and privacy notice.

Remember to keep the data management plan updated throughout the research project.

Appendix 5 Interview structure

Rough interview structure

1. Preliminary question:
    a. As an expert, how would you define the concept of data space? What does it mean to you?
2. Before we go deeper into different organization types, on a general level:
    a. Is some particular business sector especially strong for Finnish organizations regarding data spaces? (Finland's strengths)
3. Private sector
    a. SMEs and startups
        i. Greatest potentials / opportunities for Finnish SMEs and startups?
        ii. What separates Finnish companies from others?
        iii. A specific data space or domain in particular?
        iv. The role: data space participant, service provider, other?
        v. Something else?
    b. Large companies
        i. Potentials for large Finnish companies? Are they different from those of SMEs?
        ii. What separates Finnish companies from others?
        iii. A specific data space or domain in particular?
        iv. The role: data space participant, service provider, other?
        v. Something else?
        vi. what separates Finnish companies from others?
4. Public sector
    a. Potentials / opportunities for public actors?
        i. For example, the state itself, public research and funding institutions, towns, municipalities, universities etc.
        ii. Who are the beneficiaries if the state joins data space activities?
            1. Companies, citizens, national economy, EU etc.?
        iii. Something else?
5. Others, for example non-profits, NGOs, or joint alliances
    a. Potentials / opportunities for these organizations?
        i. Who benefits?
        ii. Something else?
6. General questions (discussed at suitable stages of the interview)
    a. How many Finnish organizations are in data spaces or data economy business?
    b. What can Finnish organizations achieve by going into data space activities now?
    c. What does current cooperation with international entities (IDSA, Gaia-X) look like?
    d. How could Finnish organizations get involved in data space activities at EU-level?
    e. How does Finland compare with other countries?
7. Has anything come to mind that has not been discussed yet?